Spatial Interaction Models in a Big Data Grocery Retailing Environment

By

Philip David James Wilkinson

Submitted in accordance with the requirements for the degree of Doctor of Philosophy in Advanced Spatial Analysis

University College London

The Centre for Advanced Spatial Analysis

February 2023

I, Philip Wilkinson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources I confirm that this has been indicated in the thesis.

Acknowledgements

This thesis has been helped along by the guidance, support and patience of my two academic supervisors and CASA colleagues, Prof Adam Dennett and Dr Neave O'Clery. The final thesis would not have been possible without their help and advice to shape it into what it has become. This work would also not have been possible without the funding from dunnhumby and the supervision and support from Dr Richard Congram, Dr Nathanaël Bullier, Carlos Ardid, Rosie Prior and Dr Peter Riefer. I would like to thank them for their contributions of ideas, questions and insight into the data, and ensuring that the work and conclusions reflect the reality of grocery retailing.

I am also indebted to several of my CASA colleagues and the department. Firstly for accepting me onto the MSc in Smart Cities and Urban Analytics and secondly to all those who taught me over the year. It was the skills developed in this course that gave me the foundation to undertake the current research. In particular I would like to thank Dr Andrew MacLachlan for guiding me through the application process and being a continual source of both academic and friendly support throughout the research. I would also like to thank Dr Steven Gray for not only being a good friend but also a source of solid advice, guidance and support over this journey.

Finally, I would like to say a big thank you to my Family: Mum, Dad, Grandma, Grandad, and Nan who have been immensely supportive of whatever path I take and who have always been there for me. In particular I would like to mention my Grandma and Grandad who started me off on this university journey with an early trip to Cambridge University, inspiring me to push myself and achieve what I promised I would very early on. Although my Grandad is no longer with us he has, and continues to be, an endless source of inspiration. An especially massive thank you to Gabrielle Anderson who has been there through thick and thin throughout this process and someone who without which I would not have been able to finish this piece of research. Words cannot express how grateful I am to have had your love and support over this journey. Thank you.

Abstract

Grocery expenditure is responsible for around 10% of total household spend in the UK, making the grocery retail market worth over £200bn a year in 2021. The size of this market and the nature of retailing competition makes it important for retailers to make the right decisions. One such decision is the location of their stores for which there have been a number of changes in the location, format and channel of consumer interaction along with the methods that have been employed to determine new store location. In recent years it has been suggested that the spatial interaction model is the most appropriate method for estimating new store revenue and hence location. However, previous attempts to explore the performance of the spatial interaction model in grocery retailing have been limited by access to loyalty card data. In this thesis we show that these models are unable to account for the heterogeneity in store conditions and consumer behaviour to model total store revenue. Notably, we find that at the regional scale the size of the errors are such that these models are unlikely to be used consistently in practice for estimating store revenue or locating new stores. Furthermore, that the performance achieved in previous applications are unlikely to be consistently replicated. Thus our results demonstrate that the spatial interaction model in its current form is no longer appropriate for modelling grocery store revenue. It is anticipated that these results may become a starting point for the development and application of alternative forms of models and methods for predicting grocery retailing store revenue. Notably, such new methods must be able to account for recent changes in consumer behaviour such as convenience store shopping, multi-purpose trips and the growing influence of e-commerce, alongside changes in retailers interaction strategies.

Impact statement

This research extends the existing body of literature by developing and applying a spatial interaction model at a regional and yearly scale. The results show that the spatial interaction model, in its current form, is unlikely to produce consistent results that could be used in practice, therefore illuminating potential issues in their application at this scale. This research could therefore be used in the wider academic literature as a basis for the exploration of alternative forms of model and data with the aim of estimating grocery store revenue. It also highlights some of the difficulties in the application of spatial interaction models to domains with "non-complete" data, thus opening up the discourse on model calibration and estimation under this conditions in domains beyond grocery retailing.

This research was conducted under the supervision of dunnhumby and with access to anonymous loyalty card data from a national grocery retailer in the UK. The methodology developed as part of this thesis, notably the data pipeline and modelling implementation, will continue to be used and developed by dunnhumby so as to be used to consult with their clients on new retail location stores. Notably, this research has laid the foundation for future research partnerships in this area along the lines of the exploration of new methods and data in order to predict new store location. This includes the application of spatial interaction models in countries with less mature grocery retailing sectors or the development of alternative methods for estimating the revenue of convenience stores or online channel engagement in the UK.

Contents

Figures

Tables

Chapter 1

Introduction

Grocery retailing in the UK has undergone several changes over the last 60 years to move from a market driven by small format local grocery retailers to one dominated by a few national, and in some cases international, brands. These changes have been in response to social, economic and political factors that influenced retailers decisions and offerings in terms of the location, format and channels through which they interact with consumers. While retailers have adapted to these changing market conditions they have also had to change the way in which they make decisions. An important aspect of this has been how retailers make decisions about where to locate new stores, which stores should be closed and how to improve their offerings to consumers. Over the last 60 years the methods for these making these decisions has ranged from the use of site visits and gut feel by managers, to buffer and overlay analysis both by hand and by computer, through to the adaptation and implementation of spatial interaction models to estimate store revenue. With grocery expenditure responsible for around 10% of total household spend in the UK (JRF, 2022), making the grocery retail market over £200bn a year in 2021 (Statista, 2022), it is important for retailers to make the right decisions.

In this regard, spatial interaction models have been described as a core tool in spatial data modelling that are used to model and predict spatial flows (Rowe, et al., 2022). These models have a considerable history in the academic literature, dating back to the conception of the idea in relation to retailing by Reilly in 1929, and are based on the idea that flows of goods, services or people between two locations is proportional to the size of the origin and destination and inversely proportional to the distance between them. The long history of these models has meant that their formulation and understanding has undergone many adjustment and adaptations, including the development of parallels with the laws of physical science, integration with economic utility theory, the development of a family of spatial interaction models, through to recent models incorporating a wide range of factors and model formulations. However, these models have only recently been able to be used in retailing, and grocery retailing specifically, due to limitations in practice of understanding, data and computational resources. This has meant that it has often not been clear how accurate these models are in practice and whether they are appropriate for making store location decisions.

In regards to store location decisions recent literature has only begun to explore how these models may be used in practice in conjunction with loyalty card data from grocery retailers. These applications however have only encompassed small scale exploration of these models which have either been limited by store numbers or time scale due to strict data regulations. Thus, with recent pressures in the grocery retailing sector such as the rise of convenience shopping, multi-purpose trips and the development of e- and m-commerce channels by retailers, it is important to examine how spatial interaction models perform. This is to determine whether these model formulations are the most appropriate for estimating grocery store revenues and thus whether they can make store location decisions consistently and at scale. Thus, it is within this context that the aims of the thesis are developed.

1.1) Aims and Objectives

The aim of this thesis is to advance the current understanding of the behaviour and performance of spatial interaction models in a grocery retailing setting within the UK. In order to achieve this aim, a number of specific objectives are proposed:

1. To examine and review the current spatial interaction modelling literature in terms of the development and usage of models to understand which models would be most appropriate for the application to a grocery retailing scenario in the UK.

2. To review the literature on grocery retailing in the UK to be able to identify social, economic and political pressures that have, and are currently, influencing the market and to relate these influences and market developments to models that have been used to determine store location in practice.

3. To identify issues surrounding the implementation of spatial interaction models in practice and to develop a working model based on anonymised loyalty card data.

4. To develop and apply a spatial interaction model at a regional and yearly scale so as to identify how modelling performance changes and responds at scale and whether there are any factors that influence the modelling performance.

5. To replicate the modelling implementations from the most recent and up to date papers so as to examine whether the suggested results and performance from them can be replicated on data that we have available.

6. To implement and examine alternative forms of the spatial interaction model to identify the influence of additional store based factors on modelling performance.

7. To offer potential avenues for future research to explore so as to continue the development of the spatial interaction model in reference to its application in grocery retailing.

Each of these objectives will be met by the various chapters in this thesis. Table 1 below summarises where each objective will be tackled and addressed in which chapter(s)

Table 1 - Thesis objectives and corresponding chapters

Objective	**Corresponding chapter(s)**
1. To examine and review the current spatial interaction modelling literature in terms of the development and usage of models to understand which models would be most appropriate for the application to a grocery retailing scenario in the UK.	Chapter 2: Spatial Interaction Modelling History Chapter 4: City Model Application
2. To review the literature on grocery retailing in the UK to be able to identify social, economic and political pressures that have, and are currently, influencing the market and to relate these influences and market developments to models that have been used to determine store location in practice.	Chapter 3: Grocery Retail Location
3. To identify issues surrounding the implementation of spatial interaction models in practice and to develop a working model based on anonymised loyalty card data.	Chapter 2: Spatial Interaction Modelling History Chapter 4: City Model Application
4. To develop and apply a spatial interaction model at a regional and yearly scale so as to identify how modelling performance changes and responds at scale and whether there are any factors that influence the modelling performance.	Chapter 5: Regional Model Application

5. To replicate the modelling implementations from the most recent and up to date papers so as to examine whether the suggested results and performance from them can be replicated on data that we have available.	Chapter 6: Modelling Scenarios Replication
6. To implement and examine alternative forms of the spatial interaction model to identify the influence of additional store based factors on modelling performance.	Chapter 7: Alternative Model Implementations
7. To offer potential avenues for future research to explore so as to continue the development of the spatial interaction model in reference to its application in grocery retailing.	Chapter 8: Future Modelling Implementations

1.2) Thesis Structure

As Table 1 above reveals, most of the research objectives will be tackled by individual chapters although some of the broader application objectives will be addressed across multiple chapters. Whilst chapters 2, 3 and 4 in this thesis are mostly review orientated, the introduction of different ideas and concepts or the further examination of a specific concept within each chapter will require a thorough examination of the relevant literature throughout this piece of work.

Chapter 2 begins the thesis and immediately starts by addressing the first objective through a thorough review of the development and application of spatial interaction models in the academic literature. This review highlights how the spatial interaction become the prominent retail location theory as used in practice over other theories that were developed around the same time as the principle of minimum differentiation, bid rent theory and central place theory. This Chapter then continues by tracing the history of the adaptation and implementation of the spatial interaction model, including how each iteration was developed in response to perceived issues of the model formulation at that time. It then concludes by proposing that the production constrained spatial interaction model from the Wilsonian family of models is the most appropriately designed model for the purpose of estimating grocery store revenue and location. The examination of this lays the foundation for the examination of grocery retailing in the UK and the application of the model in Chapters 4, 5, 6 and 7.

Chapter 3 then tackles the second objective of the thesis by examining the relevant literature surrounding the changes in the grocery retailing market in the UK over the last 60 years and identifying the methods that have been used over the same time period to determine new store locations. This review highlights the progression of the UK grocery retailing market through the initial period of local small store development through to the national scale concentration of the market in hands of a few brands. It also identifies recent pressures and changes in the market including the rise of convenience shopping behaviour, multi-purpose trips and the development of e- and m-commerce platforms. In doing so, the chapter also traces the history of the methods used to identify new store locations in reference to this history, and how such methods have been adapted in response to increase in data and computing power alongside changes in competition and consumer behaviour within the sector. This work is thus important to identify the need for spatial interaction models in the grocery retailing sectors and also factors that may influence model performance such as changing consumer behaviour.

Chapter 4 builds on the work of the previous two chapters by highlighting unresolved issues in the literature in regards to spatial interaction modelling implementation, so as to address objective three. This chapter does so by discussing different methods of model calibration in the literature, identifying that a Poisson Regression formulation would be the most appropriate method to use in this scenario. Furthermore, methods of model evaluation, so as to ensure a good model fit and improvements in model specification, are discussed and how this relates to those metrics used in previous research. These discussions, along with those in the previous chapters, are then used to inform the application and evaluation of an initial spatial interaction model applied at a city level scale. The results from this model are then used to illuminate potential issues in modelling specification or data integration that could affect the modelling implementations in further chapters. Notably that there are no discernable differences between the different distance decay formats and that the models perform poorly at estimating small store format revenue.

Chapters 5, 6 and 7 then provide the core contribution of this work through the evaluation of objectives 4, 5 and 6 by developing and applying spatial interactions models to large spatial and temporal scale grocery retailing scenarios. Chapter 5 builds on the foundations laid out in

Chapter 4 by developing and applying a production constrained spatial interaction model at a regional scale across three different consumer regions in the UK. This Chapter achieves this by firstly identifying the scale of the application in terms of the value of sales flows over distance, the consumer groups represented and relationship between loyalty card and total store revenue for all three regions. This is then followed up by the implementation of a system wide non-disaggregated model, an origin-disaggregated model and a model with further data integrations across all three regions for a single week. The results presented from these applications show an inability to replicate the performance seen in previous research and thus the results are further explored in this chapter through correlation analysis, the application of the models over an entire year and a cross-validation study. This further exploration showed that the results were not the consequence of factors affecting a single week, that the total revenue errors were insensitive to alternative parameters and that no single factor could be consistently identified that could potentially be influencing modelling performance. Thus, leading to questions as to the appropriateness of the current modelling formulation in estimating grocery store revenue.

In light of the results from Chapter 5, Chapter 6 attempts to replicate the analysis undertaken in the recent literature in terms of both the scale of model implementation and the calibration method to evaluate the consistency of the performance of the spatial interaction model and achieve objective 5. This Chapter revisits the recent literature in terms of differences in model specification and application to identify potential factors that could be influencing modelling performance from the previous Chapter. It then proceeds through an investigation of the different scales of modelling implementation, as measured by the number of stores, and through the implementation of a iterative calibration methodology at the regional scale. What this chapter shows is the inability of the current modelling formulation and data limitations to be consistently replicate the results seen in literature at any scale. Thus, suggesting the inappropriateness of the spatial interaction model, in its current form and data, is unable to accurately and consistently estimate grocery store revenue.

Chapter 7 then builds on the results in Chapter 6 by attempting to resolve the issues highlighted through the implementation of alternative model formulations. These models focus on the adaptation of the competing destination model, integrating store age into the original model and applying both the original and origin-disaggregated model to large basket data. In doing so, this chapter identifies the inability of these alternative modelling formulations to resolve the issues in modelling performance from the previous chapters. Specifically, while the influences of competition, agglomeration, store age and behavioural subsets are identified, implemented and discussed, none of the alternative model formulations implemented lead to improvements in modelling outcomes as given a mean error close to zero, low standard deviation of store errors or an average trip distance close to one. Thus, in light of these results, none of the models are pursued further while suggestions are offered for potential future research avenues.

Chapter 8 then looks to offer potential solutions or future research directions in response to the results presented. Such a discussion encompasses the potential for new models or data to improve on existing formulations, whilst also emphasizing the importance that evaluation criteria and the development of an open source infrastructure will have on the continued development of spatial interaction models in the literature. Alternative methods are also put forward in this chapter in light of the recent changes in consumer behaviour within the grocery retailing sector that could influence the appropriateness of the spatial interaction mode. This chapter also identifies that the complexity of behaviour modelled within the sector is likely to effect the model and scale chosen for any further analysis. This Chapter therefore addresses objective 7 whilst also touching on the discussion and analysis of objectives 1 and 3.

Finally, Chapter 9 looks to synthesize the findings of the thesis and draw some overall conclusions. The contribution of this thesis will be assessed relative to the aims and objectives laid out in this chapter, including to what extent the aims could have been said to be achieved and how they contribute to the overall objective of the thesis. While each objective will have been covered and discussed in each of the chapters, the extent to which they have been successfully achieved will vary, therefore leaving open questions that future research can explore. The chapter will therefore conclude by acknowledging the limitations of this research, how successful the thesis has been in achieving its aims and how the literature can move forward in light of the discussions presented here.

Chapter 2

Spatial Interaction Modelling History

2.1) Overview

This chapter begins the thesis addressing the first objective through a thorough review of the development and application of spatial interaction models in the academic literature. This review highlights how the spatial interaction model became the prominent retail location theory used in practice as opposed to others developed around the same time including the Principle of Minimum Differentiation, Central Place Theory and Bid Rent Theory. The review then continues by tracing the history of adaptation and implementation of the spatial interaction model, including how each iteration was developed in response to perceived issues of previous model formulations. The chapter is then concluded by proposing the use of the Wilsonian production constrained spatial interaction model in this thesis. This then lays the foundation for the examination of grocery retailing as explored in the rest of this thesis.

2.2) Retail Location Theories

From 1927-1933 the four main concepts that lie at the heart of our understanding of the location of retail activities were developed: Central Place Theory (Christaller 1933), Spatial interaction theory (Reilly 1929, 1931), Bid Rent Theory (Haig 1927) and the Principle of Minimum Differentiation (Hotelling 1929) (Brown, 1993; Clarkson, et al., 1996). Each of these theories are based on relatively simple assumptions with the aim of describing how the world should be, rather than necessarily how it is. This includes taking the standard economic assumptions that people are rational, utility maximising individuals and that economic activity takes place in a freely competitive, equilibrium seeking context (Brown, 1993). Naturally, this means that each theory operates in a simplified world by reducing much of the real world noise such as transaction costs or imperfect information, which would otherwise disrupt real estate markets from their expected equilibrium. Nevertheless, all four of these theories still contribute to, and influence, modern retail location theory and practice, regardless of having to be heavily adapted to fit within modern locational choice literature (Reigahinha, et al., 2017). This has involved considerable examination each of these theories since their inception, with different levels of support being found for one theory or another at different points in time (Brown, 1993). While this research focuses primarily on spatial interaction theory and its historical antecedents, given their considerable influence and amount of academic discourse on these different theories, it is worth initially exploring the three other locational models that exist and how spatial interaction models have come to dominate the literature in recent years.

2.2.1) The Principle of Minimum Differentiation

The principle of minimum differentiation derives from the work of Hotelling in 1929 who conceived of the idea of two profit maximising firms selling identical products and operating from fixed locations. The work begins with the premise that two sellers, A and B, locate along a linear plane, with length L, whereby a consumer would transport the goods bought at either A or B home at a cost, c, per unit distance. Initially, it is assumed that a unit of quantity of a commodity is consumed in each unit time at each unit along the length of the linear plane. This commodity is assumed to be identical from either seller, meaning that no consumer would have a preference for either other than which is cheaper by total cost, with demand being inelastic. The total cost for a conusmer would be made up of both the price of the good at the point of sale and the transport cost to travel to the store and back (distance x c). Hotelling suggested that the result of this would be that the point of indifference between the two sellers (the point at which a consumer would be indifferent between buying from either A or B) would be the point at which transport costs, and hence total cost, would be the same. Under this scenario, consumer surplus would be maximised if the two sellers located at positions 1/4L and 3/4L respectively. This is because, at these two locations along the linear plane, transport costs would be minimised within the overall system.

The solution present above however was suggested to not be stable. If one produced is initially fixed in their location, for example producer A fixed at location 1/4L, then the second producer, B, may move towards the center of the linear plane (1/2L) in order to capture a larger share of the total market. It would then be expected that in the next time period, the initially fixed producer, A, would retaliate to recapture their market share and even expand. This would be done by themselves moving towards the center of the linear plane but closer to 1/2L by distance than B had moved. It is then expected that this process will continue, with A and B moving continually closer to 1/2L, until both sellers end up located at the centre of the linear plane. Thus, while maximum consumer surplus would be achieved through reduced transportation costs at location ¼ and 3/4L, under competition both firms will likely gravitate towards the centre of the market to maximise their market share. The consequence of this will be the erosion of consumer surplus due to increased overall transport costs, while market share and/or profit for either company will remain the same. This theory therefore suggests that under perfect competition and perfect information two firms in a market would tend to cluster together.

Extending this theory then to multiple sellers, Hotelling (1929) suggest that the tendency to cluster would remain the same if the same assumptions were held. In reality we see evidence today of this clustering in shopping centres and city centres whereby multiple stores selling the same product often cluster on the same street, corner or a general area e.g. bookshops. In terms of empirical proof of this phenomenon, separating goods by their order with higher order goods being more expensive items being bought less frequently, while lower order goods being bought more frequently at lower price, there is evidence to suggest that the clustering relationship appears strongest for higher order goods such as electronics, rather than lower order goods such as groceries (Brown, 1993; Wang, et al., 2014). Thus, empirical examination of retail location, even recently, suggests that there is some support for this theory.

However, as the assumptions are relaxed, meaning that demand becomes relaxed and there is increasing product differentiation, the model begins to break down. This is because factors beyond purely transport costs, especially with advancements in transportation technology since the inception of the theory, could be suggested to influence consumer behaviour (Williams & Senior, 1977). An example of this is if we remove the assumption of undifferentiated products. Now a consumer may be willing to travel to a more distant seller to purchase a product if the gain in consumer surplus of the different product exceeds the extra cost of travelling to the more distance seller. The sweetness of cider may be one of those differentiating factor where some consumers prefer sweet cider while others would prefer dry, where consumers from either side of the market travel to the other to purchase their most preferred cider (regardless of this being a bad location strategy on behalf of either of the sellers) (Williams & Senior, 1977). The differentiation suggested here could be enabled by larger markets such as in cities where a greater population can support sub-categories of products due to increased demand, e.g. radio stations (Puga, 2010).

Thus, although there is considerable empirical support in terms of the clustering of similar firms, and this can often be experience anecdotally when exploring your local town, city or region, the assumptions supporting the theory are rarely encountered in real life. Furthermore, relaxation of the assumptions concerning market conditions and transportations tend to point towards results or conclusions that support different spatial arrangements of retail stores than those which are originally derived from the model, such as physically distant stores that are spread out across the city (Brown, 1993). Thus, suggesting that while the theory has empirical support, the forces creating the phenomenon suggest are likely to be different to those originally theorised. Hence, while there has been some work on this theory since its inception, including the adjustment of assumptions, it is often not invoked as an attempt to explain or evaluate the location of retailers in a modern city. Practically as well, the theory provides little guidance on where to locate a retail store other than to agglomerate with other similar stores in order to capture the largest market possible if your competitor cannot move. This is not practical when moving to a new location whereby the product has not been sold before or where there is considerable product differentiation already.

Nevertheless, there is considerable research examining the clustering or dispersion of retail outlets with the aim of providing greater understanding of the urban environment.

2.2.2) Central Place Theory

Central Place Theory comes from the original work of Christaller (1933) and Lösch (1940) in the 1930s and 1940s. The aim of this development was to be able to describe and enumerate the number, size and spacing of retail centres. To do this, several assumptions were made. The first was that in the area, region or country there would a uniform distribution of identical, equally affluent and fully informed utility maximising consumers (Brown, 1993). Furthermore, as in the principle of minimum of differentiation, travel was assumed to be equally priced per unit of distance, c, but instead of travelling through a linear plane, consumers could now travel across a flat plane from any direction, assuming away any geographical features. Within this space, it was assumed that consumers would simply patronise their nearest retail centre that sold the merchandise that they were seeking, and that single purpose shopping trips would take place to purchase that item. However, unlike with the principle of differentiation, instead of inelastic demand, with a single unit being purchased at each point, it was assumed that since transport costs increase with distance then demand for any particular product would decrease with distance. This would be such that, beyond a given distance from the retail centre then demand for that product would drop to 0. This would then lead to a circle of demand emanating from the point of sale. For the sale of the product to be viable then, there would have to be a minimum level of demand within that circle.

To complete this framework, it was envisaged that there were different types of goods that consumers could purchase. This included higher order goods, such as jewellery or electronics, that had higher value but were sold less frequently. These goods would require a high level of demand to be viable but individuals would be willing to travel far to purchase these goods. Consequently, higher order goods were seen to have a circle of demand with a large radius before demand reached 0. In contrast, lower order goods, such as groceries, were assumed to be those that had lower value but were sold more frequently, such as groceries. Consumers would be less willing to travel for these goods as they would higher order ones because of their lower value and higher frequency of purchase, but they would have a lower demand threshold required to be viable. As such, goods were seen to have a minimum catchment area, which is the area of demand needed to support the sale of the good, and a maximum distance a consumer would be willing to travel to purchase these goods. Combining these together, along with assuming perfect competition between retailers, it was suggested that this would result in a series of hexagonal markets. Hexagonal because firms would try to maximise sales and so overlapping circular catchment areas would lead to hexagonal patronage patterns, minimising transport costs for the consumers (McCann, 2013).

The hexagons would be different size according to the order of goods that the market would sell. Higher order goods would have a larger hexagon shaped market because of longer distance to travel and greater demand thresholds compared to lower order goods. This would lead to a hierarchy of shopping destinations overlaid over each other in a regular hexagon pattern (Brown, 1993). This can be seen to a degree in the UK with the hierarchy of shopping that we have in modern towns and cities. For example there are local shopping destinations such as the local high street, then there are city centre shopping destinations, followed by regional shopping destinations such as the Trafford Centre or Bluewater. With each increase in size of the shopping centre the order and expense of goods increase but also within each of the higher order centres you can find lower order goods as well, such as grocery retail stores in large shopping centres, suggesting that the hierarchy of stores overlaps.

Since its inception, this theory has been expanded and developed to reflect more realistic assumptions and to adapt to the modern retail environment. Regardless, several theoretical and practical issues have meant that, like the principle of minimum differentiation, there has been relatively little attention given to this theory more recently. This includes the issue of the assumption of single purpose shopping trips that only go to the nearest retail centre that sell these types of goods. There has been considerable empirical evidence, especially lately, that points to multi-purpose shopping trips and that consumers do not always patronise the nearest point of sale. This goes back to a problem identified with the minimum principle of differentiation, whereby product differentiation can lead to consumers willing to travel beyond the nearest store. The theory as well only focuses on the service sector of the economy which, while most modern developed economies are primarily service based, means that it fails to account for manufacturing and other activities in the economy. This includes failing to account for the fact that retail stores can locate near to these other activities to capture demand from workers, not necessarily where people live. Related to this is the implications of topology which also distorts the ability of consumers to travel, meaning that as the crow flies distance is not always representative of the true cost of distance (McCann, 2013). Finally, the model is based on a static environment where there is perfect information and retail markets, meaning that it fails to accommodate change which is divorced from the reality of today's modern retailing environment (Brown, 1993). Thus, while there is still interest in this model, and there has been recent research considering how applicable the model is to modern retail location, its application and conclusions are limited in the modern retailing environment. This is especially so given its limited practical implications for actual retail location decisions, other than suggesting that stores should locate close to other stores in the same goods order.

2.2.3) Bid Rent Theory

The third theory of Bid Rent Theory has held considerably more attention in the academic literature than the previous two. The foundations were originally developed by Haig in 1926 (Brown, 1993) which were subsequently developed by Alonso (1960, 1964), Muth (1969) and Mills (1969, 1970) throughout the 20th century (McCann, 2013). Originally, Haig (1926) maintained that centre of a plain of land is the most accessible point in a city, town or development and thus would be the lowest cost location in terms of transport costs. Consequently, in a perfectly competitive market, various land uses would bid for prime central sites, ranking themselves in terms of their ability and willingness to pay for the central land. This would lead, in the long run, to central sites being occupied by business that are able to pay the highest rents by putting the land to its “highest and best” use (Brown, 1993).

Such a theory could see in antecedents in the works of Von Thünen in 1827 who derived similar insights in relation to agricultural production and land usage. This was such that in an agricultural economy, production that had the least transport costs such as cattle, would locate further away from the centre, while goods that had high transport costs would locate nearer to the centre. This is because any value that was not spent on transport costs could be spent on rent instead (McCann, 2013). These insights were adapted in Haig’s (1926) to that of a city economy and further built on by the work of Alonso in 1960 and 1964. This led to Alonso’s land use model, which, based on the assumptions of a featureless plain, uniformly priced travel, free property markets, performing information and profit maximisation, bid rent curves could be derived for each land use in the city. The slope of these curves would be seen to reflect the sensitivity of that activity to changes in accessibility and substitutability of capital and labour with land (Brown, 1993). These individual bid rent curves for sectors such as offices, factories and residential land uses could be superimposed on each other relative to the centre of a city or region to create an order of precedence. Based on these orders of precedence it could then be seen where different land uses would be likely to locate including the rent and/or value of land (McCann, 2013). This suggested concentric circles of activities with business/office use in centre, following by housing and then agricultural/industrial land used, in theory reflecting the true make-up of the city.

This theory has been much critiqued in practice however, leading to a number of adaptations and changes since its first formulation. The first is that the original theory assumes a monocentric city dominant by a single centre. In reality, especially in larger cities, there can be many centres such as in London (McCann, 2013). Here, the assumption of accessibility being maximised in the centre of the city and declining equally in all directions can be distorted in reality by different transport options, segregation of land uses and the lack of a free property market (Brown, 1993). To an extent this critique has been overcome by more recent developments of the model, include several other implementations, such as the inclusion of multi-centric and non-mono-centric cities, social and institutional constraints, positive rent gradients, imperfectly informed consumers, agglomeration, variations in density and even a dynamic dimension to account for changes in land use, transport and business. This has meant that the main assumption of accessibility determining prices can still be seen in the model, even if the traditional smooth decline of value from the city centre is no longer as predominant as it once was (Brown, 1993). Consequently, this theory continues to be used in empirical examining of location of retail and its implications, showing its adaptability to account for the differences in conditions in modern retail markets. Practically as well it can be used to estimate the expected value of land in locations across a city if rent is already known. However, limits of an imperfect market and different conceptions of accessibility can make it difficult to apply this theory on a much wider scale. This is because, in order to make a realistic land valuation model, a complex model needs to be develop, integration data from a variety of data sources. Furthermore, the model is primarily focused on deriving a valuation of rent that should be paid, not necessarily how much demand there would be. Thus, while being able to suggest how much a retailer should pay, the theory is not able to say whether the retailer will make that back in the local demand. Thereby limiting its practicality.

2.2.4) Spatial Interaction Models

The final theory, that of spatial interaction modelling/gravity modelling, has received the most attention in the literature, especially in more recent applications. This is primarily due to its wide applicability, adaptability and usefulness in commercial applications and settings (Brown, 1993). The idea is based on that of Newtonian gravity in that the amount of revenue a store can derive is proportional to its size and the revenue available at an original, and inversely proportional to the distance between them. The main purpose of this theory then is to be able to determine, accurately, the expected level of demand that can be achieved at a certain point in a town, city or region for a given type of retail market. This makes it useful for retailers because accurate forecasts of demand can help with locational decisions, as if demand is greater than cost then it would profitable to locate a store in that area. If not, then there is no point in locating a store there. So how did the theory originate?

Some of the earliest works in this regard, of linking the idea of Newtonian gravity to that of the movement of individuals, was that of Ravenstein in 1885 and 1889. Here, Ravenstein used this idea in relation to internal migration decisions in England and Wales, later followed up by Young in 1924 how used similar ideas for internal migration in the Midwest USA (Carrothers, 1956). The first to link this idea to retail however, and explicitly invoke the concept of a trade-off that consumers make in terms of distance to access goods and services and the attractiveness that the particular destination may offer, was Reilly in 1929. In doing so Reilly stated that:

> "under normal conditions two cities draw retail trade from a smaller intermediate city or town in direct proportion to some power of the population of these two larger cities and in an inverse proportion to some power of the distance of each of the cities from the small intermediate city… typically, however, two cities draw trade from a smaller intermediate city or town approximately in direct proportion of the first power of the population of these two larger cities and in an inverse proportion to the square of the distance of each of the large cities from the smaller intermediate city" (Reilly, 1929, p. 16)

Reilly formulated this into an equation of the form:

$$\frac{B_a}{B_b} = \left(\frac{P_a}{P_b}\right)^N * \left(\frac{D_b}{D_a}\right)^n \quad \text{Eq. 1}$$

Where $B_{a/b}$ is the business drawn to a/b from an intermediate town T, $P_{a/b}$ is the population at a/b and $D_{a/b}$ distance from the intermediate town to city a/b. N is the exponent of population, relating to the drawing power of size of a retail centre which was assumed to be unit, while n is the exponent of the distance effect relating to the negative influence of distance on attraction. This was originally assumed to lie within a range of 1-3, not too far away from 2 as given in Newton's law of gravity, which Reilly suggested should be found through empirical validation.

Reilly argued that this idea and formula could be used to determine how much retail sales would come from an intermediate town to either of the two competing destinations, with the hope that it would be used by either retailers or city planners in determining where to locate retail outlets. This is what became known as Reilly's law (Haynes & Fotheringham, 1985) and was the outcome of empirical investigation using interview of shoppers of furniture and women's clothes in Texas, USA (Reilly, 1929). Reilly had the insight that this formulation could be used more generally with the potential to be applicable to a broad range of shopping scenarios including different categories of goods. This was only if the exponents were calibrated to the relevant data however.

While this theory could be suggested to be similar to that of Central Place Theory, in terms of determining where consumers would shop and that travel may be different for different orders of goods, a key difference is that while Central Place Theory assumes that individuals patronise the nearest store, gravity models assume that customers trade of accessibility and attractiveness (Brown, 1993). This can be seen in the quote and formulation above, where a larger retail centre further away could attract more business depending on the calculated exponent. This is the key contribution of this theory in terms of retail location (Joseph & Kuby, 2011) and remains a foundational element of all subsequent formulations and developments of this original theory.

Early examples of empirical validation of the theory and model are numerous and include explorations such as Bennet (1944), Douglas (1948, 49), Converse (1953), Strohkarek and Phelps (19488), Jung (1959), Reynolds (1952), Wagner (1974) and even Lösch (1954) (Brown, 1992). There were however other works that sought to discredit or disprove the theory as well. The main critique of the theory was the choice of variables as part of the formulation, being population and road distance, that were suggested to not be applicable to all cases, and that the assumed parameter values did not generally apply (Brown, 1993). This was despite Reilly suggesting in the original work that different parameters could be used for different applications, and that the parameter values should be calibrated according to each implementation of the model. Thus, these original critiques clearly failed to understand the ideas presented by Reilly and indeed it was subsequent attempts that failed to calibrate or adjust the parameters to reflect their application that found results that were not in line with what the theory suggest. These results were therefore not a result of an incorrect theory but rather the incorrect application of the theory.

It is this flexibility and adaptability of the model that has enabled the theories usefulness and applicability since its inception. This has included several different development and adaptations of the model to the given commercial environment that has allowed for its continued use and its increase in accuracy in predicting revenue of a retail store at a given location. Thus, give its commercial usefulness and significant literature supporting its application, it is to this model and theory that we turn to and thus develop a more in depth view of its history until its modern application and formulation.

2.3) Historical Spatial Interaction Model Development

2.3.1) The Break Point Adaptation

The first significant change and/or adaptation to Reilly's law of retail gravitation came from Converse in 1949. At this point Converse simply restated Reilly's law to be able to define the break point between two cities where a consumer would be indifferent between travelling between them. Following some mathematical manipulation, Converse derived the following formula:

$$D_b = \frac{D_{ab}}{1 + \sqrt{\frac{P_a}{P_b}}} \qquad \text{Eq. 2}$$

Where D_{ab} is the distance between city a and b, $P_{a/b}$ is the population of city a/b and D_b is the distance from city b at which consumers are indifferent between shopping at city a or b. To find the distance from city a at which consumers are indifferent, the population division can be reversed.

This resolved some of the initial critique of the model that it could only be used to find the ratio of spend from the intermediate town to either destination a or b. Now, this formula could be used to delimit a town's, retail centre's or products trading area. It could also be used without the use of surveys, in contrast to Reilly's original formulation, as it suggests where the boundary should be, not where the boundary actually is (Converse, 1949). Of course, this is assuming that there is a constant distance and population exponent which is not always the case. Nevertheless, the benefit of this adaptation was seen that it would allow retailers and newspapers to understand where they should target their advertising efforts to grow their customer base and it was used extensively to this effect to estimate trading areas of cities (Huff, 1963). Hence, we begin to see the first practical implications of the model to help retailers support and grow their business due to both its simplicity and relative ease of application as opposed to other models.

The applicability of this model extended beyond being able to used by individual retailers (which was an advancement in and of itself) but that it could also be used by planners to understand the effects of changes in the retailing environment. However, the model is a simplified version of reality as the distance exponent and population exponent in the above formulation were assumed to be 2 and 1 respectively. Furthermore, the model assumed that markets were permanently fixed, which in reality they weren't, especially during the period in which the model was first developed (Haynes & Fotheringham, 1985). Thus, while it was an advancement of the existing model in terms of its application, it suffered from some of the same issues that were highlighted in subsequent critiques of the original application of the model, indeed falling into the trap of assuming the fixed exponents that Reilly argued against.

2.3.2) The Integration of Utility

At this point, a critique that was oft repeated was the fact that the current model formulation was only able to deal with two populations at the same time and that the breakpoint was simply defined as a single line rather than a range of probabilities (Huff, 1963). Furthermore, with continued empirical use, it was argued that although practical, there was little theoretical justification to support the findings from the models, thereby suggesting that the application of the model should be limited (Huff, 1964). Thus, Huff (1963) sought to derive an improved trade area formulation by focusing on the consumer rather than the retailer, to try to explain how customer choose their shopping destination.

Huff thought that the utility of a shopping centre was determined by the number of items that the consumer desired are carried by that shopping centre and that travel time negatively affects utility. While a consumer may not know in advance whether a certain retail destination has the type of goods that they want, it is assumed that consumers would be willing to travel further the more items there may be available due to a greater likelihood of their desired product being stocked, all other things being equal. While the number of products being offered by a retailer may be difficult to determine, it was assumed that this could be proxied by floorspace, thereby taking on the same variables as the original model.

This meant that the two key tenets of the original model were retained (proportional to attractiveness and inversely proportional to distance), but derived the principles from utility theory, thereby placing the tenets on more sound economic theoretical footing. The resulting mathematical expression of this idea takes the form:

$$P_{ij} = \frac{\frac{S_j}{T_{ij}^{\lambda}}}{\sum_{j=1}^{n} \frac{S_j}{T_{ij}^{\lambda}}} \qquad \text{Eq. 3}$$

Where S_j is the square footage of a particular class of goods, T_{ij} is the travel time from origin i to destination j, P_{ij} is the probability of a consumer travelling from i to j, and λ is a parameter to be estimated empirically. Thus, the probability of a consumer patronising store j from origin i was suggested to be a function of the utility gained from visiting destination j, itself a function of size and travel time, divided by the utility gained from visiting all other potential destinations, which was also a function of size and travel time. This was a shift away from the deterministic viewpoint expressed by Reilly and Converse, and towards a more probabilistic one (Haynes & Fotheringham, 1985).

Huff therefore put the idea of gravity models in relation to human movement on more solid theoretical grounds as utility theory has continued to enjoy considerable longevity and popularity in general economic research (Joseph & Kuby, 2011). The way in which this was formulated as well allowed for the influence of multiple destinations whereby the probability of visiting a single destination depended on its relative utility compared to other destinations (Huff, 1964). It also allowed for trading areas to be gradually delineated by probabilities rather than fixed boundaries, thereby resolving two key critiques of the original model (Huff, 1964). This reflected the idea that a consumer may not exclusively patronise a single outlet for the same product, or range of products, as they may not be able to discriminate completely between on alternative or another due to imperfect information as to whether that particular outlet may fulfil their entire shopping needs (Huff, 1966). Furthermore, it shifted the focus away from population as a measure of mass, towards more shopping centre related relationships such as shopping centre size being a proxy for the amount of products (Haynes & Fotheringham, 1985). Thus, the model built on the existing foundations laid out by Reilly (1929) and Converse (1949), but was able to extend the modelling implications to address some of the critiques associated with the original model and make it more useful in its real world application.

The result of this model was that it allowed the modelling of trade areas of any given product or retail centre by generating a series of zonal probabilities radiating away from a shopping centre. This meant that individuals from any origin could have probabilities attached of going to any particular destination, allowing for multiple potential destinations, and hence for revenue sharing to multiple destinations from a single origin. On the basis of this therefore, Huff (1964) suggested that a trading area could be defined as:

> "A geographically delineated region, containing potential customers for whom there exists a probability greater than zero of their purchasing a given class of products or services offered for sale by a particular firm or by a particular agglomeration of firms" (Huff, 1964, p. 38)

This model could also be relatively easily extended to be able to estimate the amount of money travelling from origin i to destination j by including the population at i and estimates of spending on product k:

$$E(A_{ij})^k = \frac{\frac{S_j}{T_{ij}^{\lambda}}}{\sum_{j=1}^{n} \frac{S_j}{T_{ij}^{\lambda}}} * C_i * B_{ik} \qquad \text{Eq. 4}$$

Where C_i is the number of households in origin i, and B_{ik} is the estimated annual budget for consumers in the ith statistical unit for product class k, with $E(A_{ij})^k$ being the expected annual sales potential for shopping area j from origin i for product class k (Huff, 1964). Alternative alterations included defining the number of shopping visits, or the expected sales of a given item within a specific period of time. Critically however, it allowed for retail stores to be able to estimate their total expected revenue from a new site through summing up the inflows from all origins over a given period:

$$T_j^k = \sum_{i=1}^{n} P_{ij} * C_i * B_{ik} \qquad \text{Eq. 5}$$

Where T_j^k is the sum of all incoming revenue from each origins for product category k (Huff, 1964). This therefore was one of the major steps towards developing a model that could accurately estimate the revenue of a new store, or even of an existing store, into the future. Utilising this, when compared with costs and investment required, would therefore be able to suggest whether investment in a particular site was profitable or not.

2.3.3) The Physical Law Parallels

Around the same time as Converse and Huff's development of Reilly's original model, Stewart (1941-1950) and Zipf (1942-1947) were developing the foundations of the social science movement. This was an attempt to relate existing scientific theories to social phenomenon, one attempt including integrating gravity into human movement through reference to the Newtonian formulation (Carrothers, 1956). The aim of this movement was an attempt to decipher "universal laws" that were found in physical sciences that could also be applied to social sciences (Stewart, 1948). The most famous of these is Zipf's law which suggested a relationship between city size and city size rank, proposing that the size of the city is proportional to the largest city size of the country, while inversely proportional to its rank (Zipf, 1942). In terms of human movement however, Stewart (1941) originally examined student population in higher education facilities noting that "The number of undergraduates or alumni of a given college who reside in a given area is directly proportional to the total population of the area and inversely proportional to the distance from the college" (Stewart, 1941, p. 49). This was thus linked to the idea of gravity whereby the similar ideas of attraction and deterrence were seen to act at an aggregate scale for a population (Stewart, 1941). Further examination of this relationship by Stewart and Zipf continued throughout the 1940s, resulting in evidence in support of this theory coming from sources such as highway, railway and airway data for human migration in 1933-34 (Zipf, 1946), from newspaper circulation, length of items in newspapers and telephone and telegraph messages between cities (Zipf, 1946), railway express parcels (Zipf, 1946), and attendance at the New York World Fair, bus passengers and railroad tickets (Stewart, 1947). Thus, considerable empirical support in relation to the idea of gravity being able to be related to a social law.

Stewart then extended the physical analogy, beyond mere gravity, to include the concept of "potential of population" which was suggested to indicate the possibility of interaction between origins and destinations (Carrothers, 1956). This is indicted by the population at that point, divided by the distance away from the original point, expressed in units of people per kilometre or per mile (Stewart, 1947). As such, it was developed from the analogy of the potential of a gravitational or electric field, leading to the further suggestion that there is also the concept of "human gas". This was such that despite the idea of population potential, since everyone wants their own space, gravitational potential doesn't drag us all into a single central point (Stewart, 1947). He also suggested a modification to the population factor/exponent with the formulation, such that this could differ under different location circumstances. This was such that "molecular weights" could be added to population to account for the capacity of social interaction and hence would affect the density and potential of population (Stewart, 1950). These additions therefore hint at further attempts to integrate social concepts and social movement with physical laws and rules. It was at this point that Stewart and Zipf saw the development of social science as in the same early stages as early celestial mechanics in terms of the identification of rules and laws (Stewart, 1948). The aim was therefore to be able to describe social processes in an objective way when acting as a whole rather than as individuals (Stewart, 1950). While there is evidence of some of the earlier "laws" of social gravity and rank size rule however, there is limited examination and evidence presented in favour of these later developments and theory, mostly being left within the works that Stewart produced rather than going any further.

In extending these physical laws to social sciences however, while initial empirical evidence was in support of this translation, there was little theory to back up the reasons as to why this would be the case, echoing similar critiques that were seen for the work of Reilly (1929) and Converse (1949). To this end Zipf (1942) suggested that there are opposing forces of unification and diversity in resources such that an individual will want to minimise their effort of getting raw materials by locating near to them, but that if all the raw materials are not found in one location then people will also have to live elsewhere. In societies where communities produce what they consume and consumer what they produce, the system is minimising its total work, resulting in all members of the population getting an approximately equal share of the national income (Zipf, 1946). Under these conditions, any community P, will contribute to the total production of the system, C, an amount proportional to P/C, in doing so it will receive a reward from the system in the amount that is also proportional to P/C. When there are two populations that receive rewards proportionate to their populations then, the interchange of goods between these two societies would be proportionate to P_1P_2/C^2, if we ignore the factor of the easiest intervening transport distance. However, if the rank size rule depends on theoretically minimising the work of transportation, then the interchange of goods between communities will be inversely proportionate to the intervening easiest transport distance, D. Thus, the interchange in value between two communities would be suggested to be proportionate to P_1P_2/D (Zipf, 1946), with this relationship resting on the theory of work minimisation leading to goods transfer between populations. The result of this is that they derive a similar formulation as that of Reilly (1929), Converse (1949) and Huff (1964), but do so in a way that looks at the system as a whole, rather than individual choices, and more clearly derive their relationships from the physical concept of gravity and potential.

The different attempts to integrate gravity into retail location activities, and from different backgrounds and angles, highlights the perceived usefulness of the concept in explaining human movement and the requirement for a practical model to be able to explain this. The use of these models in practice, primarily in planning departments, eventually led to the development of the Entropy maximisation model by Wilson (1967, 1969, 1971). The aim of this was to put the model on a sounder theoretical and mathematical basis such that it could be used with confidence in practice, separating it from the physical analogy that up until this point had been heavily relied on and responding to the critique of limited theoretical underpinning (Haynes & Fotheringham, 1985). Of course, while Huff's model had already done this to some extent, they did so from the ground up whereas Wilson sought to do similar but from the position of top down implementation, similar to Stewart and Zipf. The basis for this therefore was noting that the formulation of existing models and their results were similar to more general models presented within the branch of mathematics known as statistical mechanics (Wilson, 2010). Thus, Wilson sought to derive the current model formulation from this basis, focusing particularly on the method of entropy maximisation.

2.3) Modern Model Formulations

Wilson (1967, 1969, 1971) saw the foundations of spatial interaction model as being able to generate estimates of the flow (T_{ij}) from origin, i, to destination, j, as a function of origin characteristics, O_i, destination characteristics, D_j, and the distance between them, d_{ij}. Thus, following on from the previous literature identified above, the simplest form of the spatial interaction model developed in analogy with Newton's law of gravity would take the form:

$$F_{ij} = \gamma \frac{m_i m_j}{d_{ij}^2} \quad \text{Eq. 6}$$

Where the gravitational force, F_{ij}, operates between two masses, m_i and m_j, that are separated by a distance d_{ij}, with γ being the gravitational constant. When translated into the notation suggested by Wilson to represent social flows, this would become:

$$T_{ij} = k\frac{O_i D_j}{d_{ij}^2} \qquad \text{Eq. 7}$$

as seen in the works of Stewart and Zipf in the social sciences movement and viewed from the top down.

The issue with this formulation however is that if both O_i and D_j were to double in population (or another appropriate variable) then the number of trips would be expected to quadruple. This is different to what we would expect in reality where the trips would only be expected to double because the formula above only represents the trips from origin i to destination j (Wilson, 1967). If the quadrupling were to occur, then at some given point the predicted flows from the model would exceed the total population that would be available at origin i, thereby leading to unsustainable predictions of flows.

Thus, in its current form above, the model would not be able to satisfy the natural constraints that would be applied to actual flows between origins and destinations. This is because the outflow from an origin cannot exceed the available population at the origin, and the inflow into a destination could not exceed its available capacity to support the population. Hence, the major flaw with the current model when trying to predict changes in flows in response to changes in population could result in logical inconsistencies. Wilson saw this and suggested that constraints could be placed on this model to ensure that the inconsistency would not appear. These constraints were such that the total outflow from an origin could not exceed the total population at the origin, and the total inflow into the destination could not exceed its capacity to support the population. These constraints could be represented by:

$$\sum_j T_{ij} = O_i \qquad \text{Eq. 8}$$

$$\sum_i T_{ij} = D_j \qquad \text{Eq. 9}$$

Equation Eq. 8 states that the total outflow from origin i to all destinations j should equal the actual total outflow from the origin i. Equation Eq. 9 states that the total inflow into destination j from all origins i, should be equal to the total inflow into destination j. These are input and output constraint which constrain the total input and output from the system. Wilson saw that these constraints are needed in the model and could be integrated through balancing factors A_i and B_j, as evidenced later in this thesis.

Furthermore, while the exponent of distance in equation Eq. 7 is taken as 2, as in the Newtonian formulation of gravity, there is no theoretical basis for assuming that this would be the case. Indeed, this is the argument presented in the original work of Reilly. Wilson thus suggested that a general function could be introduced to represent this relationship, along with d_{ij} being used to represent a general measure of impedance between i and j, rather than pure distance (Wilson, 1967).

Finally, we can also include a further constraint which can be satisfied in the model, which takes the form:

$$\sum_i \sum_j T_{ij} c_{ij} = C \qquad \text{Eq. 10}$$

This states that the total amount spent on trips in the modelled region, as predicted by the model, should be equal to the total amount of cost available (Wilson, 1967). In reality the total cost constraint does not have to be known but it can be used to ensure the other constraints hold in the model. These ideas of constraints thereby laid the foundation for the development of a gravity model formulation that could be relied of more consistent with logical conclusions form the top down.

Wilson couched these developments within the theory of entropy maximisation as derived from thermodynamics. This thereby continued the development or Zipf and Stewart whereby general scientific theories or principles would be applied to social phenomenon, particularly from the top down. Wilson (2010) described the basis for this as imagining a set of boxes for each origin-destination pair, where a single state of the system would be represented as an assignment of individual members of the population to these boxes. In one extreme, all individuals would be in one box and another would be there is an even spread across all boxes. Because all people are assumed to be identical, the mathematical permutations and combinations of these possible states tells us the number of possible number of possible combinations. In the case of the spatial interaction model with total population available T and T_{ij} being a given state of the matrix of flows from origin i to destinations j, this can be represented as:

$$W = T!/\prod_{ij} T_{ij}! \qquad \text{Eq. 11}$$

The number of combinations associated with a particular state, as opposed to all other states, can then be taken as the probability that a given state of the matrix, T_{ij}, could occur.

Borrowing once again from the field of Thermodynamics, Boltzmann showed that, subject to any constraints, one state out of all possible states is given as the most probable. This is found by maximising W in equation Eq. 11 subject to all other constraints. In the case of the spatial interaction model this is given as the constraints mentioned above which limits the total outflow from each origin and the total inflow into each destination. Given Stirling's approximation, this maximisation equation takes the form:

$$W = -\sum_{ij} T_{ij} log\ (T_{ij}) \qquad \text{Eq. 12}$$

This equation is recognised as a form of entropy, which can then be solved to extract the most likely model parameters.

After some mathematical manipulation on the part of Wilson, we can extract a formulation for T_{ij}, and a family of models which can all be used to model flows from origins to destinations. Firstly, the traditional gravity model of equationEq. 7 can be generalised to:

$$T_{ij} = KW_i^{(1)}W_j^{(2)}f(c_{ij}) \qquad \text{Eq. 13}$$

Where T_{ij} is given as a measure of interaction between zones i and j, W is a mass term, and c_{ij} is a measure of impedance between i and j, with K as a constant of proportionality (Wilson, 1971). The additional information contained within the constraints mentioned above can be integrated into this model through altering the constant of proportionality to model balancing factors.

The first model therefore is the unconstrained case. While called the unconstrained case, we do actually have a constraint in place, which is represented by the constant of proportionality seen in Eq. 13, and ensures that total flows within the predicted system is equal to the total flows in the actual system. Mathematically this is represented as:

$$\sum_i \sum_j T_{ij} = T \qquad \text{Eq. 14}$$

Where the constant of proportionality becomes:

$$k = \frac{1}{\sum_i \sum_j W_i^{\alpha} W_j^{\gamma} f(c_{ij})} \quad \text{Eq. 15}$$

This results in the total flows being constrained within the model. This can be used in the case where we know the total flows within the system, or we can at least estimate these total flows, but we do not know where the originate or where they end. Examples of this could include predicting migration in response to human or natural disasters where we can estimate the total amount of displacement, or tourism within a regional, national or international context where we may know the total number of tourists but not exactly where they start from or end up. This is therefore an information limited scenario where we do not have total known information.

The second model to be derived from this advancement is part of, what has become known as, singly constrained models. This is a production constrained model used in cases where the outflow from each origin is known, O_i, hence the total production from each origin. In this case, the constant of proportionality, k, is replaced by the balancing factor A_i, and the origin mass term, W_j, is replaced by the known outflow from the origin, O_i. This alters the model formulation to become:

$$T_{ij} = A_i O_i W_j^{\gamma} f(c_{ij}) \quad \text{Eq. 16}$$

Where:

$$A_i = \frac{1}{\sum_j W_j^{\gamma} f(c_{ij})} \quad \text{Eq. 17}$$

Here, the outflow from each origin constraint, Eq. 8, is held as the outflow from the origin is constrained to equal total outflow. These values could be based on existing known information, generated from equations or derived from a survey (Haynes & Fotheringham, 1985).

This formulation could be used when estimating retail sales as information from the census and other sources could be used to estimate the total outflow of people or revenue from each origin. For example, the census could be used to estimate the population available at origins and surveys could be used to estimate average household expenditure on specific good categories over a given time period (Wilson, 1971). Thus, the total outflow of money from an origin could be constrained within the predictions to the actual value under this model formulation. Estimates would then be derived based on a calibrated model for the parameters γ and β (Guy, 1991).

The other side of this model, and part of the single constrained model family, is that of the destination constrained model. This is used where the total inflows into each destination, D_j, is known. Here, similarly to the production constrained model, the constant of proportionality can be replaced by a balancing factor. In this case it is represented as B_j as the destination balancing factor that ensures all inflows into the destination predicted by the model reflect the actual inflows into the destination. Furthermore the destination mass term W_j is replaced by the value of inflow into each destination as D_j. This results in the destination constrained model of:

$$T_{ij} = B_j W_i^{\alpha} D_j f(c_{ij}) \qquad \text{Eq. 18}$$

Where:

$$B_j = \frac{1}{\sum_i W_i^{\alpha} f(c_{ij})} \qquad \text{Eq. 19}$$

This means that the total inflow constraint, represented by equationEq. 9, holds in this model. This model formulation can then be used hen we know the total flows into each destination or that they can be estimated from known information or equations. An example of this is being able to estimate housing demand in response to an increase, or decrease, in jobs in an area, such as the expansion of Heathrow. Thus, the total inflow into each destination within the modelled system would be constrained while the change in housing demand is explored.

The final model within this paradigm is the doubly constrained model where we have information on both outflow from origins and inflows into destinations. With this information we can ensure that both equationEq. 8 and equation Eq. 9 can be implemented. Here, the constant of proportionality can be replaced by two balancing factors, A_i and B_j, representing both the origin and destination balancing factors, and the two mass terms can be replaced by both the actual inflows and outflows observed. The model formulation for the doubly constrained model becomes:

$$T_{ij} = A_i O_i B_j D_j f(c_{ij}) \qquad \text{Eq. 20}$$

where:

$$A_i = \frac{1}{\sum_j B_j D_j f(c_{ij})} ; B_j = \frac{1}{\sum_i A_i O_i f(c_{ij})}$$ Eq. 21

The main purpose of this model then, is since inflows and outflows are constrained, is seeing how the flows may change in response to changes in the impedance factor. An example of this is when we know where people live and work and we want to understand how the population and trips can be distributed across the transport system. The inflow and outflows will be constrained while an examination can take place on the distance decay relationship within the model.

This of course requires a higher information threshold to be implemented and can often be the least used as a result. However, if information is known to allow this model to be implemented, then it is possible that all previous models can also be used, which will depend on the factors the user is trying to model or observe. Also, due to this having the greatest amount of information contained within the model, this is likely to be most accurate model in terms of being able to reflect actual flows. This often makes it a good benchmark to use when the information is available.

Wilson (1971) this described the entropy maximisation method as a model building tool in terms of allowing the family of spatial interaction models to be derived from the original formulation. The main advantages of this approach are that: the models generated are internally consistent and they can be extended, if the model can be derived in this way then it can be said the constraints give rise to the model, and there is significant history of entropy being used to study dynamic systems. Thus, the use of these entropy maximising models gives a mathematical underpinning to the utilisation of these models in practice and hence validity in their subsequent application going forward, resolving many of the issues highlighted by critiques of the original series of models.

These model formulations as a result are the ones that are most often applied in academic and real world applications and are seen as the foundation from which alterations or adaptions are built on to fit the given situation. While these developments by Wilson arguably take the top down approach of viewing the interaction problem, as seen in Stewart and Zipf, by modelling the spatial interaction system as whole, the results and model formulation are consistent with those seen in the utility maximising framework developed by Huff (1963, 1964) and subsequently generalised by Nakanishi and Cooper (1974). This is shown by work of Anas (1983) who showed their formulation equivalence along with comparing their results. Thus, for ease of use and understanding, it is from this point we take the application of these models to retail outlets and the prediction of store revenue, basing subsequent model development on that of Wilson and his entropy maximising models.

2.4) Model Adaptations

Before we continue however, since the work by Wilson is over 50 years old, there have been advancements and alterations to the model formulation and implementation that are worth discussing. The prime examples of this include the competing destinations model, the dynamic model, radiation model, geolagged model and the disaggregated form of the Wilson models. While the modelling in this these continues with the Wilson model, and utilises the disaggregated form of the model, it is worth discussing the insights and advancements suggested by these other formats of model and how they relate to the model utilised, indeed even why they are not utilised in this thesis.

2.4.1) Wilsonian Disaggregation

One of the main advancements on the Wilson style entropy models has been that of model disaggregated, originally identified by Wilson (1967) himself, which is subsequently used in the majority of model implementations when the data is available (Birkin, et al., 2017). Crucially, the models specified above treat all origins and destinations as equal in terms of the same parameters applied to all flows. This assumes therefore that each flow from origin to destination is assumed to take on the same behaviour, which for a variety of reasons is unlikely in real life scenarios. To this end, a disaggregated model was suggested to allow for different behaviours across groups with the same formulation. This is done by applying the same formula to subsets of the same dataset, expecting to see different parameter values representing different behaviour patterns.

An example of this type of thinking is when different income groups may find different brands more attractive. In terms of grocery retailing in the UK, higher income groups may be more attracted to a Waitrose than a Lidl because of the range of products they offer at certain price points, while lower income group may be more attracted to Lidl because of the product range within their budget (Newing, et al., 2015). In the formulation, this is expressed by different attractiveness parameters (γ) per consumer groups for different brands (Newing, et al., 2015). Furthermore, different income groups may have different mobilities, such as higher income consumers who are more likely to have a car, or even multiple cars, may have higher mobility and travel further, than those who depend on public transport. This can therefore be reflected in different distance decay parameters, β, for different income groups (Newing, et al., 2015).

In essence, the disaggregation of spatial interaction model allows for greater accuracy in model calibration and hence estimates. This is because the model is able to represent different behaviours, and therefore a wider range of possibilities, than a non-disaggregated model which assumes the same behaviour across all origins and destinations. In terms of a production constrained, retail sales model, this would allow for greater accuracy in estimates of potential sales from origins, i, to destinations, j, and hence total revenue estimates of the destination. Adapting the production constrained model to fit this scenario, this takes the form:

$$T_{ij}^{nb} = A_i O_i W_j^{\gamma^{nb}} exp\,(-\beta^n c_{ij}) \qquad \text{Eq. 22}$$

Where T_{ij}^{nb} is the flow from origin i to destination j, by consumer group n for brand b, α^{nb} is the attractiveness of brand b for consumer group n, and β^n is the distance decay parameter for consumer group n.

This model adaptation allows for more accurate model as long as data is available and the disaggregations make sense. Difficulties arise then when we have a subset of data, such as data from an individual retailer and/or when values are missing which must be estimated. In the former case, use of data from a single retailer can limit the accurate disaggregate potential due to the calibration of biased estimates of parameters that may not extend to other retailers, thereby limiting the accuracy of the overall model and estimates (Birkin, et al., 2017; Rains & Longley, 2021). In the latter case, where values are missing and must be estimated, estimation can be inherently difficult and bring in increased uncertainty or variability in the model, and lending values from nearby place may be limited as averages are not always applicable (Birkin, et al., 2010). Despite this, this model form has become standard practice in retail location modelling due to increased amounts of data, such as through loyalty cards and demographic classification metrics, which allow for disaggregated models to be accurately trained, with the limitations widely being accepted as not outweighing the benefits of disaggregation. Thus, this model formulation will be utilised in this thesis.

2.4.2) Competing Destinations Model

Beyond the disaggregation of the existing model, an advancement on the Wilson formulation has been the competing destination model (Haynes & Fotheringham, 1985). This development arose due to issues foreseen in the previous models in that estimates of the distance decay relationship were related to spatial structure, rather than reflecting the true underlying distance decay relationship (Fotheringham, 1983). Here, it was shown that for the use of origin or destination constrained model, the distance decay parameter was shown to reflect structure difference in the organisation of destinations or origins than in the behaviour from different origins or destinations.

Consequently, the competing destination model was developed whereby individuals were assumed to make a two stage decision process when selecting their origins or destinations. In terms of the origin constrained model, an individual was seen to choose a broad range of destinations to interact, then that individual would choose a specified destination from within that larger subset of destination. Thus, the destination selection process was seen to constitute the choice of a "microdestination" from which a "macrodestination". The principle suggested that was not taken account of in the original gravity model is that "*the more accessible a destination is to all other destinations in a spatial system, the less likely that that destination is a terminating point for interaction from any given origin, ceteris paribus*" (Fotheringham, 1983, p. 20).

On this basis it was suggested that accessible origins should show a steeper distance decay relationship than for inaccessible origins. This is because accessible origins would have more destinations to visit than inaccessible ones. Thus individuals in accessible origins won't have to travel as far as their counterparts in inaccessible origins, thereby will be unwilling to travel as far. In terms of the model formulation then, the accessibility of the origin should be taken into account, transforming the origin constrained model into:

$$T_{ij} = Z_i O_i W_j A_{ij}^{\delta_i} d_{ij}^{\beta_i} \qquad \text{Eq. 23}$$

Where:

$$Z_i = \frac{1}{\sum_{j=1}^{n} W_j A_{ij}^{\delta_i} d_{ij}^{\beta_i}} \qquad \text{Eq. 24}$$

And:

$$A_{ij} = \sum_{\substack{k=1 \\ (k \neq i, k \neq j)}}^{w} W_k d_{jk}^{\sigma_i} \qquad \text{Eq. 25}$$

In this formulation, Z_i acts as the origin constraint, and A_{ij} represents the accessibility of destination j to all other destinations available to origin i, as perceived by residence of origin i. Here σ_i measures the importance of distance in determining the perception of accessibility, where δ_i measures the strength of relationship between flows to destination and accessibility to other destinations.

This model was shown to be able to be derived from entropy maximising techniques in the same way that the Wilson models were, thereby putting them on the same footing as the Wilson models.

The theory for this was tested on 1970 airline passenger interaction data which found support for the predicted relationship and hence in favour of the adapted competing destination model. However, since this initial derivation and evaluation of the model, few papers have since attempted to implement the model due to difficulty in calibration. Indeed, at the current moment there is no generally accepted method of calibration of the competing destination models, with a grid search seen as not accurate enough compared to the calibration that can take place for the traditional model. Thus, despite its supposed increase in accuracy and theoretical basis, applications of this model have been limited.

2.4.3) Dynamic Behaviour

Another contribution and extension of the Wilson model formulation was that of Harris and Wilson (1978) who integrated dynamic behaviour into the model. They do so by drawing inspiration from the ecological frameworks for dynamics of population from the Lotka and Voltera model and the Boltzmann methodology to see how retailers would behave in response to changing competitive pressures (Wilson, 2007). The key contribution of this advancement was to be able to represent the changing dynamics of retailers which was previously a criticism of the existing model in that it could only examine static relationships (Wilson, 2010). The issue with this however was to find data that would allow the model to be fitted which meant that early tests were undertaken on synthetic data (Clarke & Birkin, 2018).

A recent application of this model includes that by Birkin and Heppenstall (2011) who do so for petrol stations in West Yorkshire. They do so by altering the mechanism of adjustment from price competition to floor space adjustment and integrate the dynamics of the model into an agent based model solution. They extend this further by accounting for discontinuous evolution, showing the effects that dynamic interaction may have on the opening, closing and expansion of petrol stations within this area. While the application of this model is novel, the fact that there have been few applications of this model suggest the difficulty in finding relevant data and accurately developing the system to apply it. Thus, application of this model is often limited in practice and cannot be used in relation to the proposed thesis.

2.4.4) The Radiation Model

More recent advancements have taken the form of developing alternative models such as the non-paramaterised model known as the radiation model, or the integration of geographically lagged variables into the geo-lagged version of the model. The first of these was developed by Simini et al,

(2012) in response to perceived issues with the Wilson model in that the distance decay relationship was picked on which fit best rather than theoretical concerns, the model requires significant data to be calibrated, and that there are regular discrepancies in model fit and calibration. As such, they developed a non-parametric model from first principles based on diffusion theory, with the aim of generating universal model that could be applied to all spatial interaction systems (Simini, et al., 2012). However, while their initial test show improved performance over gravity models, subsequent literature has found mixed results as to whether the gravity of the radiation model performs best (Masucci, et al., 2013; Lenormand, et al., 2016; Stefanouli & Polyzos, 2017). In fact, some of these result suggested that for the radiation model to perform as well as the traditional model, it had to be paramaterised anyway and so limited some of the benefits of the newer model. Hence, the traditional understanding of the model is still regularly applied.

2.4.5) Geolagged Models

Similarly, the geolagged models have been developed in response to further issues identified with the existing model in that traditional models assume independence of errors. In the results however the errors of origins or destinations physically close to each other are likely to be related. This is likely to be because of factors such as accessibility, clustering of stores, travel routes, similar characteristics of consumers or sharing information (Lee & Pace, 2005). Consequently, these models seek to integrate spatial dependence into existing models, within initial evidence suggesting that this results in improved model fit as measured by R^2 and log-likelihood value (Lee & Pace, 2005). While with this model there has been clear results that suggest improvements in model fit, the implementation of this model is often complicated and can require large amounts of data and computing power to make it work. Thus, limiting its application.

Despite these advancements in model formulation and application, several issues remain to be dealt with and reconciled within this modelling framework. One of these issues is the influence of multi-purpose or non-home based shopping trips where a significant amount of evidence suggests that the assumption of single-purpose home-based shopping within the gravity model is not well represented in reality, thus limiting the prediction power of these models and their conclusions for travel flows (Brown, 1992). Furthermore, with changing consumer preferences and lifestyles there is increased incidence of convenience shopping which departs from traditional gravity model assumptions in that this is often undertaken frequently at irregular intervals and for irregular items. This is reflected in the difficultly that current models have in being able to predict convenience store revenue where the traditional assumptions of the gravity model at the superstore level do not necessarily hold in the same way (Waddington, et al., 2018). These changing consumer preferences have also included the move towards e-commerce platforms in ordering across many retail industries which is expected to alter the distance decay and attraction relationship seen in existing models. Thus, while attempts to integrate e-commerce into these models have been made already (Beckers, et al., 2021), no satisfactory conclusion has yet been reached and thus this remains a frontier for existing models. Finally, these models also depend on data being available about competitors locations to be able to accurately model relations. However, with the current rate of expansion of stores, particularly in the grocery sector, perfect information isn't always available about the location of stores. However, if data is known about the changes in store revenue and where that change comes from then these models have the potential to be used to identify the location and size of competitors stores which has yet to be explored or utilised. Thus, there remains considerable scope for advancements of these models and for contributions to the existing literature.

2.6) Conclusion

Spatial interaction models are just one of four main retail location theories that have developed in the literature since the beginning of the 20th Century. Due to accuracy, adaptability, ease of implementation and practicality however they have become the most widely used of the four original theories and have indeed had considerably more ink spilled over them in more recent papers. While the origins of the theory come from the original insights of Reilly in 1929 in relation to furniture and clothing in Texas, the model has since adapted in a variety of ways over the period. This has included contributions that have led to the model becoming more mathematically robust and theoretically sound, such that they have been utilised by academics, retailers and governments alike. The main adaptation of this that has been used in such a way has been the developments of Wilson, who created a family of models that brought together the advancements of many previous researchers into a single framework thanks to the theory of entropy maximisations. This does not mean however that there are still not issues and or improvements that have continued to be suggested. Examples include those of the competing destination model, dynamic model, dissagregation, radiation model and the geolagged model just to name a few. However, the original Wilson model continues to be used due to its simplicity, ease of understanding, adaptability and wide applicability. Hence, its use in this thesis.

Chapter 3

Grocery Retail Location

3.1) Overview

This chapter aims to tackle the second objective of the thesis by examining the changes in the grocery retailing market in the UK over the last 60 years and to identify methods that have been used to determine new store location. This review highlights the progression of the grocery retailing market including the development of pressures and changes that are likely to influence retail store location and hence the relevance of the spatial interaction model to model current grocery store revenues. It also discusses the development of different grocery retail location methods, including how and when the spatial interaction models were adopted into practice. It concludes by suggesting that current consumer behaviour and retail channels offered are likely to negatively influence the effectiveness of spatial interaction models.

3.2) Introduction

The interest in this thesis is how spatial interaction models fit within grocery retail location decisions in the UK. To this end, before exploring the application of these models to a relevant dataset, it is worth exploring how the UK grocery retail market has changed over the last 60 years. This includes how changes in social and economic factors have resulted in the development and spread of different formats of stores and location decisions. Such a history encompasses the first development of the superstore in the 1960s, followed by the widespread adoption of the format through the 1990s. Since then, there has been a gradual shift back towards the convenience grocery format, along with the more recent rise in the importance and prevalence of e-commerce solutions in the industry.

Throughout this history, despite the prominence and development of spatial interaction models in the academic literature from the early 1970s, their practical use in relation to grocery retail location was only realised in the early 1990s. Even so, it was not until the late-2000s that spatial interaction models were used by a majority of retailers. Instead, throughout this period, a range of different solutions were used ranging from simply “gut feeling” of managers, to check listing and customer spotting, and finally to regression and spatial interaction modelling. Which, even though spatial interaction models remain the model of choice, a variety of techniques are used together in order to determine the final location of grocery retail stores in the UK.

Thus, in conjunction with changes in supply and demand factors in the sector resulting in changes in format and location, it is also worth exploring the history and application of locational choice models within the sector. The aim of this is to then understand how store formats and locations are chosen, and how spatial interaction models came to be embedded within standard practices in the industry. Such an exploration covers the changes in the superstore format and location, along with changes in the locational choice model, up until the modern grocery retailing environment we see today.

3.3) Grocery Retailing Store Location, Formats and Market Developments

3.3.1) The Development of the Supermarket

In the grocery retailing industry in the UK there has been several periods of change in both format and location of stores. This has been in response to changes in consumer habits, technology and legislation since the early 1960s. The first change to occur in the 1960s was the development of large store formats, now known as supermarkets and hypermarkets, which begin with experiments of a few retailers who hoped to increase their margins and profits (Guy, 1996). These retailers aimed to achieve lower costs and hence greater profits through the development of larger store formats. This was because the larger floor space would allow for fixed costs to be spread across a greater range of stock, thereby reducing average fixed cost per unit of stock. Furthermore, store efficiency could be increased by stocking a greater range of products with the same amount of staff dealing with greater floor space. While at the same time, with fewer large stores, as opposed to many smaller ones, a more integrated and simplistic distribution system could allow for lower costs on the supply side (Guy, 1988). Such changes were facilitated by changes in pricing legislation at that time that allowed grocery retailers to pass on lower costs to consumers through price reductions. This allowed for retailers to compete more on the price of goods rather than just the level of service offered, meaning that any reductions in cost could in theory make the retailer more competitive against other retailers (Guy, 1988).

Beyond operational changes allowed by the larger format store, the purchase of rental price per square foot of floorspace could be reduced by building these larger format stores either at the edge of town or even in out of town developments. Thus, not only was the development of a larger store focused on changing the way in which a store was managed, but also on the location of these stores. This change would rely on consumers being able to travel longer distances to reach these stores however, which, with rising car ownership across the country, was now becoming more feasible (Guy, 1988). The success of these early trials meant that retailers started to invest in even more stores of this format, following the large store out of town formula, due to increase profit margins of the early experimenters.

Consequently, in the early 1970s and 1980s, several retailers continued to adopt policies of significant store expansion centered around the development of larger store formats in suburban or edge of city sites. At this point, such sites were crucial to the continued development of these store formats because the reduced land prices around these areas, due to relatively little development, were key to ensure that there was enough land for parking space to be built (Guy, 1988). Consumers were certainly not expected to walk, cycle or take public transport when carrying their grocery shopping back home. This in itself was facilitated by a relaxed planning regime at that time, making out of town sites relatively easy to obtain and build on (Guy, 1995; Birkin, et al., 2017). Without such regulatory changes, firstly in terms of the changes in pricing strategies, and secondly in terms of planning availability, it may have been unlikely that the larger store formats would have developed in the way that they did.

3.3.2) National Expansion Plans and Effects

This was a major change in the grocery retailing industry in the UK at that time because most retailers were small, independent chains that were often geographically concentrated. This allowed them to focus on a specific market and demographic, often building up strong relationships with customers and their local community, with knowledge of the local area required for continued retailing success. However, the shift towards the larger store formats enabled retailers to consider plans for national expansion as a way of gaining national market share and increase their profit margins and overall profits (Guy, 1988). Indeed, a small area or region could only support a limited number of these larger format stores, as opposed to multiple convenience stores under the same brand.

This change, towards a national focus, was primarily driven by several major retailers. Between 1982 and 1990, the market share of the top five grocery retailers in the UK increased from under 25% in a rather dispersed, regionally focused market, to 61%, resulting in the development of an almost oligopolistic market and rapidly changing the competitive dynamics (Wrigley, 1994). This increased market share gave the retailers added advantages over and above those provided by the larger store format as they could not put pressure on suppliers to lower prices. This, combined with falling store costs per square foot in terms of both rent and resources, allowed for margins to increased from around 3-5% in the mid 1980s to around 6-8% in 1992 (Guy, 1995). Thus began one of the most profitable periods of grocery retailing in the UK, and even the world.

The benefits associated with this rapid expansion, in terms of increased margins, market share and profit meant that competition between retailers took on an entirely new dynamic as the stakes were now much higher. Where previously stores were competing primarily in their local region, such as in the South West or in the North West, with other regional retailers on the basis of quality of service, quality of product and on price, competition now played out on the national stage. With this change of arena came a change in factors that retailers competed over as well, with a clearer focus now on price and where stores would be located (Wrigley, 1994). Indeed, it could have been argued that store location now became more important than the majority of other competitive factors during this period.

With the larger format stores now thoroughly entrenched in how grocery retailing was undertaken, and this trend supposedly set for the foreseeable future, expansion of stores now accounted for a significant amount of any increase in annual sales seen by retailers (Wrigley, 1994). Thus, for every site that was made available to construct a new supermarket or hypermarket, there were several national retailers interested in opening a new store on that site. This was because of a desire of each major retailer to build tens of new stores a year, with overall levels of store creation exceeding a hundred stores a year, despite there being limits to how many stores the UK market could eventually support. This meant that retailers such as Asda, who had a traditional base in the North of the country, were now looking for space in the South of the country, and vice versa for companies such as Sainsburys (Guy, 1996). Despite these retailers still maintaining a clear regional bias to where they originally began, they were slowly shifting towards becoming national grocery retailers known across the country. The intense competition over this period, in terms of both price and location, has since been labeled the “store wars” which began to change the grocery retailing landscape across many towns and cities throughout the UK (Wrigley, 1994).

At this point, the effects of large store development by what were now national grocery retailers, began to spill over to independent local convenience outlets. There were even reports of cannibalization effects on retailers own smaller store formats as well (Guy, 1988). This included the closing of several independent grocery retailer outlets in place where large format retailers had opened up a new store, going hand in hand with the market share increase with the national retailers (Guy, 1996). While some of this could be attributed to the overall increase in grocery retailing competitiveness over this period, as lower prices seen in supermarkets put pressure on prices and margins overall, and consumer preferences changed to more convenient single trip shops with larger selections, the opening of superstores in these areas were seen as the driving force for these closures. This even spilled over into retailers own stores where cannibalization of sales led to the closing of some of their local convenience stores that operated at that time, or even in some cases the closing of some of their earlier supermarkets. However, the benefits of the new store were still seen to outweigh the costs having to close one of their own outlets, especially when the older store was poorly located originally due to poor access to population, non-relevant local demographics, or even a competitor retailer opening up locally (Guy, 1988). This was just a highlight of the significant changes that large stores were having on existing urban dynamics and locational choices.

3.3.3) The Rise of Competitive Pressures and Regulation Change

Large store development continued apace throughout this period from the 1970s to the late 1980s as ever more supermarkets and hypermarkets began to pop up across the country. However, during the period of the late 1980s and early 1990s, the number of new supermarkets and hypermarkets, even other grocery stores, started to slow down. This appeared to be because the race for any space appeared to shift to only the race for the best space (Birkin, et al., 2017). At this point, retailers were no longer looking for just any space to locate their new stores, but were now considering only highly accessible locations that were close to large population centres with disposable income (Guy, 1988). Spaces for large stores were becoming more scarce and the build up of competition, with most major retailers competing in every regional market, drastically increased land prices. Such as the extent of this competition that in some cases, the price of land for a new supermarket development site was on the order of £25 million, an order of magnitude greater than what it was just a few years ago (Guy, 1995). Thus the risk of a failed store was much higher than it had been in the past, which was combined with poor location planning by some retailers, lead to the closing of some stores that were new only a decade ago. This change also coincided with changes in locational choice methodology as more advanced, not necessarily more accurate, techniques began to be used by retailers to find the best sites. Notably, Tesco took advantage of these techniques early on, such as the spatial interaction model, and reaped the locational rewards (Wood & McCarthy, 2014). But this also enabled landlords and land sellers to take advantage of these methods as well to increase their prices based on their own sales estimates, pushing up prices further.

By the beginning of the 1990s, as the amount of new stores started to stagnate and decline, questions over saturation began to be asked. Some analysts during this period noted that there appeared to be much fewer sites available that could sustainably support a large superstore at the current profitability levels leading to questions as to where the sector was heading in the future (Alexander & Morlock, 1992). This was because, from the period of 1966 to 1996, over 900 grocery superstores had been built, along with many hundreds of smaller supermarket formats, and the increasing levels of penetration of discount superstores as well (Guy, 1996). Thus it was suggested that many of the low hanging fruit, in terms of location, had already been picked and competition was only going to increase.

At this time, while most analysts and manufacturers saw saturation of the grocery retail market being realized before 2000, many retailers remained more optimistic. Indeed, the majority saw market saturation being realized post 2000 or even later, with plenty of sites suggested to still be available that were profitable (Alexander & Morlock, 1992). This optimism was primarily seen in the behaviour and policies of Tesco, Sainsbury's and Safeway at the time, who were each still considering at least 25 new stores a year. This was despite Asda and Safeway, two major retailers in their own right, having to halt their store development programs due to financial considerations given a less than optimistic outlook for the future (Guy, 1995). Thus, there began a separation between certain grocery retailers and other stakeholders and analysts in the industry as to the expected outlook and expected growth.

The raising of these questions and concerns was also partly driven by pressure from the arrival of new deep discounters from the continent, such as Lidl, Aldi and Netto, that began to eat into the markets share of some of the larger stores by offering highly competitive prices (Birkin, et al., 2017). They were able to do this because of their innovative, at least to the UK market, formats whereby the cost of unpacking and selection was borne by the consumer rather than the store, along with cheaper non-branded or store branded products being offered as less money had to be spent on packing (Kor, 2019). Existing retailers had began to compete on quality at this time, with the aim of moving up the value chain and hence profit margins, allowing these deep discounters to enter the market and capture the lower end of the value chain. These stores also started to locate themselves in areas of major urban deprivation which larger retailers had written off as being previous unprofitable such as in the North of England or the Midlands (Birkin, et al., 2017). Two market leaders of Asda and Safeway were the most hit by these discounters as they had essentially tried to capture their target market both in terms of customer and where the stores would locate. Indeed, these two retailers began to come into financial trouble partly as a result of this new competition. This development was therefore seen as a challenge to the existing larger retailers which controlled a significant share of the market. A spotlight was now being shone on price competition, above purely locational factors, which during the recession of the early 1990s,. was on every consumers minds (Thompson, et al., 2012).

At the same time, there was also growing unease of the public about the market power that the major grocery retailers were accumulating. This is because suggested excess profits of these retailers (profit margins had continued to increase), which was suggested to come at the expense of the everyday consumer (Wrigley, 1994). Thus, there was general unrest about where the grocery retail market was heading, starting from the point of the development of the superstore format and the continued growth of a select few major retailers. This combination of factors therefore led to questions over the price that retailers were paying for sites, especially as to whether the valuations placed on land were sustainable given increased competition, potential saturation and public backlash (Wrigley, 1994). The major retailers had “hoarded” considerable holdings of land in “land banks” which if held at inflated prices, especially at the time of a broader economic recession and falling land prices, could lead to considerable financial issues for these retailers (Guy, 1995). This therefore put pressure on grocery retailers to either justify the prices they paid or there was going to market backlash in the form of stock market price reductions, counter to the growth trends they had experienced over the past few decades.

A result of this concern, which lead to even more questions over the price of land, was the beginning of tightening retail planning regulation. The basis for this change was to make the development of larger store formats in out of town areas more difficult, accompanied by broader restrictions on the development of other retail establishments as well (Wood, et al., 2006). This was in response to the idea that the development of out of town retail destinations, whether that was grocery or other retail, were having a negative impact on town centres. Thus, seeking to protect the future of the town centre in the UK. This began in the early- to mid-1990s with the release of the Planning Policy Guidance Note 6 (PPG6) in 1993 and the incorporation of the “sequential test” of new retail development in the revised 1996 version (Clarke, et al., 2012). This aimed to negatively affect large store development, particularly for grocery retail, as the new development test required developers to first consider space available in town centres, followed by the edge of town centre locations, district and local centre sites and then only as a last resort of out of town centres, where the majority of large store formats were being developed (Wood, et al., 2006). This sequential test would aim to promote the development of inner city land rather than green field location. Thus, the aim was to protect the future of town and city centre retail, stopping, or at the very least reducing, the continued increase in green-field out of town developments. This was therefore a signal by the UK government that they saw that the future of out of town retail shopping as negative implications on the future of the UK city. This was thought to have a negative impact on grocery retailing, adding to the already increasing pressure on the sector in the early 1990s, ending a period of rosy outlooks.

Despite the concerns, questions, fears and legislative consequences in the early and mid 1990s, grocery retail store development continued apace into the early 2000s. This was very much in contrast to the general trend of wider slowdown in retail property investment over the period, leading some commentators to even suggest that the “store wars” period had never stopped (Wood & McCarthy, 2014). This was because large retailers had been able to find new and innovative ways to continue their expansion. Such innovations included a shift in investment towards the expansion of existing stores, leading to increases in square foot owned by the retailers without necessarily constructing new stores, a shift towards the strategy of smaller store formats such as convenience stores, and the development of new locational techniques allowing a reassessment of previously marginal locations (Wood, et al., 2006). Thus, although the “store wars” may have never stopped, they didn’t continue in the same guise they did the early 1980s.

This shift, and comments, highlight the ingenuity of grocery retailers in the UK to be able to adapt to changing environments and to continue their growth. Many analysts suggested that the golden age of grocery retailing had come to an end, and it indeed it had in the form that it was already known for. But, growth continued apace. In particular, of the leading retail brands, Tesco was able to take advantage of the change in market conditions to adapt its retail offerings, while several other retailers such as Asda and Safeway were notably struggling, indeed even publicly stating that they were to scale back their expansion operations (Wood & McCarthy, 2014). This allowed some retailers to see this point as an opportunity to expand their market share, and continue to grow, when others were not able to. Thus, despite concerns over saturation, several retailers managed to continue with their expansion plans to some degree in a different guise. However, it is worth highlighting some of the impacts of this, as the concern over market saturation would never really go away anymore, certainly putting a dampener on the spirits of the retailers.

As mentioned previously, the development of the large format stores were seen to have negative effects on small independent convenience retailers across the country. However, even with the change in strategy, with new large stores seemingly taking a back seat, the negative effects on independent retailers were still seen (Elms, et al., 2010). The traditional unique selling point of local convenience retailers, often known as “corner shops”, were their convenient location, opening hours and instore efficiencies. These benefits however were being eroded by the development of larger store format which, while many not have been in the most convenient of locations, often were able to compete with better opening hours and efficiencies of scale (Baron, et al., 2001). As a result, independent retailers being unable to compete, many had closed down and others had been subsumed under the banner of umbrella organizations. By 2012, the market share of independents in the convenience market fell to 19.4% (Hood, et al., 2015). Thus highlighting some of the impact that the development of these stores have had on local towns and cities.

3.3.4) Changing Consumer Behaviour In The 2000s and 2010s

By 2011-2012, questions over saturation in the grocery retail market again began to raise their head. This occurred as there was a global recession occurring/recovery beginning after hitting most grocery retailers badly during this period, along with the continued expansion of deep discount retailers, the rise in internet shopping and the convenience culture of a new generation of shoppers

(Butler, 2012). Questions were being asked whether grocery retailers in their current guide were ready for the future of the market?

Such changes occurring including a shift in attitudes towards grocery shopping in the UK. Evidence from the early 2000s suggests that the main mode of shopping was still the main single grocery shop that was undertaken at the same time each week (East, et al., 1994; Popkowski Leszczyc, et al., 2004). By the early 2010s however, evidence began to suggest that changing lifestyles necessitated a different type of shopping that conformed less to regularity and more towards convenience (Buckley, et al., 2007; Hallsworth, et al., 2010). This change was characterised by more people shopping within a much smaller travel time and an increase in the proportion of shoppers who were shopping greater than three times a week (Elms, et al., 2010). Consequences were therefore expected as to the format and location of existing grocery retailers offers, including the maintenance of existing stores and the development of newer ones.

In response to this trend, and several other market pressures, since the 2010s several major shifts in grocery retailing can be seen. The first is the shift in development from large format stores to convenience formats by the major retailers once independent retailers had gone out of business. The second is the continued expansion of deep discount retailers resulting in an increase in competition, especially with price concerns becoming ever more prevalent. Finally, there has been a rise in importance and market share of e-grocery shopping, which although still relatively small to the total market share, has even led to the development of pure e-commerce pay and play retailers such as Ocado.

Firstly, focusing on smaller store formats, a convenience retail outlet can be classed as one that is smaller than 3,000 sqft, located close to consumers' homes or a large daytime population and have a wide but shallow product range (Baron, et al., 2001). Under the "small store" umbrella however we can also include high-street stores, that may in some cases be as large in 15,000 sqft, but also locate along major high streets, near to large concentrations of home or daytime population, but offer a slightly deeper selection of products than the convenience formats of retailers. These smaller store formats have thus increasingly become the vehicle for continued expansion by grocery retail outlets. They require a smaller threshold population to be profitable, can adapt to changing consumer demands for more convenience, and are viewed as a separate market from large grocery stores meaning that there are no concerns over market share size (Birkin, et al., 2017). National grocery retailers began to consider these retail formats as a more viable option for expansion, as opposed to the superstore and hypermarket format, as a response to the change in national planning regulations in the early 1990s (Wood & McCarthy, 2014; Hood, et al., 2015). This is to the extent that from 2003-2012 most of the increase in grocery stores nationwide was due to the opening of convenience formats by retailers such as Co-op, Tesco and Sainsbury's (Hood, et al., 2015). Consequently, in 2015, the grocery convenience market was reported to be worth an estimated 22% of the total grocery market (Hood, et al., 2015). Thus, as more retailers begin to invest in this format and the market share grows, competition is likely to intensify to and thus locational advantages are likely to be key (Birkin, et al., 2017). This is therefore seen as key growth area going forward, with fewer and fewer large format stores being opened.

The second key area for grocery retailers in the UK going forward is that of continued increased competition. While in the early 1960s and 70s it was retailers competing to expand nationally with large superstore sites, by the 1980s this was the same retailers competing for only the most profitable ones. Then, in the 1990s and 2000s, with the introduction of continental discounters, a new market player, and indeed even a new segment of the market, increased the level of competition. In the 2010s and beyond it remains that there is competition between retailers, discounters, convenience retailers and through e-commerce channels. This is especially so as even though there are saturation concerns, Aldi and Lidl, two of the main discounters and competitors to existing retailers plan to open over 100 stores each over the next four years (Nazir, 2021). This is likely to have further impacts on existing retailers revenues and their market shares. It thus become important for existing retailers to understand the effects of competition, to know where and when competitors are opening and figure out how to respond. For that, they need to have the correct tools to be able to estimate the full impact of these new stores going forward.

Finally, grocery e-commerce is likely to have a significant impact on the way in which individuals and families shop over the next several years. This will include through which channel we interact with grocery retailers along with where we interact with them. To this end, the adoption of grocery e-commerce has followed the general trend of broader e-commerce adoption over the least 20 years, with increasing utilisation and value relative to the overall market (Kirby-Hawkins, et al., 2019). This process has even accelerated over the least two years by the COVID-pandemic. Such growth has been facilitated by the benefits of online shopping as it reduces search costs of finding the right product, grants convenient access to product and price information, enables quick and easy comparison, no restrictions on shopping hours and no associated travel costs (Hamad & Schmitz, 2019). Furthermore, consumers have demanded greater convenience and choice, lower prices and greater store accessibility, which are all facilitated by the development of e-commerce offerings (Kirby-Hawkins, et al., 2019). Consequently, in the UK E-commerce is one of the fastest growing sectors of the retail economy (Birkin, et al., 2017), a trend which is also seen at the European, US and global levels as well (Beckers, et al., 2018). For grocery E-commerce, while uptake has been slower than the broader retail economy due to the nature of the goods sold (Van Droogenbroeck & Van Hove, 2017), it is still likely to have a significant effect on the geography of grocery retailing. For example, two competing theories of innovation diffusion and efficiency suggest that urban areas should utilise e-commerce first due to their propensity to adopt innovation or that rural areas should adapt it first as they face the highest travel costs respectively (Hood, et al., 2020). Furthermore, grocery retailers investing in e-commerce need to pay attention to delivery methods due to the mix of frozen, chilled or room temperature goods being sold and the frequency with which they are, with options such as click and collect, home delivery or locker usage all options which a retailer must consider (Vyt, et al., 2017). This is especially important as last mile delivery, estimated to be as much as 50% of supply chain costs in home delivery, shifts the entire burden of picking and delivery onto the retailer which has previously been borne by the consumer (Hood, et al., 2020). Nonetheless e-commerce is likely to continue to grow into the future and thus retailers must develop strategies to be able to compete in the new sector.

3.4) Store Location Methods in Grocery Retailing

As the formats, location and shape of the grocery retail market has changed, so have the methods by which retailers determine where to locate their stores. When experimentation with the superstore and hypermarket formats began, there were few methods that managers could reliably use to determine where to locate their new store other than “gut feel” and local knowledge. As the retailers grew however and they began looking beyond their immediate region to locate new stores it become much more difficult and costly to rely on gut feel alone, especially with the scale of investment in land and resources to open a new supermarket. This therefore led to the development and adoption of more “objective” techniques to evaluate a new store location, such as checklists, store comparisons and customer spotting. Nevertheless, these evaluation methods still remained subjective to a degree. It was not until further advancement in computer technology and mainstream GIS systems that these retailers could use methods such as buffer and overlay analysis, regression and even spatial interaction models in the late 1980s and early 1990s. Overtime these methods began to spread, with a majority of retailers now taking advantage of the increased objectivity of these models, not to mention their greater accuracy in predicting new or existing store revenue. Since the 2010s however, even more advanced expert models, along with new datasets, have begun to enter the toolkit of large retailers, including in some cases neural networks or random forest implementations of spatial interaction models. What this means is that there are a variety of techniques available for retailers to use in location planning, often being used in conjunction with each other through different stages of locational choice. Thus, it is worth exploring the evolution of these techniques to see what is the industry standard and how grocery retailers tend to make locational choices, and hence why spatial interaction models need to be evaluated in terms of their usefulness to these retailers.

3.4.1) Early Location Methods

The only technique that was really available for a long time, was that of gut feeling of a senior member of staff, often the manager or owner of the business (Clarke & Hayes, 2013). Here, managers or owners would visit potential sites that were up for sale, look around, and determine whether they had a "good feeling" about a particular sight. This is of course highly subjective and relied heavily on the decisions of an individual to decide store locations but it was highly prevalent before the development of GIS or other commercially available statistical methods (Clarke, 1998). Nevertheless, when the potential cost is relatively low (at least compared to todays standards or large superstores), the consequences of a wrong choice were that an individual may go out of business.

The use of this technique was despite the early development of academic theories such as Reilly's law of retail gravitation or central place theory. This is because there was relatively limited awareness of these developments, understanding of their conclusions or resources available to be able to implement them in practice. An individual corner store owner looking to expand to two shops is unlikely to have looked towards academia for any ways to locate their second store. They are more likely to have a good knowledge of the local area and community and thus have a rough idea of where would be a good place to locate a new store given the expected demand.

Prior to the 1970s, and the development of the superstore format, this technique was heavily relied on for location decisions, and it was even utilised by retailers throughout the 1970s, 80s and 90s. This is because the majority of retailers throughout this period still did not have access to more advanced analytical techniques due to resource constraints or the management team being unsupportive of locational planning departments/individuals (Clarke & Hayes, 2013). Even today, many small retailers who can't afford the investment in modern techniques can rely on gut feeling or knowing the local area if they own, or look to own, one or two small stores. It can even play a part in the decision making strategy of much larger firms when many potential options are presented after more complicated analysis has been completed.

While there was some success with these methods, especially in cases where managers had significant experience with retail location decisions and the new store was local, there were also some failures as well. This could occur in situations were the manager did not take account factors that were clear to see or the location decision was beyond their local area of expertise (Clarke, 1998). In some cases, there was not a clear understanding of what made an area conducive to food store performance, as luck can often play a part in decisions that are taken infrequently, or site visits were not long enough to get a full appreciation of the new site. A level of failure in this regard was accepted at a time of relatively limited competition. However, as competition intensified and the costs of poor location decision making increased, the influence of this method waned a the sole basis for locational decision making. Instead, it began to be used in conjunction with other methods and techniques, and may be the technique of last result when there is no single or well defined solution. Indeed, in cases such as convenience store location analysis, where more advanced techniques break down, it has been suggested that gut feeling and site visits remain a prominent selection method, even by some of the larger grocery retailers in the UK (Hood, et al., 2015).

3.4.2) The Search for More Objective Techniques

As the stakes for locating a store rose, more advanced and objective measures to evaluate where to locate a store were sought to be able to improve the site location analysis field, both generally and by individual retailers. One of the earlier techniques in this regard was that of check listing. This technique is used by simply generating a set of established criteria, such as the number of people in a certain demographic within a given travel time, or the distance to the nearest competitor, which is then used to evaluate the location for a new store (Robinson & Balulescu, 2018). While some of the criteria used in practice were often subjective in nature, indeed in some case there was no basis for selecting a given criteria, often the choices were based on past experience of the retailer or were accepted market standards (Robinson & Balulescu, 2018). For example, if an owner had a successful store in an area with a certain density, then when looking for a new store they may include that on their checklist for criteria to have, even if that wasn't necessarily the factor that made the store successful.

The benefit of this technique was that it set a base level of criteria for a store to be considered to be located in the area and thus it was hoped would reduce the incidence of poor store location on the basis of gut feeling alone. It was also seen as a quick way to dismiss potential store locations which became especially useful as retailers began to expand nationally and many sites had to be evaluated quickly (Robinson & Balulescu, 2018). The reason why this technique was not used earlier was because it was often difficult to identify and collect relevant sources of data to support this technique, at least on an objective level. While it would be easy to count the number of houses along a street, it was another thing to reliably estimate the number of population within a given walking distance. Thus, in reality, the final checklist criteria was often still highly subjective. Nevertheless, by the late 1980s, checklisting was seen as the most commonly adopted approach for grocery retail store location, suggested that considerable degree of faith was placed in this technique to make good store locational decisions (Simkin, 1990).

At the same time as check listing was being adopted, another technique was also being utilised, and is still used today, is that of “customer spotting” (Applebaum, 1966). This technique involved interviewing a representative sample of customers in a store to obtain their address, demographic information and their shopping habits. Locating these customers then on a map enabled a retailer to roughly delineate a stores trade area through a series of circles, or other shapes representing the underlying geography, to calculate the percentage of total customers coming from each given distance. Zones could then be created from this dataset to determine the primary (70%), secondary (20%) and tertiary (10%) trade areas by distance. This could then be used to roughly determine how much population would be needed within a given distance to support the store, to identify a potential target demographic to aim for or to see how competitors influence whether customers visit the store or not. This could then be used for store location analysis by finding site locations with similar characteristics to determine the potential viability of the new location (Applebaum, 1966; Clarke, 1998). While the interviews could be done without any extra data sources, and indeed the distance metrics would be easy to calculate, supplementing this to create a check list of criteria would suffer from the same issues as above, namely that access to data to support this method was difficult to find and evaluate. Nevertheless, this method could be used in conjunction with checklist analysis to identify potential locations for new stores.

3.4.3) The Integration of Information Technology

Broadly, these two techniques fall under the umbrella of what are called, analogue methods. Essentially what this means is that retailers would use this information to be search for similar locations to those of well performing stores where there wasn’t already a store. These methods were often relatively costly and time consuming to implement due to the resources required for data collection and implementation, especially when it came to mapping the results. However, these methods received a boost in the mid-1980s when technology advances and improved data dissemination methods enabled GIS to become more widespread in many retail organisations (Clarke & Hayes, 2013). This is because technologies advances allowed data from a variety of sources to be integrated into a single platform that could be used to more easily visualise the data and make decisions from, especially in relation to mapping. However, due to the cost of using these techniques, it was often only the larger retailers who could truly take advantage of these new technologies, through in house capabilities or the use of consultants, being the beginning of the gap in capabilities between national and independent grocery retailers.

These new technologies, combined with original techniques, led to the development of buffer and overlay analysis with GIS. Here, checklists, or customer spotting data, could be implemented within software relatively easily and quickly, or the demographic profiles of areas could be integrated into visualisations to understand where customers came from (Benoit & Clarke, 1997). The main benefit from this was the speed and ease of visualisation of the data which made the method much more accessible than ever before. This has allowed the two previous methods, in combination with buffer and overlay analysis, to be continued to used up until today because of their ease of implementation in automated systems and that their results are relatively easy and simple to understand for managers and decision makers (Roig-Tierno, et al., 2013). This has only become easier as technology has advanced, and indeed they remain the most commonly used methodologies in the initial analysis of retail store location today (Reynolds & Wood, 2010).

3.4.4) The Adaptation of Regression and Spatial Interaction Models

However, since then, these methods have tended to be used alongside more sophisticated methods, rather than on their own, as technology and the amount of available data has improved. This has meant that they are not the only tools at the location planners disposal anymore (Clarkson, et al., 1996). The technology advancement that enabled the development of buffer and overlay analysis has allowed for the utilisation of more complex methods as well, such as regression or spatial interaction modelling (Clarke, 1998). In the case of the former, the adoption of regression methodologies has allowed store to correlate geodemographic variables with a store's performance that then allowed new store performance to be predicted rather than just estimated (Clarke & Hayes, 2013). This therefore is able to act as more sophisticated and objective analogue method as it allows for direct comparison between existing stores and new sites on the basis of statistically significant influence factors (Birkin, et al., 2017). If the variables identified showed explanatory power then confidence could be put into the use of these techniques in evaluating new store performance.

However, there are several issues with using regression for evaluating a new store location. The first is the fact that these models are often unable to adequately handle the nature of spatial interaction flows. This is in relation to the geographical factors that influence customer flows, including the adverse effects of competition on performance, which can be difficult to adequately account for in a regression model (Birkin, et al., 2017). Furthermore, regression models are often unable to model the impact of sales on existing stores due to cannibalisation, thereby limiting their use in situations where new stores are located close to existing ones. Finally, due to the geographical nature of the relationship, several explanatory variables are often highly correlated, leading to incorrect or confounded conclusions from the results (Birkin, et al., 2017). Despite this, these models continue to be used, and are still used today, because of their perceived objectivity and more complicated analysis leading to actual estimates of potential store revenue. This is especially so for highly segmented markets such as clothing and restaurants as the negative effects associated with these models are minimised due to the way in which the market behaves (Mendes & Themido, 2004).

As to the second method, of spatial interaction models, competition for sites and consumer patronage because more fierce in the 80s and 90s, required the need to adopt more advanced and accurate models than there has been in the past (Birkin, et al., 2017). While experimental spatial interaction models had been in used as early the 1980s, it was often difficult to collect enough data to calibrate and validate the models through survey evidence alone (Guy, 1992). This was still despite considerable advances in data, computational developments and changes in the retail sector at that time (Clarke & Birkin, 2018). Furthermore, there was also limited software available to run these spatial interaction models, at least at the level required for accurate predictions (Benoit & Clarke, 1997). Thus, if retailers wanted to utilise these forms of model, they often had to be pay for expensive, custom made software implementations which many could not afford, leading to different extent of usage of these models and software between sectors of retailer and also between retailers themselves (Reynolds & Wood, 2010).

However, over this period from the early 1990s through to the beginning of the 2010s, spatial interaction model usage began to have a big effect on retail store location choices beyond a few early adopters (although the early adopted benefitted significantly from utilising these advanced techniques before anyone else) (Clarke & Birkin, 2018). This is because, as data, computers and broader technology continued to advance, it became clear that the application of spatial interaction models, at least when applied appropriately and correctly, produced significant value in terms of investment savings and profit benefits (Clarke & Birkin, 2018). For example, Tesco made early use of these methods in the mid-1990s, resulting in forecasts of their store turnover with 10% mean error, aiding them in their expansion plans and supporting them to become the dominant retailer that they are today (Mendes & Themido, 2004). The utilisation of spatial interaction models in a commercial retail setting thus began to be used more widely. However, even by the 2010s they were still not used by a 1/3rd of stores (Reynolds & Wood, 2010). This is primarily because of their often academic nature, leading some to use less sophisticated models, and their high data requirements to be able to be calibrated accurately. Thus, while these may be some of the most accurate models available, not all retailers are able to utilise their benefits, this is especially so for many of the smaller retailers where resources may be limited.

3.4.5) The Fourth Wave of Retail Location Methods

Despite spatial interaction models not being fully utilised across the entire sector, in some cases retailers are exploring other, even more advanced analytical techniques, in their location planning departments. This includes methods such as artificial neural networks, random forest models, or other data science techniques that are filling gaps where spatial interaction models not be working as expected or where large datasets enables their usage alongside traditional models (Robinson & Balulescu, 2018). Some commentators have even suggested we are entering a third, or fourth, age of retail location planning due to continued advances in technology and data (Clarke & Hayes, 2013). The split is often the first age of "gut feeling", followed by analogue method usage supplemented by the beginnings of technologies advances, then the adoption of analytical techniques such as regression and spatial interaction models, followed by the fourth age of data science methods. This fourth stage thus include the integration of new datasets into existing models such as mobile phone location data or social mediate data that allows us to extend the models beyond the analysis of a single retailer (Clarke & Birkin, 2018).

In reality, the use of the "fourth age" techniques is often limited by data, experience and computing power available, with only a small minority of firms attempting to utilise these techniques for locational analysis (Reynolds & Wood, 2010), despite the percentage of firms using this continuing to grow (Hernandez & Bennison, 2000; Reynolds & Wood, 2010). If this is indeed the case, it suggests that these methods are potentially seen as the way forward, beyond the spatial interaction model, and their use is likely to continue to grow as their understanding and ease of implementation improves. Of course, there also needs to be continued improvements in computing power and data collected by the retailers, but this has been an ever expanding process. Adoption of these methods would likely mirror the use and implementation of gravity model formulations, whereby in the early 1990s a minority of firms were able to use this technique, but since the turn of the decade a majority were able to do. The question is whether this would actually be the case or not as many of these new techniques are as of yet unproven in practice.

In grocery retailers, the current grocery location planning workflow utilises most of these techniques in conjunction with each other to make the final decision. It is noted that the simpler techniques, such as buffer and overlay analysis, are used to relatively easy filter through potential sites due to their ease of implementation and quick understanding, while more complicated techniques could be used on the narrowed down list due to their complexity and increased resource use required (Clarkson, et al., 1996). In a department analysing tens of potential locations a week, only a small amount of those can be critically analysed to the full extent of all the methods available. Furthermore, using multiple locational models in conjunction with one another is likely to reduce the risk of a single erroneous result affecting any judgement, thus more confidence can be placed in the final conclusion (Clarkson, et al., 1996).

This means that, with access to more data and improved technology and software, there is still a role to play for more traditional techniques. As such, despite advancements in adoption and usage, there is still a role to play for local knowledge in an understanding of local retail markets. On the ground knowledge can support advanced analytical techniques to ensure that the best locational choice is made (Hallsworth & Jayne, 2000). This is the reason why many local real estate advisors still have their job, as local knowledge is still useful! The strength of these techniques in informing the final decision can also be influenced by a variety of factors within the organisation. For example, an experienced location planning team, with patient and understanding managers, that are well respected within the firm are more likely to be influential in deciding final locations than those in firms who do not support the location planning team or whose managers do not understand the work that they do (Hernandez & Bennison, 2000; Reynolds & Wood, 2010). Thus, it is not only how accurate these models that affect their usage, but also broader institutional factors within the sector (Wood & Reynolds, 2011).

In this sense, our interest remains primarily with the use and implementation of spatial interaction models. This is because they are often used as the final methodology for store site evaluation once more simplistic techniques have ruled out several sites (Reynolds & Wood, 2010), they can accurately forecast store revenue, and they have received considerable attention in the academic literature. Their implementation however is often of the more basic spatial interaction model formulations because despite continued advances in the academic literature they are often difficult to implement in practice due to the required expertise and data (Hernandez & Bennison, 2000), being reflected in the low (although increasing) utilisation rates of actual retailers (Reynolds & Wood, 2010). This focus on spatial interaction model is despite more advanced models such as neural networks and random forests being identified in the literature as if spatial interaction models struggle to be implemented in practice then these methods are even less likely due to their more complicated nature (Hernandez & Bennison, 2000). Thus the focus is on the current implementation of spatial interaction models, how they can be improved and where they can be utilised in practice.

3.5) Conclusions

The UK grocery retail market in the UK has evolved since the first introduction of the supermarket in the early 1960s in response to changes in demand, supply and regulation. This has included the successful first few experiments of the supermarket format in response to changes in pricing and land regulation allowing retailers to increase their profit margins. Following these examples, several retailers adapted the concept on a much wider scale, expanding from regional competition in the convenience store market to national competition through the supermarket format. Such was the dominance of a few firms in this regard that the top five grocery retailers in the UK expanded their market share from 25% to over 60%, locating stores all across the country. This continued expansion led to concerns however, with the government introducing regulation to favour brownfield development, concerns over too much market power for the key retailers and increases in competition from intercontinental discounters. The main retailers thus had to adapt by expanding their existing stores, developing a convenience store format and begun forays in the world of grocery e-commerce. This allowed them to continue to expand their offerings at a time of general economic difficulties in the UK until the beginning of the 2010s. Challenges that they face today in that sense are generally different to what they have faced before, with continued competitive pressures from expanding discounters, changes in consumer preferences towards convenience retailing and the growing importance of e-commerce. This has meant that the models they choose to put their faith into for where to locate their stores are more important than before.

Throughout this period of change and upheaval, the techniques that grocery retailers have used to determine where to locate their store has also changed. While competition was mostly local with small format stores, managers or owners using their local knowledge could use their "gut feeling" to be able pick new successful store locations. However, with increasing competition and costs of constructing a new store, more advanced and objective techniques were sort. In the early 1970s and 80s, before computing was widespread, the industry looked to techniques such as check listing and "customer spotting" to determine where to locate a store. These methods remained subjective in their use and often struggled with a lack of data, but their use made sure that it was no longer down to just the manager or owner to make these decisions. Advances in technology in the mid 1980s enabled these techniques further by allowing for advances in data integration and visualisation. Such was their simplicity and insight that in fact they are still used today through what is known as overlay and buffer analysis. Nevertheless, new technology also enabled the use of more advanced and objective techniques such as regression and spatial interaction modelling. With the costs of developing a new store increase ever higher, it became crucial for large retailers to adopt these new techniques, with those that did early on reaping rich rewards. Nevertheless, with continued advanced in technology and data availability, some retailers have gone even further, utilising large datasets and data science techniques such as Neural Networks or Random Forest algorithms with loyalty card or GPS data to model store location. While a variety of techniques are likely to be used in practice, it remains important for these firms to know how to accurately model store revenue, to which extent it is useful to evaluate the performance of industry standard spatial interaction models in relation to challenges that the industry is currently facing.

Chapter 4

City Model Application

4.1) Overview

This chapter builds on the work of the previous two chapters by highlighting unresolved issues in the literature in regards to spatial interaction modelling implementation, tackling objective three. To this end the chapter begins by discussing different calibration methods in the literature, identifying that a Poisson Regression formulation would be the most appropriate method to use in the current thesis. Furthermore, methods of model evaluation are also discussed in relation to metrics that have been used in previous research and how they can be used to validate the modelling performance achieved here. These discussions, along with those in the previous chapters, are then used to inform the implementation of a spatial interaction model applied at a city level scale.

4.2) Introduction

The foundation laid out by Wilson and his family of gravity models has been the building block of the majority of retail location analysis since his original contribution. This has included both in the academic literature, where the majority of contributions continue to build on this pivotal work, and in practice where spatial interaction models are one of the most used models. This does not mean however that implementing them on datasets is without difficulty. Indeed, several issues are highlighted in the literature that often act as a barrier to the actual implementation, evaluation and usage of these models. For our purposes, this primarily concerns how to calibrate these models, which is no mean feat, and how to evaluate the models. The debates are fundamental to understanding how to apply these models in practice and how they influence the model formulation. Thus, it is worth examining these issues in turn, before applying these models to a small subset of the data that we have available. The aim of this is to then inform the future direction of this research and to highlight potential issues that may affect the modelling within such a data rich environment.

4.3) Spatial Interaction Modelling Calibration

4.3.1) The Wilsonian Family of Models

Since the first development of these models significant attention has been paid to how these models can be calibrated. This is important because if the models are to be utilised in practice, we need to know that that parameter values are those that accurately reflect the underlying data and phenomenon that we are studying (Batty & Mackie, 1972). As with the original critiques of Reilly's law showed, a model that isn't accurately calibrated to the situation which is wishes to be applied cannot be relied on to draw reliable conclusions. This would limit its usefulness and indeed its practicality if they could not be accurately calibrated.

Introducing the forms of the models again we have:

The unconstrained model

$$T_{ij} = KW_i^{\alpha}W_j^{\gamma}f(c_{ij}) \quad \text{Eq. 26}$$

The production constrained model

$$T_{ij} = A_iO_iW_j^{\gamma}f(c_{ij}) \quad \text{Eq. 27}$$

The attraction constrained model

$$T_{ij} = B_jW_i^{\alpha}D_jf(c_{ij}) \quad \text{Eq. 28}$$

The doubly constrained model

$$T_{ij} = A_iO_iB_jD_jf(c_{ij}) \quad \text{Eq. 29}$$

For the unconstrained model, calibration includes estimated of the mass parameters, α and γ, which measures the relationship between flows and the "mass" of the origins and destinations respectively, and β, which measures the strength of the distance decay relationship in the flows. In the singly constrained models (the production and attraction constrained model) this includes estimation of α or γ alongside β, and in the doubly constrained model we are only interested in the calibration of the β parameter. Thus, in each model as we go from the unconstrained to the singly constrained to the doubly constrained, more information is contained within the data itself meaning less work has to be done on the calibration end, but calibration still needs to be undertaken to ensure a good model fit.

While early models, such as Reilly's model, would have used simple search techniques for the calibration of a single parameter to best fit the data, we now have considerably more data and more parameters to calibrate. This means that we need to ensure that the models are calibrate accurately and effectively as otherwise the models will be rendered useless. The importance ascribed to this within the spatial interaction domain can be seen in the considerable amount of literature to finding the best technique for calibrating these models.

4.3.2) Linear Regression Calibration

One of the first methods of calibration for the family of these models suggested taking logs of either side of the equation in question to linear model. In terms of the unconstrained model, this would transform the equation to:

$$ln\, T_{ij} = ln\, k + \alpha\, ln\, W_i + \gamma\, ln\, W_j - exp(\beta d_{ij}) \tag{Eq. 30}$$

Transforming the equation in this way would have allowed the parameters to be estimated through standard ordinary least squares (OLS) regression, a well-known and easy to use technique at that time (Fotheringham & O'Kelly, 1989). Such formulation could also have been shown to be derived through the equivalence of the multiplicative competitive interaction (MCI) model and entropy maximisation model, as shown by Anas (1983), Nakanishi and Cooper (1974), as regression was heavily used in the early implementations of the MCI model as well.

This method of calibration has been applied to a variety of empirical examinations since then. For example, in order to calibrate a production constrained SIM to the East Midlands Economic Planning Region, OLS was used (Gibson & Pullen, 1972). Using this method of calibration meant that they were simple to apply and additional explanatory variables could be easily added to the model, due to the nature of regression. This was useful in economic applications of the models, such as in trade analysis or knowledge transfer, where adding additional variables was seen to add flexibility to the model specification and account for factors other than traditional variables used in spatial interaction modelling (Burger, et al., 2009). As a result, linear regression was one of the most widely used techniques for calibration of these models (Ewing, 1974), and remains one of the main reasons why it is still used today.

Despite its early and continued popularity, there are several issues in the way in which this method is used to calibrate spatial interaction models. Firstly, if there is considerable variation in the individual flows in the data, log linearising the model and then using OLS regression will give undue weight to small flow observations in the untransformed scale. This means that the estimated parameters in the model, when translating the results back to the original scale, will likely be skewed, as opposed to if the parameters were estimated directly from the data (Cesario, 1974). Secondly, OLS makes the assumption that the variance of the dependent variable remains constant regardless of the size (normality of errors assumption). In terms of flows however there is likely to be considerable variation with the variance increasing with the size of the flows, skewing both the results and the parameters from the OLS model (Cesario, 1975). This is related to the fact that log linearising the original model assumes that errors are log-normally distributed around the estimate, where in reality, due to the nature of flows, there is little reason to assume this is the case (Flowerdew & Aitkin, 1982). This often means that the OLS assumption is often broken and thus can lead to inaccurate and skewed results in terms of the parameter values.

When it comes to how the data is fed in to the model and values are predicted, the model is trained on log-normalised values, Jensen's inequality implies that estimates of log T_{ij} are not the same as logged values of T_{ij}, meaning that the estimates from the model when transformed back to the original scale are unlikely to be accurate (Santos Silva & Tenreyro, 2006). In this particular case, this will mean that there will be consistent underprediction of small flows and overprediction of large flows, therefore failing to reflect the true variation in flows represented in reality (Flowerdew & Aitkin, 1982). Secondly, the log-linearised model cannot deal with zero flows, as you can't take a log of 0, meaning that an small offset value is used. When there are few of these values, then this does not make too much of a difference to the model, however when the number of zero values is large, then this can have a significant effect on the coefficients of the model which can lead to incorrect conclusions (Flowerdew & Aitkin, 1982). Finally, when it comes to transforming the data back to the original scale, the results from the least squares model are unable to ensure that the constraints hold (Cesario, 1975). While Fotheringham and O'kelly (1989) suggested that in the unconstrained case k could be adjusted after model calibration to ensure that the total constraints fit, in singly and doubly constrained models it is likely that, even after adjusting the balancing factors, some of the origin and destination constraints are likely to be consistently under or overpredicted (Cesario, 1974). This therefore limits the use of the constrained forms of the model as if the constraints can't be enforced then the extra information is relatively redundant in the model implementation. Thus, OLS calibration of spatial interaction of the models is likely to lead to inaccurate results and inappropriate conclusions, arguing that OLS should not be used to calibrate these models in practice.

A final limitation of the log-linear model method of calibration includes the fact that the models are inherently non-linear, especially doubly constrained ones. This means that estimation of the parameter values using this technique is unlikely to result in estimation of the true parameter values (Wilson, 1971). Batty and Mackie (1972) explore this in some depth, stating that on this basis linear estimation method is not appropriate for spatial interaction models. They state that this is because of the difference between non-linear equations which can be linearised through transformation, thus allowing them to estimated via methods such as OLD, and those that can't. They show that while the unconstrained form of the model could be argued to fall into the former camp, the constrained models cannot. This is because of their balancing factors which if estimated by OLS would result in incorrect estimates of the parameter values.

4.3.3) Maximum Likelihood Estimation

After showing that OLS cannot be used, Batty and Mackie (1972) use the earlier work of Hyman (1969) and Evans (1971) to show that the models could be calibrated through existing maximum likelihood estimation methods and that this would ensure correct parameter estimates. For this, they convert the production constrained retail model to a probabilistic format and highlight several methods for estimating the parameters including: first-order iteration, Newton Raphson method, Fibonacci sequences, simplex search and quadratic search. Using these methods of calibration ensures that the parameters are estimated directly, rather than through their log-transformed values, which removes the bias created by focusing on small flows associated with log-linear calibration methods (Cesario, 1974). Furthermore, while least squares estimations through the log-linear format only holds constraints approximately, maximum likelihood estimation ensures that the constraints are met in their original form (Cesario, 1975). Thus, an alternative to the OLS method of calibration was presented by Batty and Mackie (1972), which resolved several of the issues highlight with the OLS method.

In presenting this alternative however, Batty and Mackie (1972) highlight the fact that specifying and applying these algorithms, especially after developing a probabilistic form of the model, can often be a complicated and lengthy process. This means that in reality, while they resolve the issues of OLS, they are often difficult and apply in practice, thereby acting in contrast to a key benefit of OLS, that it is easy to use. Utilising such a method would often mean that specialist knowledge would be required which is often not available in practice, especially in commercial domain, where time and resources are often constrained. Thus, while such methods can reach the correct value for parameter estimates, and are likely to more accurate models, their use in practice has been hampered by their complexity and limited applicability. We can see this highlighted in the existing literature where very few spatial interaction model implementations have attempted to use any of the formulations presented by Batty and Mackie, despite computer programs openly available that can implement them. It remains difficult to overcome the initial implementation step of implementing the model in a probabilistic format. Thus, other methods of calibration have been sought in order to overcome the issues of OLS and maximum likelihood calibration.

4.3.4) Poisson Regression

Following this period, there was a lull before any viable alternative was suggested. However in 1982, Flowerdew and Aitkin suggested that Poisson regression could be used to calibrate spatial interaction models. Notably, in response to the issues of log linearising the model, as highlighted above, they suggested that Poisson regression would be more appropriate for several reasons. Firstly, if the flows represented people, then these could only be positive integer values, which suggests that a discrete probability distribution should be used, like that of Poisson distribution. Secondly, Poisson regression can deal with zero values without having to add an offset value, thereby reducing the need for an extra step in model preparation and ensuring that there is bias or skew as a result. Finally, Poisson regression assumes that the variance is proportional to the mean, which is a natural assumption of the flow distribution, as the value of individuals increases then the variance is also likely to increase (Flowerdew & Aitkin, 1982). Beyond these benefits, Poisson regression was also seen to have a relatively easy implementation at the time as it was a well known distribution that has been around for a while, similar to OLS regression and in contrast to maximum likelihood estimation (Tiefelsdorf & Boots, 1995). More recently this is represented by representation in several open source software and programming packages that are relatively easy to use (Oshan, 2016). Thus, Poisson regression was suggested to be able to overcome of the issues highlight in the earlier methods of spatial interaction model calibration.

Taking these results, we can assume a Poisson distribution of flows which would suggest that the probability that k individuals would be recorded as moving between and origin, i, and destination, j, could be modelled as:

$$Pr\ (n_{ij} = k) = \frac{e^{-\lambda_{ij}}\lambda_{ij}^{k}}{k!} \qquad \text{Eq. 31}$$

This can then be translated to the spatial interaction model, where λ_{ij} is the expected value from the Poisson probability distribution. We can specify the calibration of the model in a Poisson regression form by assuming that the estimate of flow value, λ_{ij}, is logarithmically linked to a linear combination of the logged independent variables. The original application of this methodology was to the unconstrained model which could be specified as:

$$\lambda_{ij} = exp\,(\beta_0 + \beta_1\, ln\, P_i + \beta_2\, ln\, P_j - \beta_3\, ln\, d_{ij}) \quad \text{Eq. 32}$$

To calibrate this relationship then, an iteratively reweighted least squares procedure could be used to estimate the parameters. Results from from Nelder and Wedderburn (1972) and McCullagh and Nelder (1983) show that estimation in this case is equivalent to that of maximum likelihood estimation. This thus means that the results from the Poisson regression model would be equivalent to those in Batty and Mackie (1972), assuming that Poisson distribution holds for the underlying data. Thus, the benefits of parameter estimation from maximum likelihood would be achieved, along with avoiding the issues of linear estimation, doing so in a more easily implemented methodology than those presented by Batty and Mackie (1972).

To see how well this calibration method performs, Flowerdew and Aitkin (1982) use British migration flows and found that the Poisson regression model outperformed that of linear calibration as measured through the chi-squared statistics. An issue with this however is that the linear model that they calibrated was not totally constrained, while their Poisson regression was, leading to an unfair advantage and thereby nullifying their conclusions. Fotheringham and Williams (1983) thus rectified this by comparing a constrained model through both OLS and Poisson regression. In doing so, they confirmed the suggestion of Flowerdew and Aitkin (1982) that the Poisson model performs between that a lognormal model calibration through OLS, but this was in terms of the R^2 value. Diving deeper into their results, it can also be seen that the Poisson regression model also showed an improvement in the prediction if large flows as opposed to the log-normal model, reducing the tendency to underpredict. This therefore shows the useful application of the Poisson model to spatial interaction models, overcoming some of the key issues of the linear model and the maximum likelihood estimation method, while providing a better fit to the data.

While this was originally applied in the case of an unconstrained model, subsequent papers by (1987), Davies and Guy (1987) and Flowerdew and Lovett (1988) extended the application of the Poisson regression methodology idea to that a singly constrained model, in particular to the production constrained model. The importance of this is that both OLS and maximum likelihood estimation had difficulty in estimating these models because of the non-linearity of the models. While OLS could not deal with non-linearity in the model as already mentioned, maximum likelihood estimation, despite producing accurate parameter values, also struggled because of the balancing factors (Batty & Mackie, 1972). These often meant that calibration of the model through maximum likelihood estimation become complicated because in solving the multiple equations derived from the constrained models often meant that several well known procedures for maximum likelihood estimation would get stuck in a local rather than global minima (Davies & Guy, 1987). This meant that sometimes maximum likelihood estimation did not produce accurate parameter estimates, especially for the more complicated models.

However, Flowerdew and Aitkin (1982) suggested that Poisson regression could be extended to the single constrained model and beyond. This could be done by adjusting equationEq. 32 by replacing $\beta_1 \ln P_i$ in the model with origin specific factors such as a_i:

$$\lambda_{ij} = exp(\beta_0 + \beta_3 \, ln \, D_j - \beta_3 \, ln \, d_{ij} + \, a_i) \quad \text{Eq. 33}$$

This a_i factor essentially acts as a categorical variable for each origin and would thus ensure that the origin and total constraints are held within the model if data about the origins are known (Guy, 1987). Applying this to the same dataset as the original paper, Flowerdew and Lovett (1988) show that estimates produced by this model show related, but not identical, parameters to those produced by an entropy maximising model, and show overall good model fit. Thus, complete accuracy is seen as a potential trade-off for ease of application and speed of implementation with the results being seen as close enough to justify its continued use.

These results suggested that the model could also be extend to the destination constrained model by replacing the destination mass terms with a destination dummy variable, or to the doubly constrained model through a dummy variable for both the origins and the destinations (Davies & Guy, 1987). Indeed, subsequent applications of the spatial interaction model through Poisson regression has shown that this is the case (Burger, et al., 2009). Thus, the Poisson regression formulation is seen as an easy and interpretable way to calibrate spatial interaction models, removing the issues associated with the linear regression estimation procedure while also reducing the complexity associated with entropy maximisation or maximum likelihood estimation procedures. This is done while still producing the same outcomes in terms of performance and parameters as maximum entropy and maximum utility models (Tiefelsdorf & Boots, 1995). Furthermore, just like the previous models, this could be extended in ways such using origin specific parameters, the incorporation of other dependent variables if their relationship is multiplicative, and it could be adjusted to represent the competing destination model (Guy, 1987). Thus, using the Poisson regression methodology allows the same flexibility of the original formulation of the models, while retaining simplicity and providing increased accuracy in application to flow data. This has therefore led to this methodology becoming common practice in the literature in terms of application of spatial interaction models.

Despite Poisson regression being recognised as one of the best methods for calibrating spatial interaction, there are still areas of debate in the literature. The first is over the assumption of the mean being equal to the variance as its relation to reality. While such an assumption was seen to free the model from the even more restrictive assumption of a constant mean, as seen in linear regression, it is still seen as a constraint in the model that must be assumed for the Poisson regression model to hold (Davies & Guy, 1987). This has therefore led to some to suggest the use of a negative binomial model when overdispersion is seen within the Poisson regression model. This is because this would suggest that the variance in the data is larger than the mean value and thus violates the assumption of the Poisson model and lead to incorrect estimates (Burger, et al., 2009). Specifically, overdispersion is likely to lead to incorrect significance levels of parameter estimates which may suggest significance where there is none and hence incorrect conclusions (Burger, et al., 2009). Thus, in some cases it could be argued that the Poisson regression model in some cases would be inappropriate.

However, the negative binomial regression model is known to have its own issues when it comes to spatial interaction modelling. In the case where flows are non-discrete units, such as revenue estimates or weights, then the negative binomial model cannot be used as the estimates from the model are scale dependent (Kritsztin & Fischer, 2015). This means that depending on the scale selected, millions of pounds, thousands of pounds, hundreds of pounds etc. then the results from the model will be different, leading to the model not being robust under different specifications and hence leaving the door open of manipulation of the results on behalf of the researcher or incorrect conclusions. In contrast the Poisson regression is seen as robust enough to be able to deal with these issues. This is because, despite Poisson regression assuming non-negative integer values, it has been shown that Poisson regression can still be used for non-discrete units such as flow of monetary values in the case of the retail model (Motta, 2019). Thus, negative binomial methodology is not commonly utilised within the existing literature with the limitations of the Poisson model accepted due to its robustness.

The second main critique of the Poisson regression usage is in the case of excess zeros. While zero flows can be dealt within in Poisson regression, because it allows for zero flows, if there are a greater number of zero flows within the results than would be suggested by a natural Poisson distribution then it is suggested that this could be due to different processes operating on zeros and actual values. An example of this is suggested in the trade arena where zero flows may be indicative of no resources rather than influences of distance, or in the case e-commerce sales where zero flows could indicate a lack of provision rather than a lack of choice (Santos Silva & Tenreyro, 2010). In this situation, it has been suggested that a zero-inflated Poisson or negative binomial model could solve this issue. This is because these models (despite the above critique of the negative binomial model) separates out two processes by firstly determining whether a value should be zero and then whether those zero values fit within the traditional model (whether that is Poisson or a negative binomial model) (Burger, et al., 2009). Determining when such a model should be used can be done using the Voung statistic alongside the likelihood ratio statistic that test for overdispersion (for selecting the negative binomial model).

Again however, issues affect the choice of this model. It is acknowledged that different statistics testing for these conditions can point to the use of different models, whether that is the base model or the zero inflated model. This means that it is often acceptable to use the base Poisson specification in most scenarios as it has been shown to be well behaved, even in the case of excess zeros (Santos Silva & Tenreyro, 2011; Kritsztin & Fischer, 2015). In this case, while excess zeros could be identified and could affect the model, Poisson regression is still seen as robust under these conditions and the parameters estimated from it to be valid. This has meant that, while alternative forms of the Poisson regression specification can be used under different scenarios, it remains the most common method for dealing with spatial interaction models currently and is thus utilised in this work.

This therefore highlights the continued debate in the literature as to the most appropriate calibration method for spatial interaction models. Indeed, while calibration via log-linearisation and OLS is common, especially so within the economic discipline, several issues have been highlighted with its implementation, suggesting that its use can lead to incorrect or inappropriate conclusions. Maximum likelihood estimation was suggested to resolve many of these issues, leading to correct estimation of the parameters, but the difficulty of calibration the spatial interaction models through this methods is highlighted through the lack of usage of this method in both practice and in the literature. Thus, Poisson regression has been suggested to be the best method for calibrating spatial interaction models, with support in the literature finding that results derived through this method are more accurate than OLS and equivalent to those found through maximum likelihood estimation. This is not to say that use of the this method does not have its own issues, but that the ease of implementation and general robustness of the method are suggested to outweigh any of concerns or additional issues that arise with otherwise suggested models.

4.4) Spatial Interaction Modelling Evaluation

Although calibration is a key part of the application of these models, and there has been some debate in the literature presented above, we also need to understand how to evaluate and validate the output of these models in relation to the data that we have. What this means is that we need a reliable measure, or group of measures, that are able to tell us how accurate the models we calibrate are (Wilson, 1976). This is done through goodness of fit statistics which often have two main purposes. The first is to examine the accuracy of different models in relation to the same dataset, or the accuracy of the same model across different datasets, while the second is to determine whether there is a significant difference between actual and predicted values (Knudsen & Fotheringham, 1986).

For our purposes, spatial interaction models tend to produce two key outputs: the trip length frequencies and the overall origin-destination matrix, both of which can be used to evaluate how well the model fits the data (Black & Salter, 1975). In particular, the former output can be used to see how well the model performs on the overall system, as replicating trip distance is seen to replicate overall behaviour, while the latter can be used to measure how well the model represents individual behaviour and flows. Thus, ways to accurately measure the performance of these outputs relative to the data are needed, alongside being able to compare the performance of different models against the same data.

4.4.1) Trip Length Frequency Metrics

In terms of trip length frequencies, a common evaluation metric for this is the average trip distance (ATD). The aim of this is to evaluate whether the average trip distance predicted by the model is close to that of the observed average trip distance (Newing, et al., 2015). The basis for this is that if the model can be seen to replicate the overall trip behaviour in the data, as represented by the average trip distance, then the model is likely to be able to well represent the overall system behaviour (Batty & Mackie, 1972). This metric takes the form:

$$ATD_{pred} = \frac{\sum_{ij} \hat{T}_{ij} d_{ij}}{\sum_{ij} d_{ij}} \quad \text{Eq. 34}$$

$$ATD_{obs} = \frac{\sum_{ij} T_{ij} d_{ij}}{\sum_{ij} d_{ij}} \quad \text{Eq. 35}$$

$$ATD = \frac{ATD_{pred}}{ATD_{obs}} \quad \text{Eq. 36}$$

Where $\hat{T}_{ij}$ represents predicted flows, T_{ij} is the observed flows, and d_{ij} is the distance between the origin and the destination. The closer this result is to 1, then the closer the predicted average trip distance is to actual trip distance and hence the better the model is believed to replicate actual spatial interaction. As such, this metric has been used in most modern models to see whether the model represents the overall behaviour of the system (Newing, et al., 2015). The issue with this metric however is that there are no commonly accepted standards or ranges which this metric should take for a model to be classed as a “good” or an “acceptable” model. Thus, this metric is often limited to being used to suggest which model is better able to replicate the same data, rather than being able to compare the same model over different datasets or how well the model represents the underlying data.

4.4.2) Origin-Destination Matrix Metrics

It becomes more complicated however when attempting to evaluate how well the predicted individual flows represent the flows in the data. This is because there are many common metrics that are used, such as the correlation coefficient, the coefficient of determination, SRMSE and the chi-squared statistic, for which there is much debate in the literature as the best statistic to use for spatial interaction models. Firstly, one of the most commonly used statistics to evaluate the accuracies of spatial interaction models in comparison to the actual data is the correlation coefficient (r) (Wilson, 1976). However, its use in practice is often critiqued. This is because the metric is used to measure the degree of linear dependence between two random variables, whereas spatial interaction flows often do not follow a linear relationship. In all cases, it is expected that the predicted interaction should be positively related to the actual interaction because of the formulation of the model, suggesting that high r values would be meaningless (Wilson, 1976). Thus, it is suggested usage of this statistic, and using it to support an individual model, could lead to incorrect conclusions because of the nature of the model.

Related to the use of the correlation coefficient is also the coefficient of determination, often referred to at the R^2 value, which is commonly associated with linear regression results. This metric is primarily used to measure the amount of variation in the dependent variable that is captured by the model, with values closer to 1 representing a better model fit. In relation to spatial interaction models however, it suffers from similar issues as the correlation coefficient as mentioned above. Furthermore, previous research has noted that the R^2 is insensitive to variation in the model specification, meaning that you cannot tell which model is “better” through the use of this statistic, or can in some cases yield artificially high values (Knudsen & Fotheringham, 1986). This means that it can be difficult to interpret and evaluate the meaning of high R^2 values often present in spatial interaction models. Furthermore, it has been argued that it has limited application across datasets as the value is a function of the variance of observed data, meaning that the results of the value will depend on the data applied (Knudsen & Fotheringham, 1986). Thus, in theory it should not be used to compare different models, or even the same model, across different datasets.

However, despite these issues, while the correlation coefficient is less commonly used in practice, the coefficient of determination is often applied. This is because of its ease of interpretation, in that a higher R^2 suggests a better model and that values over 0.8 are suggested to indicate a good model fit, and its use in several other domains of modelling relationship. As such, for those coming from different disciplines, or areas of modelling, seeing and understanding what the R^2 means for a models should be relatively easy. Thus, the coefficient of determination continues to be used in most modern applications of spatial interaction models and their evaluation, despite its limitations (Newing, et al., 2015). This does not mean that it should be used alone however.

Due to these limitations, it has been argued that the R^2 value could be supplemented, or even replaced, by the more appropriate pseudo R^2 value, especially when Poisson regression is used. This metric is based on the likelihood function from the model and takes the form:

$$R^2_{pseudo} = 1 - \frac{ln\,\hat{L}(M_{full})}{ln\,\hat{L}(M_{intercept})} \quad \text{Eq. 37}$$

Where $\hat{L}$ is the likelihood of an estimation model, M_{full} representing the current full model and $M_{intercept}$ representing the model only with the intercept (Oshan, 2016). The aim of this metric, as the R^2 for linear regression, is to show the improvement in explanatory power as a result of the application of this model, but to do so as opposed to a simple base model that contains only an intercept value. As such, higher values are associated with greater explanatory power, with 1 representing a complete model.

This metric could be further adjusted, like actual R^2, to account for model complexity by including the number of regressors, K:

$$R^2_{pseudo} = 1 - \frac{ln\,\hat{L}(M_{full}) - K}{ln\,\hat{L}(M_{intercept})} \quad \text{Eq. 38}$$

This implementation thereby attempts to compensate for the fact that more independent variables, regardless of whether they are important or not, will result in a greater pseudo R^2 value. Thus, this adjustment means that the value will only increase if the new independent variables actually add explanatory power to the model. These metrics are therefore suggested to be more appropriate for spatial interaction models when Poisson regression is used, as opposed to the R^2 value, because of the non-linear nature of the data. This metric can also be reliably used to measure how well the model explains the variance in the dependent variable, and thus can also be used to compare models across the same dataset.

Another key measure often employed in spatial interaction modelling evaluation is that of root mean squared error (RMSE) and standardised root mean squared error (SRMSE). The former is given as the standard deviation of the residuals of the model, with lower values indicating greater fit as the prediction points lie closer to the actual values. The issue with this however is that it does not lend itself well to the comparison across different datasets, at the value of the RMSE relates to the underlying variance of the data (Oshan, 2016). That is, regardless of the same model being used and the same calibration methods, a dataset with greater variance would produce a greater RMSE value. Thus, the RMSE value can often be standardised by accounting for the underlying variance within the data, which allows for the comparison of models on the same dataset and also across datasets (Black, 1991). This metric has a broad applicability across many different models beyond spatial interaction models and is a commonly used evaluation metric (Fotheringham & O'Kelly, 1989). This means that, as with R^2, individuals coming from different domains or models should understand how to interpret this metric in relation to spatial interaction models. Specifically, the SRMSE value has a lower limit of 0, indicating a perfect model, with a commonly assumed upper limited of 1 (although in reality the upper limit depends on the distribution of actual flows (Knudsen & Fotheringham, 1986)). Thus, the RMSE and SRMSE values are commonly used in the spatial interaction modelling domain due to its ease of use, applicability to spatial interaction models and that it can be commonly interpreted from a variety of domains.

Beyond these statistics, the chi squatted test is seen as goodness of fit statistic that fits within the idea of origin destination matrices. The idea is that this could be used to test between the observed frequencies of the number of trips and the modelling frequencies or the appropriate values in the trip length distribution (Wilson, 1976). Thus, this could be used to determine whether or not the null hypothesis, of whether the model does not fit the data, or not. The benefit of this is that it could be used to test whether the model predictions are significantly different to the actual flows, and importantly it is suggested to be sensitive to underpredictions in both small and high flow magnitudes (Knudsen & Fotheringham, 1986). Thus, using this statistic could be used to suggest whether the modelling results are significantly different from the actual data, and thus whether the model is valid or not.

However, there are noted to be several issues with the usage of this statistic in relation to spatial interaction models. Notably, in order to use the statistic a threshold level of flows is required, which may result in aggregations needed to be performed, which will distort the actual flows taking place and thus will not be representative of the actual system (Knudsen & Fotheringham, 1986). This is especially so in cases where there are a high number of zero flows in the system, which is often likely in spatial interaction modelling, thereby limiting its practicality. Furthermore, the statistic has often been shown to generate relatively high values, as with the R^2 values, suggesting that there could be issues in the modelling procedure and provides relatively little information to evaluate the models themselves (Wilson, 1976). Thus, while there are some instances of the use of this statistic to evaluate spatial interaction models, the use of this metric is generally limited and not often used.

Beyond the distance and statistical based metrics presented above that have been variously used to evaluate spatial interaction models, Knudsen and Fotheringham (1986) identified a set of goodness of fit statistics known as information based statistics, that could be used to evaluate spatial interaction models. These statistics have their basis in in Kullback and Leibler' information statistic:

$$I(\boldsymbol{P}:\boldsymbol{Q}) = \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}\, ln\,(p_{ij} \backslash q_{ij}) \qquad \text{Eq. 39}$$

Where m and n are the number of origins and destinations respectively and p_{ij} and q_{ij} are elements of a posterior discrete probability distribution, **P**, and a prior discrete probability distribution, **Q**, respectively. Here p_{ij} and q_{ij} are defined as:

$$p_{ij} = t_{ij} / \sum_{i=1}^{m}\sum_{j=1}^{n} t_{ij}$$

$$q_{ij} = \hat{t}_{ij} / \sum_{i=1}^{m}\sum_{j=1}^{n} \hat{t}_{ij} \qquad \text{Eq. 40}$$

Where t_{ij} is the observed flow between i and j, while $\hat{t}_{ij}$ is the estimated flow. When applied to spatial interaction models, this would have a lower limit of zero, which corresponds to the perfect set of predictions, and an upper limit of infinity. The benefit of these type of statistics would be that the significance of the metric could be found through its relationship to the minimum discriminant information, thereby being able to suggest whether there is any significant difference between the modelling output and the actual data (Fotheringham & O'Kelly, 1989). Knudsen and Fotheringham (1986) highlighted that this could be useful to evaluate a single model against the data, although they suggest that this could be dependent on the dataset that is fed into the model, with no clear distinction as when it is useful or not. Thereby leading to questions over its usefulness, indeed shown by the lack of relative usage of such statistics within the spatial interaction modelling literature since.

Similarly within this information theory domain is that of Akaike information criterion (AIC) which takes the form:

$$AIC = -2 \ln \hat{L}(M_{full}) + 2K \quad \text{Eq. 41}$$

Where a lower value is taken to indicate a better model fit. This is another information based statistic, that has become more common recently due to its usage in the data science domain. This model could suggested to be used for model selection, as above, but is limited on its ability to compare across datasets or spatial systems (Oshan, 2016). This has been used in some of the more recent adaptations of the spatial interaction model, notably in the radiation model implementation, but still has relatively little usage in the traditional spatial interaction modelling literature.

Alternative metrics also include the modified Sorensen Similarity Index (SSI) that has become increasingly popular in the literature that deals with non-parametric models such as the radiation model (Lenormand, et al., 2012; Masucci, et al., 2013). This is defined as:

$$SSI = \frac{1}{nm} \sum_i \sum_j \frac{2 \min(\mathrm{T_{ij}}, \widehat{T_{ij}})}{T_{ij} + \widehat{T_{ij}}} \quad \text{Eq. 42}$$

Where *n* represents the number of origins, *m* the number of destinations, T_{ij} is the flows between origin i and destination j, and $\widehat{T_{ij}}$ is the predicted flow between origin i and destination. This metric is bounded between values of 0 and 1, with values closer to 1 indicating a better model fit (Oshan, 2016). However, the implementation of this is often highly sensitive to the number of zero flows within the data which could affect the relative value and hence assumed model performance (Oshan, 2017).

Similar to this is the Common Part of Commuters (CPC) metric, which is often known as the Sorensen Dice metric, and is a well-established measure to compute the similarity between real and generated flows. When the generated total outflow is equal to the real total outflow, as in a constrained Wilsonian model, CPC becomes equivalent to the accuracy of the model, which is the fraction of trips and destinations which are correctly predicted by the model (Simini, et al., 2021). This is implemented as:

$$CPC = \frac{2\ \sum_i \sum_j \min\left(T_{ij}, \widehat{T_{ij}}\right)}{\sum_i \sum_j T_{ij} +\ \sum_i \sum_j \widehat{T_{ij}}}$$ Eq. 43

Where T_{ij} is the flow between origin i and destination j, and $\widehat{T_{ij}}$ is the predicted value between origin i and destination j. The CPC value itself is always positive and contained between 0 and 1 where 1 indicates a perfect match between the generated flows and the ground (Ying, et al., 2019). As such, it has been used more recently in data science applications to spatial interaction models and radiation model implementations (Masucci, et al., 2013; Piovani, et al., 2018), for example Lenormand et al. (2012) used this calibrate their model to achieve the maximum value. A drawback of this however is that small percentage deviations in the predictions of large flows can have significant impact on the value of the coefficient (Hilton, et al., 2020).

This exploration has therefore highlighted that there are a wide variation of potential goodness of fit indicators that could be applied to the evaluation of spatial interaction models, and that to date there is no generally well accepted singular indicator that is commonly used (Black, 1991). This has meant that despite considerable issues identified with R^2 values and SRMSE values as presented above, they are both often employed in modern spatial interaction models due to their ease of implementation, general understanding of their purpose and interpretability relative to what they mean for spatial interaction models (Newing, et al., 2015). This means that using these statistics would be consistent with the history of spatial interaction modelling in the literature. However, more recent application of the models have started to use information gain statistics or alternatives alongside R^2 and SRMSE values in order to provide further evaluation of the existing models, and acknowledge the differences in the useful of the different metrics. This therefore follows Knudsen and Fortheringham's (1986) suggestion that a variety of statistics should be used to evaluate spatial interaction models. Therefore, in progressing with the rest of this paper, a mixture of R^2, SRMSE, and ATD values are used to interpret and understand models, supplemented with other metrics as and when they are available.

4.5) City Level Implementation

The main contribution of this thesis will be an evaluation of the effectiveness of spatial interaction models on a spatial-temporal scale that has not been made available before. However, before getting to that scale, these models must be first implemented and evaluated on an existing scale to be able to identify any challenges that may arise for the large scale implementation. Thus, with the literature identified above in terms of both spatial interaction modelling and grocery retailing, the first application of the models will be that of an individual city scale.

4.5.1) Data Overview

For this thesis, we have access to anonymised loyalty card data from a single UK national grocery retailer which currently holds a significant portion of the UK market share and is responsible for thousands of stores nationwide. The dataset that we have from this retailer includes their loyalty card scheme which has been running for over 20 years, with our access covering the last five years. The benefit of this scheme for this research and the retailer is that these loyalty cards can be used to obtain information about customers in terms of their home location, where they choose to shop, when they choose to shop and what they choose to buy. While from the original data the retailer can create customer classifications, track where spending comes from and identifying different shopping habits of the consumer, the interest for this research is to focus on the locational aspects of the data. This will be used to examine where households and their spending comes from at an aggregated scale, where it goes to and what format of store they tend to shop at. It is acknowledged for the purpose of this research that examining data from a single retailer is likely to lead to bias in model calibration (Rains & Longley, 2021), it is a limitation that we must accept and which is unlikely to be overcome by any researcher in the future. This is because the grocery retailing industry in the UK is a highly competitive industry, and access to a data source that combines data from a variety of different retailers is highly unlikely. Thus, it can be accepted for the purposes of this research, and for others in the field, that access to a single retailer is the best data source we are likely to get access to, at least for the foreseeable future.

The underlying anonymised loyalty card data from the retailer is linked to an address at which the loyalty card has been registered and contains information about where the consumers shop, when they shop and what they purchase. Such data would be highly granular, allowing an insight into individual households around the geography of their shopping habits. However, in accordance with private regulation, the data provided for this thesis has been aggregated to ensure that no individual could be identified. This is because, even with anonymised data, the level of granularity of loyalty cards data showing shopping trips from origins to destinations may allow for identification of individuals. Thus, the data in our case has been aggregated to the output area scale for England and Wales. This scale of geography was purposefully created for the use with census data and is the lowest geographical scale at which census estimates are to be provided, which is accepted in the literature as the greatest level of granularity that could be achieved without links to individual address (Newing, et al., 2015). We are able to utilise this scale of geography, and hence census data, due to the highly disaggregated nature of the underlying anonymised loyalty card dataset and the large market share that the grocery retailer has across the UK. This data thus contains information on the number of households and sales value that travels between an output area and a destination store per week from the beginning of 2015, including the value of shopping that is broken down into high level categories. However, in accordance with data protection regulation, data cut offs were established by the retailer to ensure that no part of the data could be used to identify an individual. This means that if the number of households fell below a given threshold (5 households) for that individual week, then that data was not included in the final dataset. The effect of this is suggested to be of minimum consequences to the analysis however as the output area that this affects are expected to be infrequent shoppers that, considering the scale of the remaining data, should not affect any of the parameter estimates from the model.

The output area scale of geography was chosen for this aggregated primarily because it is the smallest scale of census aggregation in the UK, which ensures that census level data, and thus demographic information, can be attributed to these individual areas, without the need for further aggregation or adjustments. The original design of these areas but the ONS was such that they would have similar population sizes and that they represent relatively socially homogenous groups of individuals based on tenure of household and dwelling type, while still being large enough to ensure data confidentiality (ONS, 2013). For the 2011 census, there were a total of 171, 371 output areas for England and 10,036 for Wales, with an average population count of 309. Thus, given the data that this research has access to, a significant amount of loyalty card revenue is likely to be captured even after the GDPR cut-offs have been implemented. In terms of modelling, this scale is also small enough to be able to create detailed disaggregated models that can assume homogeneity in each output area (Newing, et al., 2018). In contrast, larger geographical scales such as LSOA, MSOA and the postcode sector scale are likely to have less homogenous populations, while variations in geographical distance (relative to the centre of the centroid) would likely significantly influence the accuracy of the models (Aiello, et al., 2020). Thus, the output area scale was chosen as best scale for the data to be aggregated to, in support with most modern implementation of the spatial interaction models in the UK (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021).

A benefit of working with data aggregated to this scale is that, given their design towards relative and social homogeneity and the link with census data, researchers have previously been able to classify output areas in several different groups based on demographic characteristics (Gale, et al., 2016). This was done for the 2011 census data following the example laid out by Vickers and Rees (2006) for the 2001 census. The classification of output areas, according to demographic values, was constructed by firstly selecting a subset of census variables that weren't significantly correlated with each other but were deemed to be of socio-economic interest, implemented a k-means clustering algorithm to standardised versions of these variables, evaluating the most accurate number of clusters and then consulting with relevant parties to identify the best performing results. The consequence of this was the creation of 8 supergroups, 26 groups and 76 subgroups for the 2011 census, allowing for disaggregated spatial interaction models to be developed. These models are primarily built in the 8 supergroups, as shown in table 3 below, because further disaggregation would lead to inaccurate and inappropriate models.

Table 2 - 2011 Census output area supergroups (Gale, et al,. 2016)

Supergroup number	Supergroup title
1	Rural Residents
2	Cosmopolitans
3	Ethnicity Central
4	Multicultural Metropolitans
5	Urbanites
6	Suburbanites
7	Constrained City Dwellers
8	Hard-Pressed Living

With this information, initial models could then be constructed on the basis of a small subset of anonymised loyalty card data to be able to test the implementation of spatial interaction model on the given dataset. This was firstly performed for stores in a single city within the South West region of the UK. The purpose was to make sure that the model implementation and development was correct and that the calibration and evaluation methods chosen were appropriate to the application. If deemed to be so, then the idea is that these initial models could then be scaled up to explore the limits of spatial interaction models in the face of spatially and temporally large datasets.

For this implementation, it has been acknowledge that behavior can be expected to be different across regions and cities, following the debate that spatial structure can influence individuals decisions and behaviors (Kerkman, et al., 2017). Thus, a single city was initially chosen in order to minimise these expected influences on the results and to ensure a fair implementation of the model. The city selected for this first evaluation was chosen because previous research by dunnhumby, the funding partner for this research, suggested that there was limited leakage of loyalty card data, in the form of grocery revenue, from the city and thus lending itself to being modelled consistently by a spatial interaction model.

4.5.2) Modelling Methodology

These spatial interaction models, based on the data available, aim to examine the flow of households and/or money from an origin to a destination. Beyond the anonymised loyalty card dataset that we have access to however, we can also use information that is available from the census and other sources. For the purpose of this research, we can use known information about the origin to be able to create more accurate and reliable models as we can reasonably estimate the expected outflow of either households or money from each output area in the UK. This is because, while data from the anonymised loyalty card dataset can be aggregated at the output area level to see the total revenue leaving an origin, census data could be used to estimate the total population and hence revenue available in an origin. For this, the number of households in each output area could be obtained from the 2011 census (assuming no change in household numbers for now), and the average expenditure on groceries for each household could obtained from the living costs and food survey (ONS, 2021). The latter of which can be broken down by output area classification, so as to more accurately provide estimates of the expenditure from each origin. This can be estimated as:

$$O_i^{kt} = e^{kt} n_i^{kt} \qquad \textit{Eq. 44}$$

Where O_i^{kt} is a measure of total expenditure available in origin i by household type k during the year t, e^{kt} is a measure of the average weekly grocery spend for a household of type k in year t, while n_i^{kt} is a measure of the number of households within origin i of type k in period t. While up to date information on the number of households is not available for each year, meaning that n_i^{kt} becomes n_i^{k2011}, up to date information is available for expenditure estimates as the living costs and food survey estimates expenditure each year.

This information can therefore be used to implement a production constrained spatial interaction model, also known as the "retail model". This is because we have information about the expected outflow from each origin, in this case the output areas in the UK, which we can use to constrain the model and hence improve the accuracy of the estimates. The model therefore takes the form:

$$T_{ij} = A_i O_i W_j^{\gamma} f(c_{ij}) \qquad \textit{Eq. 45}$$

Where:

$$A_i = \frac{1}{\sum_j W_j^{\gamma} f(c_{ij})} \qquad \textit{Eq. 46}$$

Whereby T_{ij} represents the flow from origin i to destination j, A_i is the balancing factor, O_i is the outflow of revenue from the origin, W_j is a measure of attractiveness of the destination, $f(c_{ij})$ is the distance decay relationship and α and β are the parameters to be estimated by the model. The form of this model is the most commonly used model in retail analysis because the census information presented above is widely available, meaning that more accurate models can be created and estimated (Newing, et al., 2015; Waddington, et al., 2018). Hence, this thesis follows well accepted conventions in applying this type of model.

The production constrained model can then be calibrated using Poisson regression, as presented in section 5.1, due to both the ease of application, interpretability and generally accepted accuracy in modelling spatial interaction. The advantages of this method includes avoiding the issues associated with OLS regression such that Poisson regression can deal with zero flows, which are highly prevalent in the dataset, constraints can be implemented in the regression framework, rather than post-hoc, and the estimates represent actual flows rather than log flows (Flowerdew & Aitkin, 1982). Furthermore, compared to more complicated maximum likelihood estimation or entropy maximisation methods, there are ready built Python libraries for ease of implementation, such as statsmodels api (Seabold & Perktold, 2010) and SpInt (Oshan, 2016), while retaining similar levels of accuracy (Tiefelsdorf & Boots, 1995). The estimates of the models are then evaluated against the anonymised loyalty card data using measures of R^2, SRMSE and ATD, as mentioned in section 5.2, due to their ease of implementation and wide acceptance of usage (Newing, et al., 2015). Once the models have then been calibrated to replicate the behaviour exhibited in the anonymised loyalty card data, they are then scaled up using the calibrated parameters to be predict total store revenue, which is the ultimate objective of this model implementation.

4.5.3) City Level Data

As mentioned, the first model implementations focused on a subset of the original data to ensure that the model calibration and evaluations were working as expected. The city chosen for this is home to multiple stores, included different formats, that were well represented in the data and that exhibited relatively little data leakage in terms of loyalty card revenue going, or coming from, elsewhere. The data accessed for this city, according to the seven stores in the area, contained flows from 416 individual output areas around the stores, with eight of the seven output areas supergroups represented (no Multicultural Metropolitans), as seen in table 4 below. When we begin to scale this up however, to model the total revenue, we take the maximum distance that revenue is seen from an output area in the loyalty card dataset and draw a circle with that radius around each store. We thus assume that non-loyalty card consumers display similar behaviour to those with loyalty cards, and model total flows as those potentially travelling from each of those origins to each store. From this, in the total dataset, we model a total of 1,203 origins as seen in the Total column of Table 4 below.

Table 3 - Output Area distribution at the city level

OAC	Loyalty Card Count (percentage of total %)	Total Revenue Count (percentage of total %)
1 (Rural Residents)	25 (6.01%)	132 (10.97%)
2 (Cosmopolitans)	32 (7.69%)	101 (8.40%)
3 (Ethnicity Central)	4 (0.96%)	14 (1.16%)
4 (Multicultural Metropolitans)	0 (0%)	0 (0%)
5 (Urbanites)	64 (15.38%)	248 (20.62%)
6 (Suburbanites)	67 (16.11%)	249 (20.70%)
7 (Constrained City Dwellers)	56 (13.46%)	145 (12.05%)
8 (Hard-Pressed Living)	168 (40.38%)	314 (26.10%)
Total	**416**	**1203**

To apply the spatial interaction model to the data, the functional form of the distance decay relationship had to be identified ($f(c_{ij})$ in Eq. 45 above). To this end, Figure 1 below shows both the sales value and the number of households travelling to the stores over distance within the dataset. This figure primarily shows that there is indeed a distance decay relationship to be seen, with both the number of households and revenue falling over distance, but not necessarily what form the relationship takes. To examine this, both the exponential ($e^{-\beta d_{ij}}$) and power ($d_{ij}^{-\beta}$) distance decay functional forms were chosen to be applied to see which performed best on the dataset. This is because, while it has been noted in the literature that the exponential model is expected to fit intra-city movement better (de Vries, et al., 2009), evidence also suggests that the power law distance decay function could also be applicable (Chen, 2015). Furthermore, later implementation of the model on larger datasets struggle with the exponential distance decay implementation, so for the purpose of consistency in model exploration and evaluation, both forms of the distance decay relationship are utilised for initial tests. This is not to discount more complex distance decay relationships however, but they were deemed too complex to be able to apply on this dataset with any reliability and thus were not utilised.

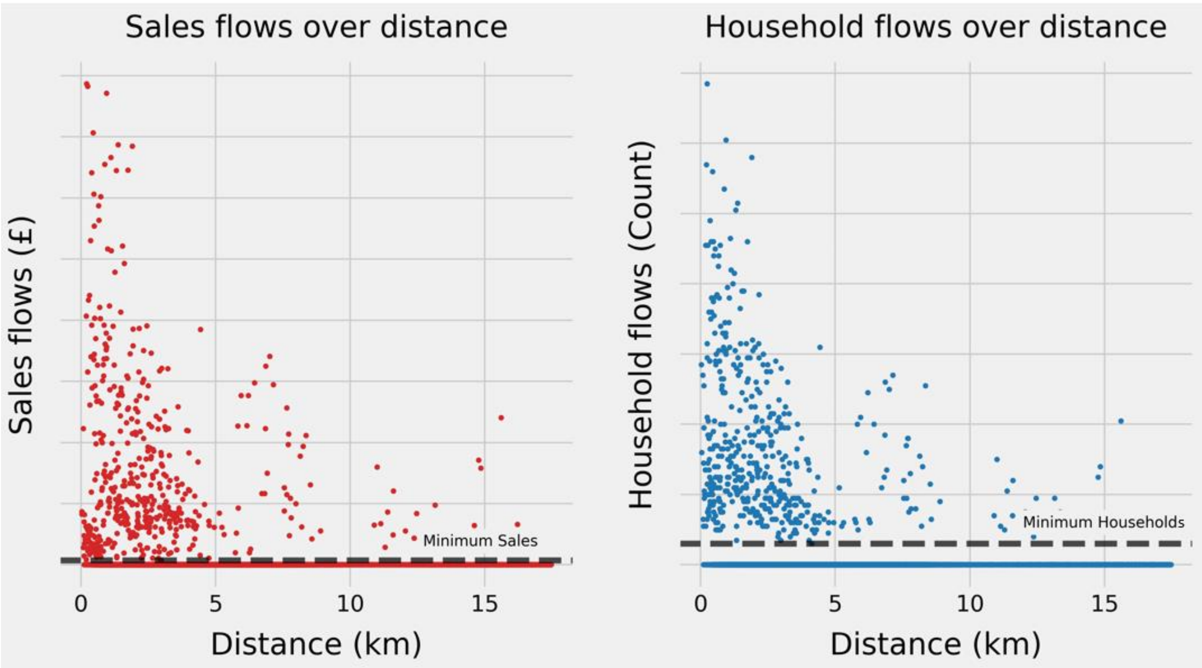

Figure 1 - Amount of flows from origin to destination from the *anonymised* loyalty card data for a) number of households, b) Sales value in £s

Now that we have seen what the data looks like for this example, we are firstly interested in how well the spatial interaction models can replicate the known dataset of anonymised loyalty card data. At this point, it is important to acknowledge that not all transactions occurring at an individual store are represented by loyalty card data. In some cases the percentage of total sales represented by loyalty card spend could be anywhere between 5 and 70%, which can vary by area, time of the year and store format (Rains & Longley, 2021), similar ranges to what is seen in this dataset. Thus, the purpose of the initial model implementation on loyalty card data is to be able to train the spatial interaction model, and to then use the learned parameters to scale up the model to predict total store revenue or even other store revenue. Thus, individual model performance is originally tested against how well it can replicate the anonymised loyalty card dataset, and then if satisfactory, scaled up to model total revenue of the stores and then performance is remeasured. At this point, the model is initially tested on a single week in 2017, consistent through all basic model implementation, so as to remove the influence of any seasonal or yearly effects on the results, such as the influence of holidays or other events.

4.5.4) Loyalty Card Performance

Using the production constrained model, on the anonymised loyalty card dataset, the residuals from the model can be seen in Figure *2* below, while Table 4 shows the performance metrics for the models. Firstly, figure 8 shows the residuals from the model in terms of the model predictions minus the actual flow values against the actual sales value. From this we can see that that the majority of large flows of revenue go to the hypermarket format, followed by those to the supermarket format, while smaller value flows from origins. Furthermore, it is clear that there are a greater number of flows from origins to the larger stores, in line with what we would expect as these larger stores have greater attractive power. Thus, they should reach and influence a larger number of origins than the smaller stores. In terms of the model performance, the residuals suggested that both models perform as expected in terms of variance increasing with expected values, therefore supporting the use of the Poisson regression model. Although, the fact that the residuals are small for larger value flows suggest that these points have significant influence on how the model trains and performs, suggesting that the regression model could be favouring these larger value flows. This therefore suggests that the model perform better on larger stores, considering there is considerable variation for the smaller formats with notable consistent underprediction in both models.

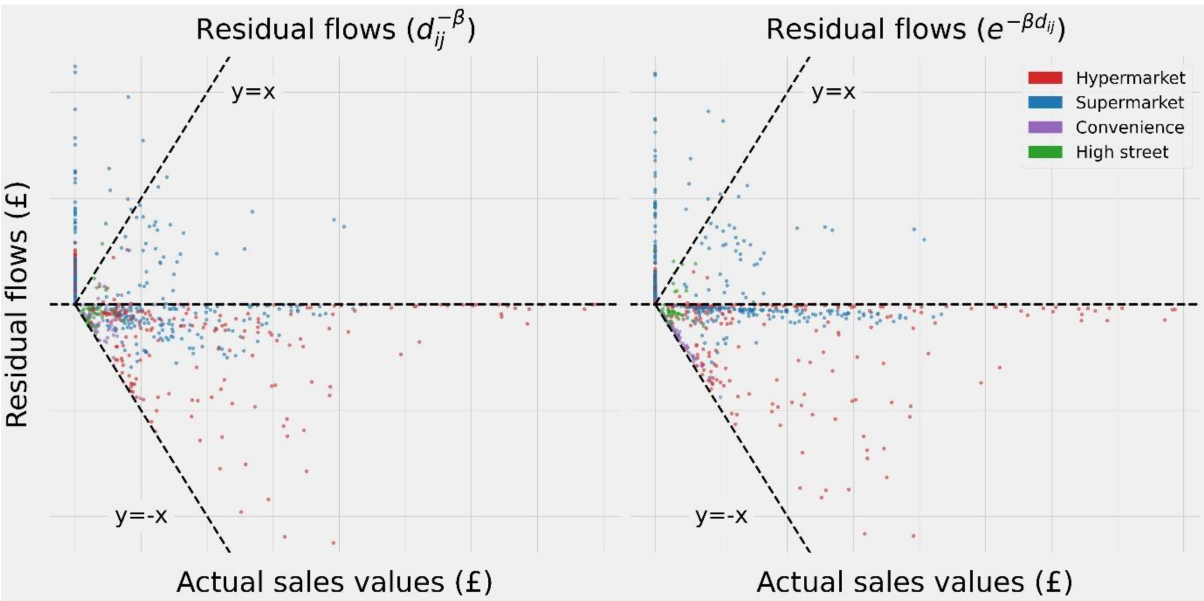

Figure 2 – City level residual flows for the inverse power decay and exponential decay forms of the model across each store format type

Table 4 – City modelling metrics for both the inverse decay and exponential decay model formats for a single wee

Metrics	$d_{ij}^{-\beta}$	$e^{-\beta d_{ij}}$
R^2	0.931	0.929
Pseudo R^2	0.937	0.938
SRMSE	1.537	1.563
SSI	0.042	0.041
CPC	0.855	0.860
AIC	315106	309335
ATD	1.027	1.000

From table 5 above, we can see that the values for R^2 and SRMSE values that while both models perform well, in terms of representing the anonymised loyalty card data, it appears that the inverse power model performs better. This is because for this dataset we can see that the inverse power model has both a higher R^2 value and a lower SRMSE. However, the exponential distance decay model results in an ATD value that is much closer to 1, suggesting that it is better able to fit the behaviour shown in the underlying data in terms of the willingness to travel. The fact that these measures suggest differing performance between models is an issue inherent in the spatial interaction literature, whereby different measures can support different model implementations. On this basis it is therefore difficult to conclude clearly or definitively which model is better on this dataset. Nevertheless, we can continue with the dataset exploration and evaluation.

We can see the geographical representation of these predictions in Figure 3 below, so that we can try to understand how these different models are behaving. Here figure 9a shows the predictions from the inverse power law model while figure 9b shows the results from the exponential decay model. Each of these shows the value of the predicted flow as against the maximum sales from each origin seen in the anonymised loyalty card data. In theory no value should exceed 100% because these are origin constrained models so that all value should be assigned to each origin seen in the data. Consequently, we can see that both models appear to show similar distributions of sales from each origin, and that the majority of these estimates sales cluster around the two larger format stores, in line with what would expect. We can also see that there is little revenue appearing around the five smaller format stores in the centre of the city. For this it is acknowledged that these stores are unlikely to be serving local residential demand, probably focusing on day time population, which are unlikely to be represented here as living around these stores (Waddington, et al., 2019). Nevertheless, the distribution of revenue around all of the stores shows high modelled revenue closer to the stores with lower revenue further away, along with greater revenue closer to larger stores, thus showing results that are consistent with the spatial interaction model specification. The high R^2 values in Table 4 suggest that these results are in line with what we would expect to see from the actual loyalty card data, thus we can be confident in this distribution.

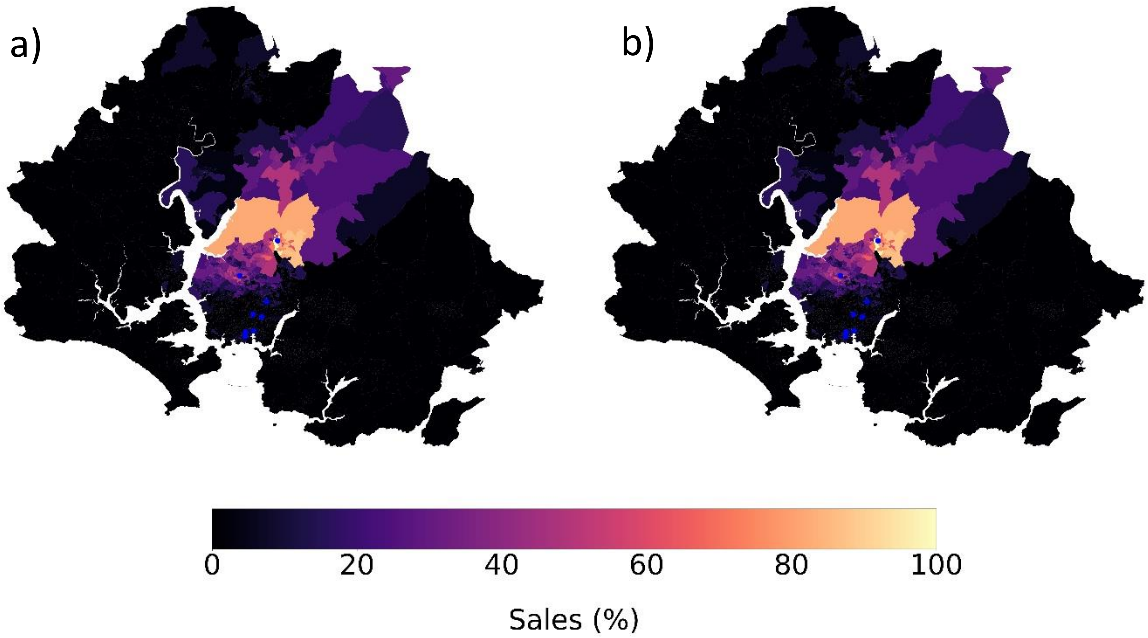

Figure 3 - Predicted flows by model a) Inverse power law, b) Exponential distance decay, as a percentage of the maximum sales revenue from a single origin in the anonymised loyalty card data

At this point it is difficult to separate the two models from other. However, regardless of how well the model is able to replicate the individual flows seen in the anonymised loyalty card data, we are primarily interested in how well the model can predict the store level revenue. This is because, for the retailer, the main data of interest is how much revenue an individual store will make from the surrounding region. If the store is making less than it needs to be profitable then that store can be closed down or replaced, while if a location is seen to be profitable, then a store should be developed in that area. We can thus firstly see how well the model is able to predict the total loyalty card store revenue for an individual week, as can be seen in *Figure 4* below.

This is done by adding up all the flows from each origin to each destination from the anonymised loyalty card data and the model predictions, with the difference in revenue between the model and the actual revenue being represented here. This is shown by the percentage of revenue over or under the predictions are against the total loyalty card revenue i.e. a positive percentage shows overprediction while a negative percentage shows underprediction. From this, we can see that both models perform satisfactorily for the two larger stores (Within 20% error bounds) while it struggles to represent the convenience and high street stores consistently. This therefore reflects the results seen in Figure 2 and Figure 3 above in terms of the residuals from the model and the flows. This is likely to be due to the differing nature of convenience and high street shopping that deviates from the underlying assumptions of the gravity model. This is because shopping at convenience or high street stores is often made up of irregular, small shops that can include impulse shops, alongside the smaller percentage of total revenue being represented by loyalty card data, thereby making it more difficult to model through a spatial interaction model (Waddington, et al., 2018). In terms of separating the models however, it shows that the different distance decay parameters have different performance across each store. Here, the inverse power decay model appears to have closer results to the majority of store total revenue (in 6 out of 7 stores) and thus it could be suggested that it performs better at the store revenue level.

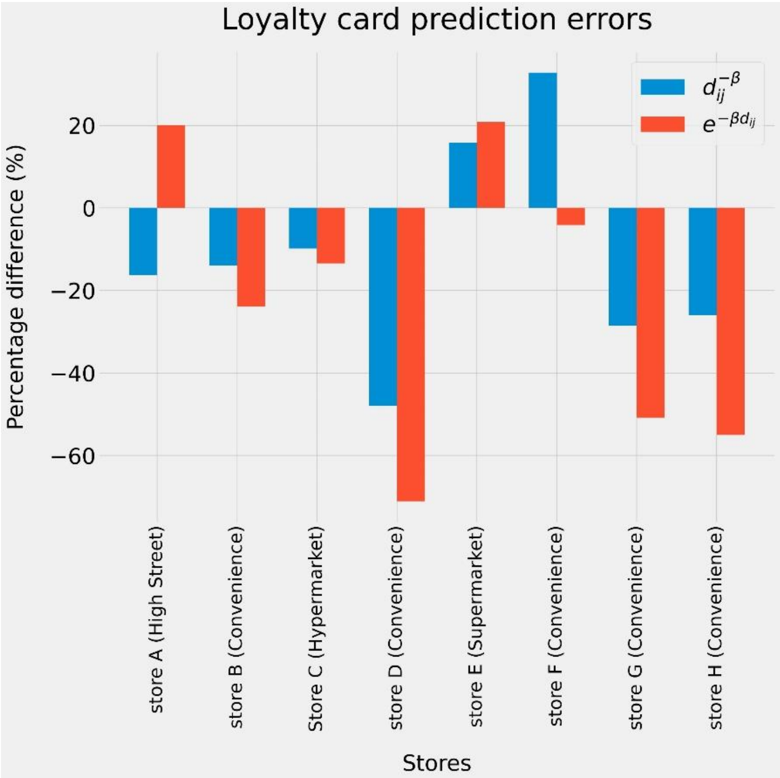

Figure 4 - Weekly store revenue estimates for loyalty card total revenue

4.5.5) Total Revenue Prediction

The issue with this, as already mentioned, is that loyalty cards do not account for all the spending at each store, and that this can vary across each type of store as well. Thus, our aim is not only to be able to replicate the anonymised loyalty card data, but to scale this up to predict total store revenue.

To do this, parameters from the trained model are inputted into the origin constrained model from Eq. 45 above. Potential origins that could patronise these stores are identified as those that have centroids within a radius of 17.5km of the original stores. As already mentioned, this distance comes from the maximum range observed in the anonymised loyalty card dataset. The potential revenue available at each of these origins is then taken to be the average weekly spend on groceries for that output area classification multiplied by the 2011 census number of households. To be able to identify competitor store for the calculation of the balancing factor, each of these origins identified (seen in the total column of Table 3) are then given a radius of 17.5km as well to identify potential competitors in that range. The balancing factor is then calculated as in equation Eq. 45 above. The results from these models is then expected to be able to calculate the true total flow of revenue from origins to destinations, which can be used to see where the expected flows originate from and what the predictions of total revenue are.

Visualisation of the resulting flows can be seen in Figure 5 below, showing where the models predict that total revenue will come from as a percentage of the maximum revenue seen in the anonymised loyalty card dataset. Compared to the flows in Figure 3 the revenues are greater and come from further away, in line with what we would expect when scaling up the revenues to represent the true flows. We also see the same distance decay relationship as for the loyalty card data, following the expectation of the gravity model and the assumption of store patronage. The main difference here however is that the inverse power model seen in Figure 5a shows lower revenue estimates from origins close to stores but show revenue coming from further away, while Figure 5b shows higher revenue estimates from closer origins. This therefore primarily shows the difference in the distance decay specifications in that the inverse power law decays slower than that of the exponential distance decay and thus revenue is estimated to come from further away. The issue however is that the trend shown in figure 11a may be unrealistic due to the use of physical distance from origin to destination, rather than travel time, as revenue is shown to originate from origins that are unlikely to patronise these stores in real life across the river, thus it could be suggested that Figure 5b shows more realistic real flow values from these origins.

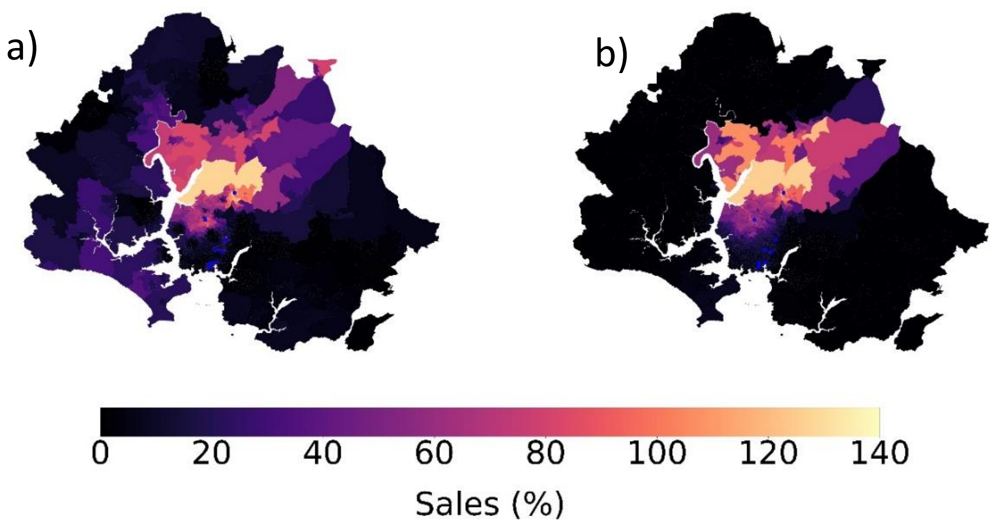

Figure 5 - Total store revenue predictions by a) Inverse power law model, b) Exponential distance decay model as a percentage of the maximum total store revenue from an origin seen in anonymised loyalty card data

We cannot compare these predictions against total revenue flows because we do not have data available on that. Thus, as mentioned previously, our focus is primarily on how well these models can predict total store revenue. The values for this can be calculated the same way as with the loyalty card before with predicted flows being aggregated to the individual store level. Beyond the anonymised loyalty card data, we also have access to the total store revenue data, which we can then use to compare against our prediction. The results from this can be seen in Figure 6 below. As expected, scaling these flows up to predict total revenue results in predictions that are further away from the actual store revenue than the anonymised loyalty card data the model trained on. This is because scaling emphasises and enhances the variation and performance issues in the underlying model. Thus, from these models it can be seen that all the predictions are above or below total revenue for the stores by at least 20%, suggesting an overall poor model fit.

However, it must be highlighted that both models perform significantly worse at predicting total store revenue for the convenience and high street stores. This is likely to be because of the reduced percentage of total revenue captured by the anonymised loyalty card data, meaning that there is less data to train on to get accurate results, that makes it more difficult to scale up to total revenue prediction. Furthermore, as already mentioned, that shopping at the small format stores represents different behaviour from shopping at larger stores. As such, it could be expected that training the model on the complete dataset will have biased performance towards the larger format stores, as indicated by the residuals of the models in figure 8.

In terms of the performance of the different models, from these results it appears that the inverse power law has closer predictions for the majority of stores (5 out of 7). This may be because of the distance that these flows are predicted to come from with some flows coming from a distance much greater than would be suggested than the local geography. While this therefore suggests a preference for the inverse power law decay model, due to the poor overall results, it is actually difficult to draw a definitive conclusion on which model is preferred from these results.

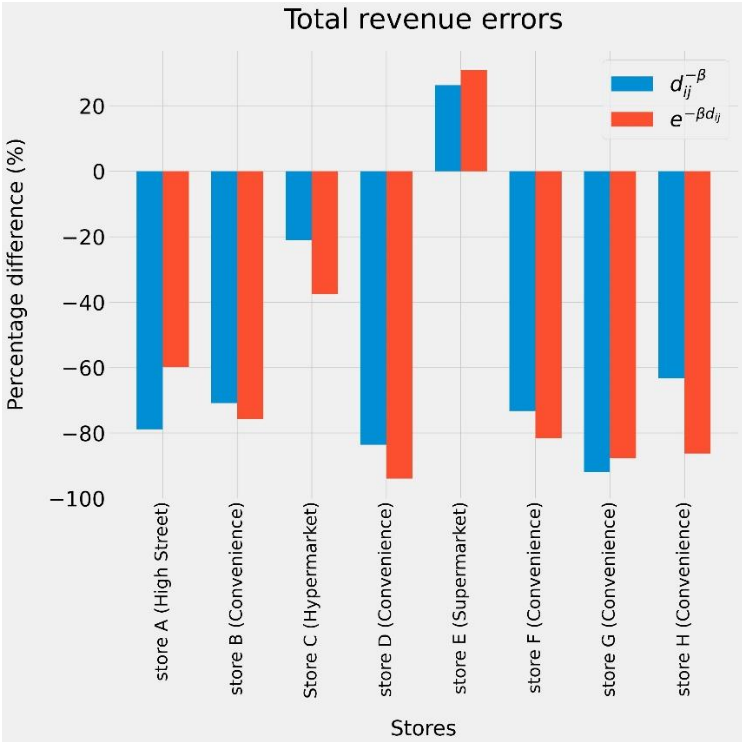

Figure 6 - Total store revenue estimation for the inverse power law model and exponential distance decay model in terms of percentage difference to actual store revenue

4.5.6) Yearly Analysis

Beyond the results of a single week, it is important to ensure that these results weren't the result of a single influence that could affected that week. This is because store revenue and the percentage of revenue captured by loyalty cards can vary significantly throughout the year (Newing, et al., 2015), potentially due to changing behaviour and preferences to shop throughout the year. Therefore, the model was repeated for all weeks throughout the year. The results for this are represented in Figure 7 below at both the loyalty card scale and total store revenue. From this we can see that the results at the yearly scale are consistent with the performance of the model at the individual week. Despite some error bars showing considerable variability in model performance, we can see that as previously suggested at the store level the inverse power distance decay model appears to perform better. This is because at the loyalty card scale, predictions for loyalty card revenue are better for size out of seven stores, while at the total store revenue scale, this model shows more accurate results for five out of seven stores.

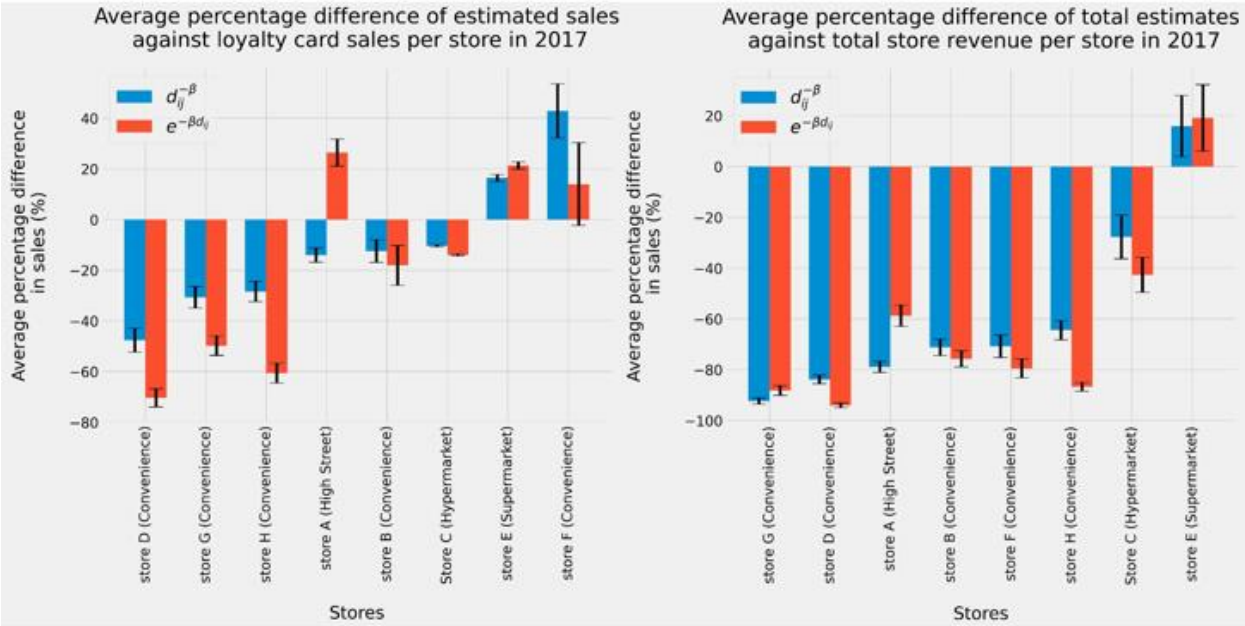

Figure 7 - Yearly store revenue estimates compared to a) loyalty card sales, b) total store revenue

What we can also see from these plots is that the error bars for the larger stores (supermarket and hypermarket), while smaller for the loyalty card data, are considerably larger for the total store revenue in comparison to the high street or convenience stores. This could suggest that while the model trains better and more consistently on the larger stores, because it focuses on these stores more, this performance may not necessarily translate to predicting total store revenue. This could be due to greater variability in larger store revenue, variability in loyalty card coverage or difference performance of this model. It is likely however as a result of the difference in loyalty card coverage, and the affect that has on predictions, because although there is greater variability in large stores, the predictions are far more accurate than for any of the smaller format stores, highlighting that predictions for the smaller format stores are consistently bad.

4.6) Conclusion

This initial exploration and evaluation can help in determining how to proceed with implementing spatial interactions on a large spatial-temporal scale. The first result is that these models can indeed be implemented and evaluated on the current dataset, and that this is likely be able to scale. How far they can scale however is another question that can be answered going forward, with potential limits to computing power affecting how many stores the new models may be able to be implemented on. The second is that when scaling up the model from the anonymised loyalty card data to the total revenue estimation, especially at the store revenue scale, any errors are likely to be magnified. Therefore it is important to pay attention to how these models train on the underlying dataset and to try to identify where the differences between loyalty card and total revenue arise. This could likely lead to explorations of how to improve the model beyond its initial implementation. Thirdly, the models implemented here perform significantly better at modelling the larger store formats. This is consistent with existing literature, which acknowledges that smaller store formats (convenience stores and high street store) revenue is governed by different behaviours (Waddington, et al., 2018). With this in mind therefore, the next stage of implementation and evaluation of the models, can focus on trying to model only large store revenue. Finally, the performance of the different distance decay in the production constrained model on this data does not appear to be clearly separated. It is therefore worth implementing both types of model if possible on the larger dataset.

Chapter 5

Regional Model Application

5.1) Overview

The previous three chapters introduced the history of the spatial interaction model development, the history of grocery retailing in the UK, current discussions in the literature as to calibration and evaluation and an initial implementation of the spatial interaction model on a small city level dataset. This Chapter builds on this work by applying the origin constrained spatial interaction model to region scale anonymised loyalty card datasets for large format stores across a single week and a whole year. The purpose of this chapter therefore is twofold. Firstly, to examine how these models perform when applied on a larger spatial and temporal scale than they have done previously. Secondly, to examine whether the findings and accuracy reported by other similar models can be replicated at this scale. The final results show that the performance achieved in the literature is not able to be replicated at the regional scale with the current dataset.

5.2) Introduction

The previous chapter introduced the application of a non-disaggregated spatial interaction model that was applied to a city level scale using anonymised loyalty card data from a major grocery retailer which contained a variety of store formats. This application was indicative of how the models have been developed and applied so far in the literature due to the limited access that academics have had to commercial datasets. In this regard, Newing et al. (2015) applied a spatial interaction model to four large format stores in Cornwall across an entire year while Waddington et al. (2019) applied a spatial interaction model to 48 stores for a single week in Yorkshire and the Humber, 16 of which were large format grocery stores. This therefore suggests limitations of both geographical and temporal scale in access to loyalty card datasets. However, these limits do not arise do not arise because of a lack of loyalty card data. Indeed, many grocery retailers in the UK have developed their own spatial interaction models (Reynolds & Wood, 2010). Rather, the limitations are because of a lack of access to this data by academics due to concerns over privacy and commercial sensitivity and that commercial research outputs have not been shared (Newing, et al., 2015; Waddington, et al., 2018). Thus while grocery retailers are suggested to use gravity models in their store location analysis (Wood & Reynolds, 2011; Clarke & Birkin, 2018), the accuracy and geographical and temporal extent of these models are unknown. This means that academics and commercial teams must base their claims on models presented in the existing literature based on limited datasets (Newing, et al., 2015; Newing, et al., 2018; Khawaldah, et al., 2012). The empirical literature that does exist suggests that spatial interaction models can achieve a 5% error rate across a few stores over an entire year (Newing, et al., 2015), or a 10% error rate across tens of stores for a single week (Waddington, et al., 2019).

However, we believe that the limited empirical evidence in the literature over large store networks, and a lack of critical engagement with the limitations such as heterogeneity of store size, consumer groups served and changes in consumer behaviour, is likely to lead to existential problems in the application of spatial interactions models in grocery retailing. Specifically, the results from the literature raises two main questions. The first is, how well do these models scale? Based on the results in the existing literature, when scaling up the model from four large format stores to sixteen (including 32 small format stores), while the average error increases only from 5% to 10%, the range of errors increases from 10-15% up to 33% (Newing, et al., 2015; Waddington, et al., 2019). Therefore, if a potential aim of the application of spatial interaction models in grocery retailing is to be able to create a national level model (Davies, et al., 2019; Beckers, et al., 2021), then it becomes necessary to examine whether the error range and average error continues to increase with the number of stores modelled. If it does so, then it leads to the further question as to what scale do these models become unable to be used in practice?

The second question in this case then is how does shopping behaviour and the accuracy of the spatial interaction model vary over a year at this scale? In Newing et al. (2015)'s analysis, the four store subset analysis examined the model performance every week over a whole year, finding that while the average error rate was within 5%, for an individual week the error could range up to +/-15% of actual revenue (Newing, et al., 2015). In contrast, the sub-regional application for 16 large stores (and 32 small format stores) only examined the errors for a single week where in which the large format store error ranged up to 30% (Waddington, et al., 2019). Thus, it could be suggested that if the same stores were examined over the entire year the average error could decrease but the range of errors overall could increase. Therefore, if the weekly errors were to scale with the number of stores modelled, then it would also be useful to examine how the average error and scale of errors varied over an entire year.

It is therefore these questions that this chapter aims to analyse:

1) How well do spatial interaction models perform at a regional scale?
2) How well do spatial interaction models perform over an entire year at a regional scale?

This analysis is facilitated by access to a spatially and temporally large dataset that allows for the evaluation of the performance of the spatial interaction model at a regional scale over an entire year. Examination of these questions is achieved firstly through the application of the non-disaggregated model presented in the previous chapter to three regional scale datasets for a single week. Then, in line with the models applied in the literature (Newing, et al., 2015; Waddington, et al., 2019), an origin disaggregated model is developed and applied to the same three regions and compared with the base model. The results from this analysis show large ranges of errors for individual stores, such that the origin-disaggregated model implementation is unable to replicate the performance seen in the previous literature. This implementation is therefore supplemented by access to drivetime, brand attractiveness and new household estimate data to align the application with those found in the previous literature (Newing, et al., 2015; Waddington, et al., 2019). Even with access to this data, the model at the regional scale is still unable to replicate the performance seen in the previous literature.

Consequently, further analysis is performed by first analysing the relationship between individual store errors and store level characteristics. This includes individual store attributes, store level revenue and the surrounding area attributes. No clear relationship could be identified from this exploration, thus the performance of the model across the entire year is examined relative to the individual week chosen. From this it is identified that while the mean store error varies considerably across the year for one region, for the other two the single week results are in line with those seen across the rest of the year. This therefore suggests the results are not a consequence of factors affecting a single week. Finally therefore, the robustness of these errors is explored in response to the parameter pairs that are trained on the other two regions. With the performance indicating that the errors are not responsive to the parameters, but rather than scaling up process. The main contribution of this chapter therefore is the replication of the existing model formulation to a regional and yearly scale analysis. The aim of this is to be able to inform the debate as to whether a national scale model is a realistic goal that could be achieved for grocery retailing modelling (Beckers, et al., 2021), with the conclusion suggesting that it could not with the current model format.

5.3) Regional Level Overview

5.3.1) Value of Sales Over Distance

The city level results from the previous chapter focused on how well the model was able to perform in replicating the loyalty card flows for a single week. For the purpose of consistency, the regional level application and all subsequent applications will also explore the results of the spatial interaction model to the same week on the new datasets. For this, three consumer regions[1] were chosen to be analysed, as opposed to a single region, to ensure that the results were robust across a range of potential conditions (Newing, et al., 2015). To this end therefore, the value of sales flows over distance can be seen in Figure 8 below for all large format stores within the three regions which are examined. This figure shows a clear decrease in the sales value over distance for all three, as at the city level, therefore providing support for the idea that a spatial interaction model may still be appropriate at this scale. The main difference between this and the city level is the distance for the sales and the number of overall flows. For Region 1, the flows can be seen to extend up to a distance of 20km, whereas the city level application only saw flows up to 17.5km, and Region 2 and 3 only show sales up until around 14km and 12km respectively. This difference between the distances seen in each regions highlights potential differences in both behaviour and geography. For example, consumers in the first region may be willing, or even have to, travel further for their grocery shopping than in the other two regions. This may be influenced by factors such as distance to the closet store, limits on geography such as coasts, rivers and mountains, as well as the overall density of stores in the region relative to the density of consumers.

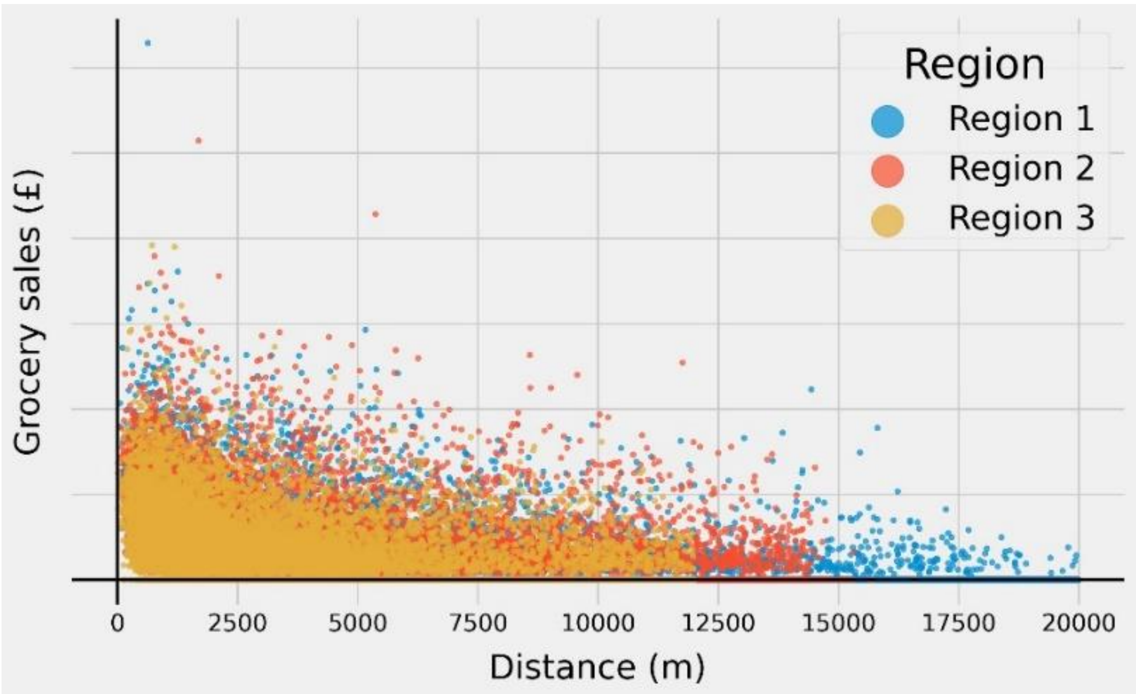

Figure 8 - Grocery sales over distance for each region

[1] Consumer regions are defined by our partner retailer based on the nine English Regions and Wales as can be seen in Figure 45 in Appendix A on page 240.

5.3.2) Disaggregating Customer Origins by Geodemographic Typology Type

This plot also shows that there are a much greater number of origins in the regional datasets than at the city level. This is to be expected due to both the increase in the number of stores (from 7 to 29, 47 and 60 respectively), alongside the exclusive focus on large format stores that have more revenue and consumers are more likely to travel further to. To this end, for each region, the number of output areas, and to which output area supergroup they belong to, can be seen in Table 5 below. This shows both the number of output areas seen in the anonymised loyalty card data and the total number of output areas from which revenue could reasonably be expected to flow to the stores. This latter count was derived in the same way as at the city level, by first identifying the distance to which 97.5% of loyalty card customers were willing travel to stores in each region, given as the crow flies distance. This distance was then used to draw a radius around each of our partner organisations stores, from which each output area within each store circumference/catchment was determined as being a potential origin that could potentially visit that store. This meant that for Region 1, while there were 3449 origins seen in the anonymised loyalty card data, this was then extended to 5700 output areas. The total count of output areas varies between regions both because of the number of stores and also because of the different distance that consumers were willing to travel.

From this table it can also be seen that the distribution of output areas within each output area supergroup varies between regions. Notably, while Region 1 is dominated by the Rural Residents supergroup, Regions 2 and 3 have greater representations of Hard Pressed Living and Suburbanites[2]. This may therefore influence differences in region wide behaviour and explain the distances over which consumers are willing to travel. In this case, Rural Residents, given they are living within a rural area, may be willing to travel further distances to shop than either Hard Pressed Living or Suburbanites who are found in or around cities. In contrast, all three regions have relatively low representation of Cosmopolitans or Ethnicity Central supergroups, suggesting that these regions, or target consumers, are not that diverse. Nevertheless the percentage of each output area supergroup in each region does not vary considerably between the loyalty and total revenue estimates. This suggests that the consumers seen in the anonymised loyalty card data are representative of the surrounding areas and the overall distribution of potential customers in each region.

[2] Descriptions of these supergroups can be found at the ONS (2015)

Table 5 - Output area count for each region in terms of output areas in the *anonymised* loyalty card dataset and total flow dataset across the eight different output area supergroups

Supergroup	**Region 1**		**Region 2**		**Region 3**	
	Total	Loyalty Card	Total	Loyalty Card	Total	Loyalty Card
1 (Rural Residents)	1775 (31.14%)	1291 (37.43%)	1204 (13.86%)	1044 (15.34%)	1664 (9.35%)	1056 (13.15%)
2 (Cosmopolitans)	197 (3.45%)	15 (0.43%)	257 (2.96%)	116 (1.70%)	579 (3.22%)	115 (1.43%)
3 (Ethnicity Central)	15 (0.26%)	0 (0.00%)	64 (0.73%)	16 (0.24%)	0 (0.00%)	0 (0.00%)
4 (Multicultural Metropolitans)	4 (0.07%)	2 (0.06%)	286 (3.29%)	170 (2.50%)	1882 (10.47%)	712 (8.87%)
5 (Urbanites)	1253 (21.98%)	643 (18.64%)	1102 (12.68%)	882 (12.96%)	2837 (15.78%)	1273 (15.86%)
6 (Suburbanites)	808 (14.18%)	527 (15.28%)	1950 (22.44%)	1638 (24.07%)	4401 (24.47%)	2197 (27.35%)
7 (Constrained City Dwellers)	533 (9.35%)	237 (6.87%)	745 (8.57%)	505 (7.42%)	1755 (9.76%)	597 (7.44%)
8 (Hard-Pressed Living)	1115 (19.56%)	734 (21.28%)	3082 (35.47%)	2433 (35.75%)	4864 (27.05%)	2079 (25.89%)
Total	**5700**	**3449**	**8690**	**6804**	**17982**	**8029**

5.3.3) The Relationship Between Loyalty Card and Total Revenue

For this application, the main aim of the spatial interaction model is to be able to predict total store revenue. This is because while the model is trained on loyalty card flows, loyalty card data accounts for anywhere between 5-70% of total revenue, varying by area, time of year and store format (Rains & Longley, 2021). Thus, if loyalty card data does not represent the total store revenue, then the model has to be able to estimate total revenue flows. For this the relationship between loyalty card store revenue and total store revenue is presented in Figure 9 below. From this it can be seen that while the strength of the relationship varies between regions, in general there are strong positive correlations between loyalty card grocery sales and total grocery sales. This therefore supports the idea that if the spatial interaction models are able to replicate total store loyalty card data well, they should be able to predict total store revenue as well.

Figure 9 - Total grocery sales against total loyalty card grocery sales for each store across each region along with the Pearson correlation coefficient

5.4) Regional Application Analysis

5.4.1) Base Model

The first step then is to apply the base spatial interaction model to all three regions. The purpose of this is to see how the same model that is applied at the city level scale performs at the regional level. As before, this model takes the form of the production constrained spatial interaction model, also known as the "retail model", which constrains the total amount of outflow from each origin. This then takes the form:

$$T_{ij} = A_i O_i W_j^{\gamma} e^{-\beta d_{ij}} \qquad \text{Eq. 47}$$

Where:

$$A_i = \frac{1}{\sum_j W_j^{\gamma} e^{-\beta d_{ij}}} \qquad \text{Eq. 48}$$

Within which T_{ij} represents the flow of expenditure (pounds sterling) from origin, i, to destination, j, A_i is the balancing factor that ensures that the origin constraints are met, O_i is the estimated expenditure available from each origin[3], w_j is the measure of store attractiveness and d_{ij} is the as the crow flies distance between origin, i, and destination, j. From this γ and β are the two parameters that are estimated in model calibration, representing the strength of attractiveness and distance decay from the anonymised loyalty card data. Here, for consistency with previous papers (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021), the exponential form of the distance decay relationship is examined[4].

For this application, as with the city level model, Poisson regression was used to calibrate the model parameters from the anonymised loyalty card data. The estimates from this regression were first used to model loyalty card flows from each origin to the stores with the total then being aggregated together at each individual store. These parameters were then used to predict the total revenue flow from each origin by using the estimated revenue available from each origin and the potential stores which they could reasonably be expected to visit. The latter is estimated by using the same method as determining potential origins, but this time from each origin to determine potential stores within the given circumference. The predicted total revenue flows were then aggregated at the individual store level in order to predict total store revenue. The results of this can be seen in both Table 6 below in terms of the metrics related to the flows from origins to destinations, and Figure 10 below which shows the errors for the store revenue aggregations for both the loyalty card and total revenue.

Table 6 - Performance metrics for the non-disaggregated base model for each region in comparison to the loyalty card flows

Metrics	**Region 1**		**Region 2**		**Region 3**	
	Loyalty card	Total Revenue	Loyalty card	Total Revenue	Loyalty card	Total Revenue
R^2	0.947	0.764	0.947	0.701	0.963	0.713
Pseudo R^2	0.932	N/A	0.919	N/A	0.957	N/A
RMSE	170.64	939.25	152.20	1186.60	81.83	1057.54
SRMSE	0.562	3.102	0.601	4.688	0.661	8.542
ATD	1.000	1.073	1.000	1.121	1.000	1.057

[3] Calculated using Eq. 44 as in the previous chapter.
[4] The inverse power distance decay was also examined, as can be seen in Appendix B, but due to the relatively small differences in model performance between the two, only the exponential form of the model is discussed here.

MAE	62.835	390.196	54.281	466.782	17.801	362.443
AIC	1356964	N/A	30743446	N/A	1830943	N/A
SSI	0.206	0.155	0.176	0.121	0.113	0.064
CPC	0.896	0.595	0.893	0.504	0.928	0.397

The first thing to note here is that in Table 6 the R^2 value and the ATD value for each region for the loyalty card flows show better values (a higher R^2 and an ATD closer to 1) than for the initial city level model. This suggests that, despite these models being applied to regional dataset, with more stores and output areas, the modelled loyalty card flows better replicate the underlying data than at the city level. This is likely to be because, although there is a greater range of potential flows, these are only for large formats stores. Thus, this suggests that flows to large stores only, even when there are more stores, are easier to model and replicate with a spatial interaction model than a smaller application with a variety of formats. This is therefore in line with previous research that suggests that shopping in small format grocery stores is different than shopping at large format stores (Waddington, et al., 2018; Waddington, et al., 2019). However, when scaling up the flows to model total revenue as opposed to loyalty card flows, it can be seen that there is a drop in the accuracy of all metrics relative to the underlying anonymised loyalty card data. This is because the estimates are scaled up because loyalty card is only able to represent a subset of the total flows. Thus, total revenue flows from origins to destinations have to be modelled. What is important however is that the ATD metric increases from close to 1, to marginally higher, suggesting that more revenue is being drawn from further away, in line with increase the total amount of revenue to stores.

This is then translated into store revenue for both the loyalty card and total revenue flows by aggregating the flows at the store level, for which the model performance can be seen in Figure 10 below. From this figure it clearly be seen that while the model in each region performs well at modelling total store loyalty card revenue, this performance does not translate to the prediction of total store revenue. Indeed, for each region the average error for the loyalty card model ranges from 0.85% to 1.62%, while this average error range increases to -3.59% to 15.89%. This can also be seen in the range of overall errors where the majority of loyalty card errors cluster between +/-10% for all three regions, ranging in total from -30% to just over 40%, while the total revenue errors mostly cluster within +/-25% and range from 60% underprediction to 120% overprediction. This can clearly be seen in the distributions along the x and y-axis of the plot whereby the loyalty card errors are consistent and clustered centrally, while the total revenue errors are more widely distributed. This therefore suggests that while the models are able to replicate loyalty card revenue well at this scale, they are not able to model total store revenue consistently. These results are therefore not in line with those seen previously in the literature (Newing, et al., 2015; Waddington, et al., 2019) and thus suggest that the model does not scale well.

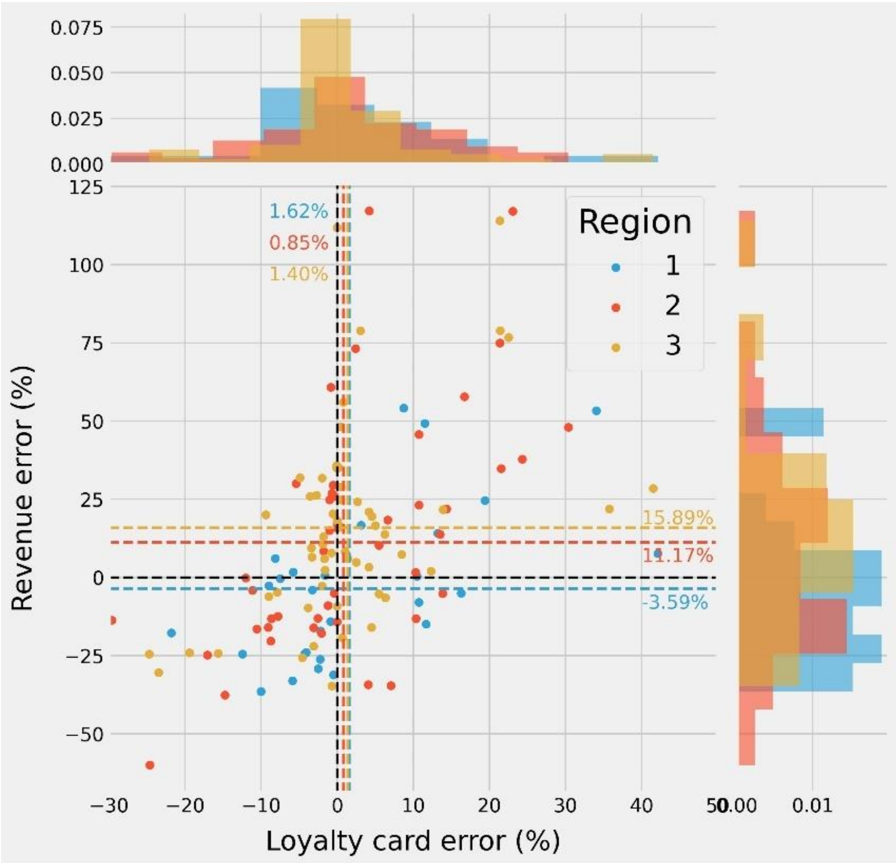

Figure 10 - Results from the non-disaggregated model application across all three regions in terms of the store level errors for both the loyalty card and the total revenue in terms of the percentage error

5.4.2) Origin Disaggregated Model

However, the results presented above are produced using the base spatial interaction model where the parameters for both the attractiveness and distance decay are the same for all origins and destinations. This is unlikely to produce the most accurate spatial interaction model because of different socioeconomic customer groups and the associated non-stationary spatial relationships that would influence the values of both of these parameters. For example, the distance decay relationship for a low socio-economic group in a rural area may be expected to be different to a similar group in an urban area due to differences in transport accessibility. To rectify this an origin disaggregated spatial interaction model can be implemented (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021), which allows for the attractiveness and distance decay parameters to be calibrated across different socio-demographic classifications. In this thesis, this is done across the eight different Output Area Supergroup classifications (Gale, et al., 2016; Waddington, et al., 2019)[5]. Due to differences in socioeconomic conditions, these groups are expected to have different propensities to shop at our partner retailer, due to factors such as price and quality of products, alongside having different willingness or ability to travel, due to access to different resources and means of transportation (Birkin, et al., 2017). Such a model could not be implemented at the city level due to limited loyalty card representation for each supergroup which would have led to inaccurate and inconsistent estimates of parameters values. This origin disaggregated model therefore takes the model form:

$$T_{ij}^{k} = A_{i}^{k} O_{i}^{kt} \mathrm{W}_{\mathrm{j}}^{\gamma^{\mathrm{k}}} \exp\left(-\beta^{k} c_{ij}\right) \qquad \textit{Eq. 49}$$

Where

$$A_{i}^{k} = \frac{1}{\sum_{j} W_{j}^{\gamma^{k}} e^{-\beta^{k} d_{ij}}} \qquad \textit{Eq. 50}$$

With each parameter and variable taking the same interpretation as before, but now each *k* represents the output area classification supergroup for which each γ and β are calibrated separately.

An issue with this implementation, however, is that the coverage for certain output area supergroups is limited in the anonymised loyalty card data. This can be seen in Table 5 above, as already discussed in section 5.3, where there are limited numbers of output areas within supergroups 2, 3 and 4. This means that for most weeks across each region there is not enough data for the parameters to be accurately or reliably calibrated to generate reasonable estimates. Thus, to ensure that some flows are still estimated for these supergroups and output areas, if the parameters cannot be calibrated then the system wide parameters from the base model are assumed to apply when modelling total revenue[6]. While the number of output areas for these supergroups is not small enough to assume that they will generate zero flows to the stores, their number is small enough such that attributing value from these origins using the system wide parameters is not expected to considerably affect the final model performance.

Then, as with the base model, the results in terms of replicating the loyalty card flows and the total store revenue can be seen in Table 7 below and Figure 11 below. What is interesting to see however

[5] The OAC is used the classification for different demographic groups due to its use in previous literature, the ability to link estimated expenditure for each group, and that the classification was freely available.
[6] The parameter values for both the system wide and disaggregated model can be seen in Appendix A in Figure 46 and Table 14.

is that while most metrics show improvements relative to the base model for all three regions, as highlighted by those in green, these improvements are only relatively small. Indeed the comparison of most metrics from the disaggregated model to the base model (those in brackets) are smaller than the comparison of metrics between regions. This therefore suggest incremental improvements in being able to model the underlying loyalty card flows from the application of the disaggregated model.

Table 7 - Performance metrics for the origin disaggregated model for each region in comparison the actual *anonymised* loyalty card data (and the metrics from the base model for each region)

* Disaggregated metrics reproduced in *Appendix A Table 12*

Metrics	**Region 1**		**Region 2**		**Region 3**	
	Loyalty card	Total Revenue	Loyalty card	Total Revenue	Loyalty card	Total Revenue
R^2	0.949 (0.947)	0.769 (0.764)	0.952 (0.947)	0.706 (0.701)	0.966 (0.963)	0.671 (0.713)
Pseudo R^2	*	N/A	*	N/A	*	N/A
RMSE	166.513 (170.638)	934.345 (939.245)	145.190 (152.203)	1182.189 (1186.601)	78.833 (81.831)	1125.172 (1057.542)
SRMSE	0.550 (0.562)	3.086 (3.102)	0.574 (0.601)	4.671 (4.688)	0.637 (0.661)	9.088 (8.542)
ATD	1.002 (1.000)	1.074 (1.073)	1.000 (1.000)	1.131 (1.121)	1.000 (1.000)	1.023 (1.057)
AIC	*	N/A	*	N/A	*	N/A
MAE	61.746 (62.835)	385.061 (390.196)	51.406 (54.281)	464.558 (466.782)	16.618 (17.801)	366.267 (362.443)
SSI	NaN (0.206)	0.155 (0.155)	0.177 (0.176)	0.122 (0.121)	0.114 (0.113)	0.064 (0.064)
CPC	0.898 (0.896)	0.598 (0.595)	0.898 (0.893)	0.506 (0.504)	0.933 (0.928)	0.394 (0.397)

These results then also translate to the modelling of total loyalty card revenue and total store revenue as seen in Figure 11 below as well. From this figure, it can be seen that while there may be some differences in the average errors for both scales of revenue prediction and for individual stores, these improvements or changes are relatively small compared to the size of the errors seen in the base model. This is highlighted by the fact that the range of errors within each region has not considerably improved as the result of the application of the origin disaggregated model as would be expected if the base model was unable to reflect the underlying behaviour of each group within the region. Therefore, the development and the application of the origin disaggregated model, with the current data and methodology, has been unable to resolve the issue of the large errors in total revenue prediction at the regional scale.

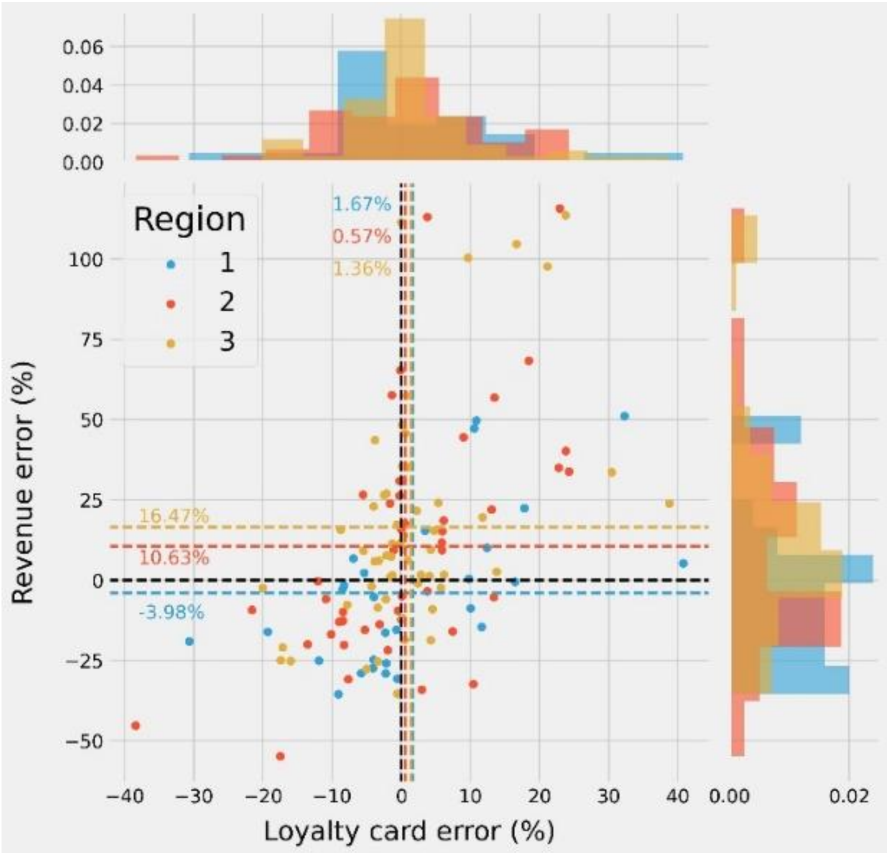

Figure 11 - Results from the origin disaggregated spatial interaction model across all three regions in terms of the individual and average store errors for both the loyalty card and total revenue predictions

To this end, the relative comparison between the base and the origin-disaggregated model, in terms of their ability to model total store revenue, can be seen in Figure 12 below. This figure highlights that while there are differences in mean store error for each region, these changes are minor adjustments relative to the scale of mean error to 0. In this sense, the largest difference between each model mean is only 0.58%, which is relatively small compared to the overall range of model errors. It can also be seen that for most stores in all three regions, the store error from the base model is reflected in the error for the origin disaggregated model. This is shown by the close to linear relationship between the two model outcomes for each region and the similar distributions on the x and y-axis. Therefore it is clear that, despite expectations, the application of the disaggregated model at the regional scale has not lead to considerable improvements in the ability of the spatial interaction model to predict total revenue at this scale.

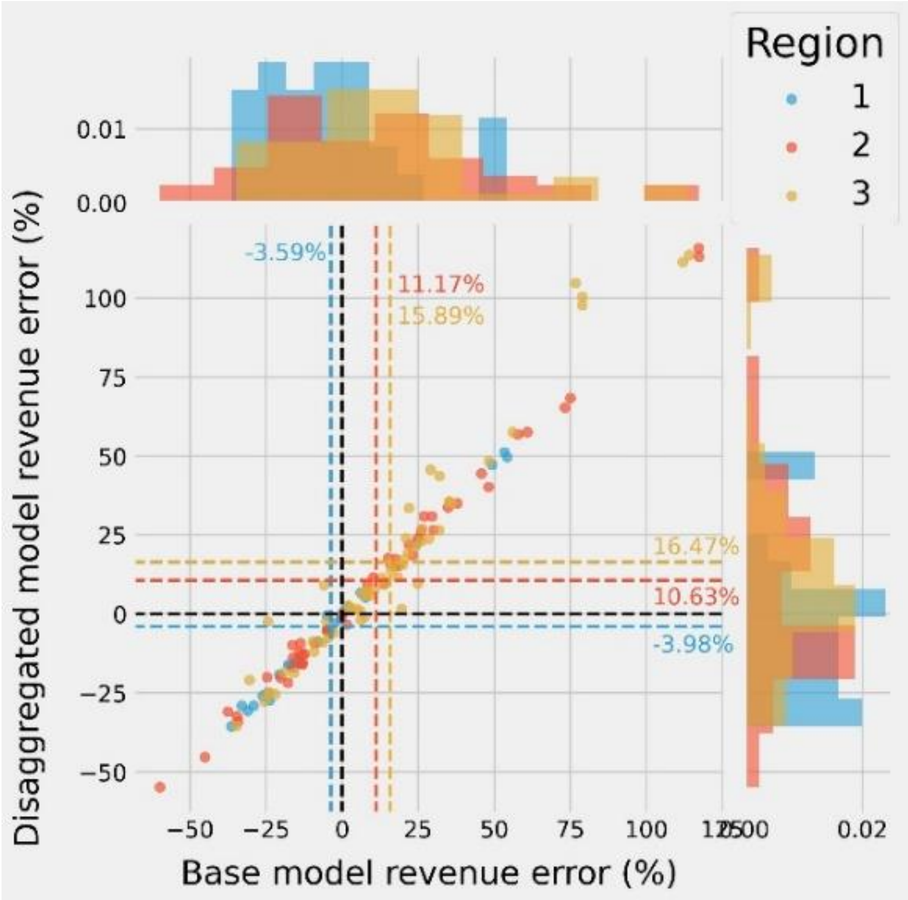

Figure 12 - Base model and origin disaggregated model errors against each other for all three regions

5.4.3) Additional Data Integration

From these results it is clear that while the spatial interaction models for each region can replicate store loyalty card well, they are currently unable to accurately and reliably estimate total store revenue at this scale. While for the most part the average store error and range of errors for replicating loyalty card revenue are in line with previous results (Newing, et al., 2015; Waddington, et al., 2018), the range of total store revenue errors are considerably larger than expected and those presented in the previous literature. This range of errors is consistent across all three regions therefore suggesting that this is not due to factors that may only be affecting a single regions performance. However, it is worth acknowledging that these model implementations have not accounted for variables that have been integrated into recent applications of the models in the literature which may be affecting the errors (Newing, et al., 2015; Waddington, et al., 2019).

5.4.3.1) Travel Time

The first of these factors that have not been implemented in the models presented above is the influence of travel time. So far, all the models implemented have been used as the crow flies distance measured from the centre of the output area to the centre of the store. This may potentially affect the performance of the model on a store level as for example, a store that is accessible only through a North-South corridor but which under the current implementation is drawing revenue from East-West could potentially be overpredicted. In a similar vein, a store that is highly accessible by car, such as along a major road running alongside the stores entrance, would likely be underpredicted when using as the crow flies distance as compared to other stores households may be able to travel further, quicker, than they would in a city. Therefore, drivetime data was brought in to potentially rectify for this a single region using the Open Source Routing Machine (OSRM) API (Giraud, et al., 2021). This was used to estimate the average travel time a car would take from the centre of each output area to each store they could potentially visit.

5.4.3.2) Customer/Retailer Preference

Secondly, while the parameters in the model are calibrated using data from our partner retailer for each origin output area supergroup, these parameters are not adjusted to account for each group's preference for different destination retailers (Newing, et al., 2015). This could affect the model's performance for individual stores as in some cases they may be located in areas whose surrounding demographics prefer to shop at competing retailers or prefer our own partner organisation. In the former case, the total revenue would be overpredicted for that store as consumers would prefer another local store, while in the latter case revenue would be underpredicted because our partner organisation would be more attractive than other surrounding stores. For this then, data on retailer preference could be adapted from the paper by Newing et al. (2015), originally from Thompson et al. (2012), as seen in Figure 13 below. These values represent the relative attractiveness of each retailer for each output area supergroup from the 2001 Output Area Classification (Vickers & Rees, 2007). To apply this in our current model, which used the 2011 Output Area Classification, the 2001 Output Area Classification was linked to the 2001 output area dataset. A lookup table provided by the ONS was then used to link the 2001 output area codes to their 2011 equivalents (ONS, 2020). This then allowed the values from the figure below to scale the attractiveness parameter for competing stores relative to the values calculated for our partner organisation.

Brand (retailer)	OAC supergroup						
	1	2	3	4	5	6	7
	Blue collar	City living	Countryside	Prospering suburbs	Constrained by circumstances	Typical traits	Multicultural
Aldi	0.9980	0.9970	1.0051	0.9987	1.0025	1.0005	0.9952
ASDA	1.0076	0.9912	0.9904	0.9970	1.0023	0.9992	1.0013
Co-Op	1.0020	0.9990	1.0157	0.9922	1.0008	1.0000	0.9894
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	1.0020	1.0005	0.9990
Sainsbury's	0.9904	1.0121	1.0013	1.0088	0.9942	1.0028	0.9997
Tesco	0.9992	0.9987	1.0071	1.0010	0.9965	0.9990	0.9985
Waitrose	0.9811	1.1000	1.0061	1.0124	0.9843	1.0023	1.0068
Iceland	0.9997	0.9982	1.0058	0.9975	0.9991	1.0001	1.0021

Figure 13 - Relative attractiveness of retailers for each output area supergroup from Newing et al. (2015) Table 1 Pg 227

5.4.3.3) Estimating Origin Grocery Expenditure

Finally, while the anonymised loyalty card data for this model comes from 2017, the number of households in each origin that is used to calculate the estimated available expenditure comes from the 2011 census. This could therefore affect the models performance for individual stores in areas where population has changed during that 6 year gap. This is expected to be especially important for new stores who opened up where populations could have increased due to new construction, or in places where population has decreased such as due to an ageing population or migration. To this end, there are no yearly updated or estimates of the number of households in each output area. However, there are yearly estimates of the output area population, as provided by the ONS. This can therefore be used to estimate the number of households in each output area in 2017. This is done by firstly dividing the 2011 output area population by the number of independent households from the census estimates to get the population per household, and then dividing the 2017 population by the estimated population per household[7]. Thus, leading to an updated estimate of the total number of households in each output area in 2017 and hence an updated estimate of the revenue available from each origin.

These additions can therefore be integrated into the existing model formulation for the origin disaggregated model. The adjusted model formulation then becomes:

$$T_{ij}^{kn} = A_i^k O_i^{kt} \mathrm{W}_{\mathrm{j}}^{\gamma^{\mathrm{kn}}} \exp\left(-\beta^k c_{ij}\right) \qquad \textit{Eq. 51}$$

[7] $h_{2017} = pop_{2017}/(\frac{pop_{2011}}{h_{2011}})$ where *h* = households, pop = population

Where:

$$A_i^k = \frac{1}{\sum_j W_j^{\gamma^{kn}} e^{-\beta^k d_{ij}}} \qquad \text{Eq. 52}$$

And:

$$O_i^{kt} = e^{kt} h_i^{kt} \qquad \text{Eq. 53}$$

Within which the parameters and their interpretation are the same as they were before, but now *n* represents the influence of brand on the attractiveness parameter by output area supergroup, and *t* represents 2017 for both expenditure estimates and the number of households. This adjusted model implementation therefore reflects the latest advancements in the existing literature (Newing, et al., 2015; Waddington, et al., 2019).

5.4.3.4) Implementation in the Model

The results from this model adaptation can be seen in Figure 14 below which shows the change in individual stores performance in the model from the addition of new datasets in Region 2. This includes the replacement of as the crow flies distance with drivetime, then the addition of the relative attractiveness of different retailers with drivetime, followed by the updating of the number of households with drivetime data, and finally a model that integrates all three additional datasets into the model. There are three main takeaways from these results. The first is that moving from as the crow flies distance to the drivetime model results in an increase in the average store error from 10.39% to 13.49%. This would therefore suggest that the stores from our partner retailer are more accessible relative to their competition by car than they are by as the crow flies distance. This is because, with total available revenue staying the same, the amount of revenue assigned to our partner retailers stores has increased on average. The second key insight is that drivetime has the biggest influence on average store error, and indeed for most stores this causes the greater change in error, as opposed to adjusting the relative attractiveness by brand or updating the estimated number of households. This is then closely followed by the number of households, with the adjustment of relative brand attractiveness having the least overall effect on model performance. Finally, the main finding from this model adjustment is that even adding in these new data sources has not resolved the overall issue of poor model performance in terms of replicating total store revenue. In this case, the complete model even moves the average error further away from the ideal 0% and shifts the distribution towards the higher percentage errors. Therefore, even the latest model from the literature is unable to replicate total store revenue at this scale and has not resolved the issues from the base model presented above.

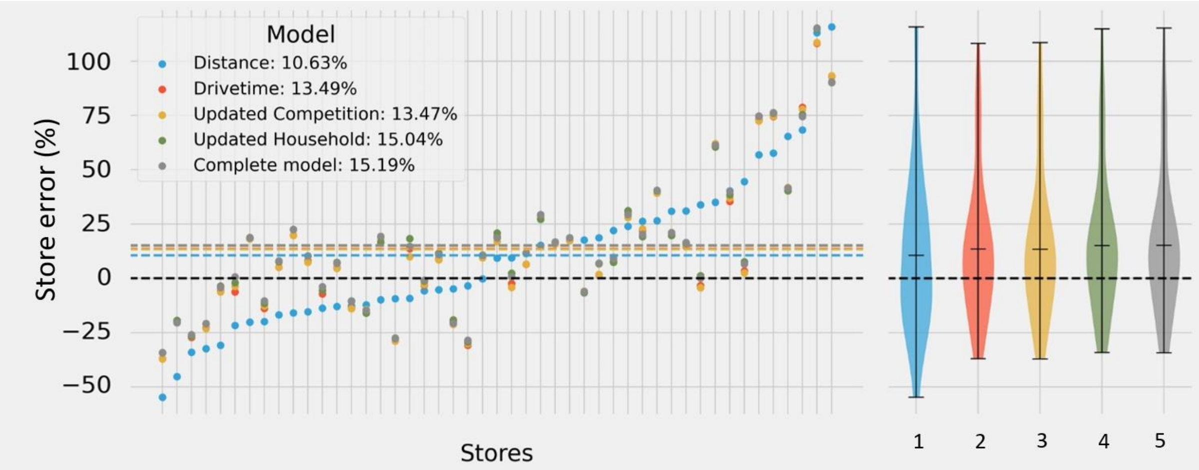

Figure 14 - Individual store error and the overall distribution in region 2 in response to additions of new data. 1) The origin disaggregated model as already presented, 2) The origin disaggregated model with drivetime data between origins and destinations, 3) The origin disaggregated model with drivetime and updated attractiveness values to account for competition, 4) The origin disaggregated model with drivetime and updated household count, 5) the origin disaggregated model with all new datasets integrated.

These results therefore suggest that the current model implementations are unable to account for the variation in underlying store conditions across an entire region when scaled up to predict total store revenue. In all model implementations presented above (Figure 7), the average store error for the anonymised loyalty card data is within +/-2%, with most store errors within +/-20% of actual revenue, which would be in line with the results with previous model implementations (Newing, et al., 2015; Waddington, et al., 2019). However, since loyalty card revenue accounts for between 5 and 70% of total store revenue for each store (Rains & Longley, 2021), these estimates have to be scaled up to model total store revenue. In doing so while each of these model implementations across each region has an average error within +/-15% of total store revenue, the range of store errors is considerably larger. This is such that less than half of the stores for each region fall within a +/-15% error band, with ranges of up to 60% underprediction to 120% overprediction. This range would therefore render these models effectively useless in practice as retailers would not be confident in choosing a store location when the actual revenue is likely to be at least 15% different to the actual predicted revenue from these models (Newing, et al., 2015; Waddington, et al., 2019).

5.5) Factors Affecting Store Performance

The results in the previous section therefore suggest that at the regional level the spatial interaction model in its current format and the current inputs into that more are unable to accurately or consistently scale up to predict store total revenue. This is consistent across all three regions and all model implementations that were developed. Thus it is worth exploring whether there are any store characteristics or underlying conditions that are not captured within the current spatial interaction model that could potentially be influencing these results or are related to model performance. These potential factors can be broken down into three categories of: individual store characteristics, store revenue attributes and surrounding area characteristics.

The first of these sets of factors are individual store characteristics that could influence how attractive an individual store is. This includes net store size, total indoor floorspace, gross store size, total store space including external spaces and car parking, and the age of the store in months. While net store size is included in the current model as a measure of attractiveness, this is a measure of total floorspace as opposed to the floorspace dedicated to grocery retailing. Furthermore, store attractiveness is also likely to be influenced by a variety of factors beyond just store size (Newing, et al., 2020), such as distance to the street, store frontage, other services located in the store and age (Birkin, et al., 2017). To this end, data was only available on gross square footage and store age, for which any relationship these variables have with store performance may indicate that there may be a more appropriate measure of store attractiveness than is currently used. This is analysed alongside store revenue attributes which consists of: total store sales, total grocery sales, the percentage of total sales that are grocery sales, total loyalty grocery sales and the total number of baskets. These factors are related to what the model is trying to predict, in terms of total store grocery sales, and represent the way in which a store receives its revenue. Thus, any relationship with these variables may suggest a bias in model performance towards well performing or underperforming stores alongside how a store may be deriving their revenue. Indeed, if any relationship is discovered between store error and the total grocery revenue then this would suggest that there is bias in the final model estimation of revenue. Finally, there is also surrounding area characteristics such as the number of output areas from which revenue is expected to come from and the average distance to these output areas. These characteristics indicate the level of density surrounding a store and the concentration of potential revenue around the store. Thus, any relationship with these variables may indicate that a stores immediate geography is not accounted for in the model implementation, including whether a store locates in a dense or sparse area such as a store targeted towards urban or rural populations.

The correlations between the individual store errors and store level characteristics for each region and model can be seen in Table 8 below, while the individual relationships for each store can be seen in the scatter plots in Figure 15 below. Firstly it can be seen that the correlations between store characteristics and model errors vary across both model implementation and regions. In this case, the variation in correlation between regions is greater than the variation within region, which is in line with the results presented in the previous sections whereby the base and the origin-disaggregated model produce similar results in terms of total store revenue errors. In terms of the between region changes in correlation, while for some attributes the relationships do not change much, for example store size (sqft), others change considerably such as for total store sales. While the former may be expected because the models are trained and implemented using this attribute, the latter suggests that across the regions there may be different underlying conditions that affect individual store performance. However, what can be seen is that there are very few strong correlations between any characteristics and store performance. Indeed, the strongest correlation is that of 0.3685 for the base model implementation in region 2 with the age of the store in months. This characteristic is also the only characteristic for which there is a consistent correlation across all model implementations.

Table 8 - Pearson correlation statistic between store characteristics and model errors

Correlate		Region 1		Region 2			Region 3	
		Base model	Disaggregated model	Base model	Disaggregated model	Complete model	Base model	Disaggregated model
Store characteristics	Store size (sqft)	0.0300	-0.0016	-0.0488	-0.0231	-0.1376	-0.0292	0.0530
	Gross store size (sqft)	0.0575	0.0245	-0.0864	-0.0631	-0.1692	-0.0733	0.0070
	Age of store (months)	-0.2171	-0.2307	-0.3685	-0.3584	-0.3451	-0.2764	-0.2710
Store revenue	Total store sales	-0.0072	-0.0448	-0.1603	-0.1307	-0.1359	-0.2620	-0.2198

	Total Grocery sales	-0.0580	-0.0953	-0.1742	-0.1451	-0.1331	-0.2979	-0.2537
	Total loyalty card grocery sales	-0.0352	-0.0648	-0.1167	-0.0900	-0.0236	-0.3227	-0.3337
	Percentage of sales of grocery revenue	-0.2308	-0.2130	-0.0790	-0.0869	0.0718	-0.2867	-0.2748
	Total number of baskets	-0.0319	-0.0676	-0.1719	-0.1372	-0.1195	-0.1346	-0.0826
Surrounding area	Number of output areas	0.1329	0.1011	-0.2182	-0.2225	-0.2065	0.0660	0.1312
	Average distance to output areas	-0.0322	-0.0152	0.2595	0.2480	0.3347	0.0915	0.0613

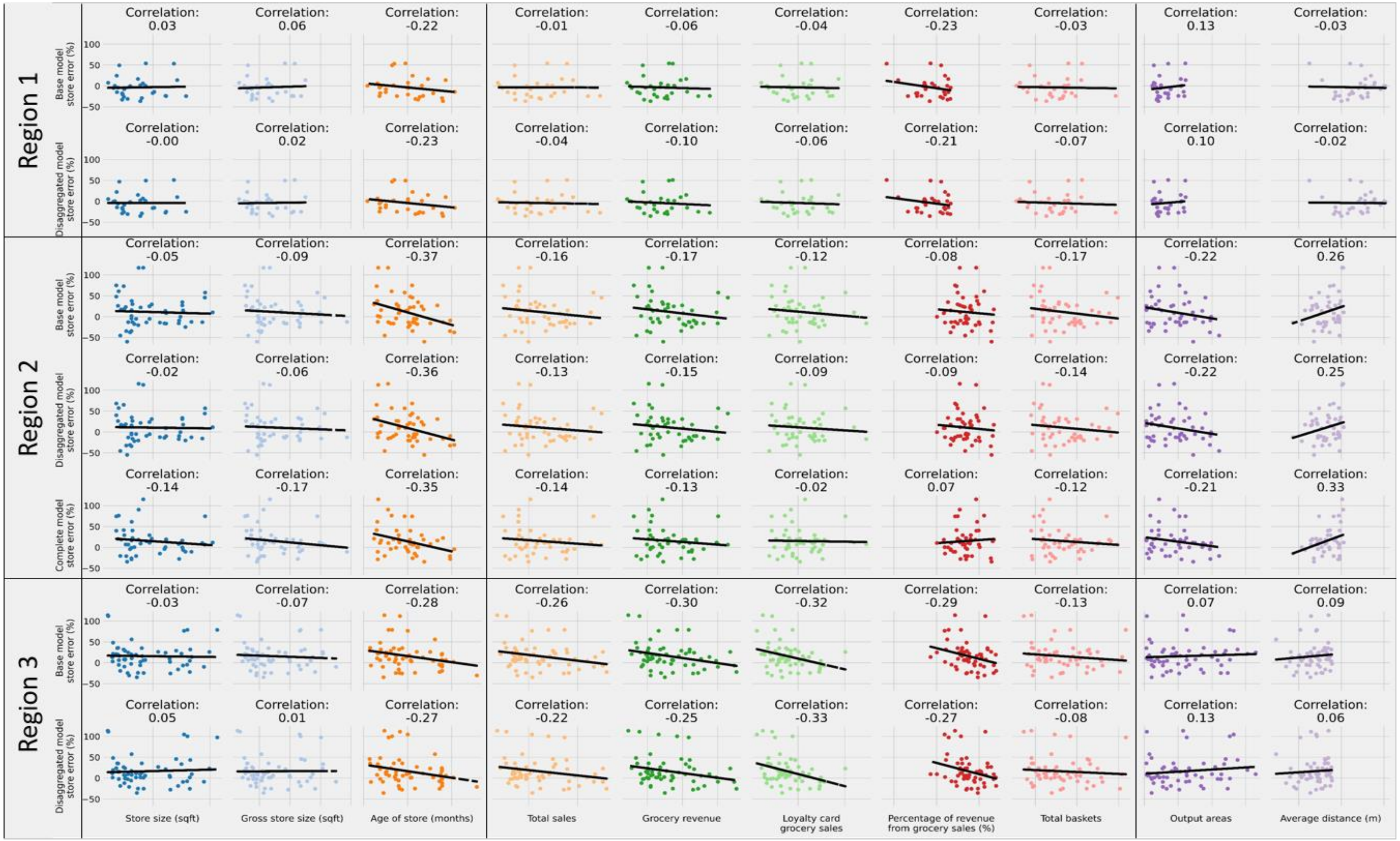

Figure 15 - Scatter plot of model errors and store level characteristics including the line of best fit and the Pearson correlation statistic

For this then, the underlying relationship between each models errors and the characteristics can be seen in Figure 15 above. From this, in conjunction with the results in Table 8, it appears that the only consistent characteristics which are correlated with store errors across all regions and model implementations is that of the age of the store, total store sales and total grocery sales. In terms of the latter two variables, it can be seen from the figure that while there are consistent correlations across all model implementations and regions, there is also consistent variation around the line of best fit. This therefore suggests that the relationships are not clear, but could indicate that the model is failing to account for some store attribute that may be influencing model performance. Indeed, the model is aiming to predict grocery revenue therefore any relationship here is likely to indicate poor model performance or bias in model prediction. Nevertheless, the relationship between store errors and the age of stores is consistent and clear for all regions and models. This shows a negative correlation between the age of stores and the individual store error, suggesting that an older store in our model tends towards underprediction, while a younger store tends towards overprediction. This therefore may suggest that a younger store may be less attractive than just its store size may suggest while an older store may be more attractive because it has been a part of the local community for longer. This factor is however unaccounted for in our model. Unfortunately consistent data for competitors stores was not available for this and so could not be integrated into the measure of attractiveness[8]. Thus, leaving this question open for potential future research.

5.6) Yearly Variation

Nevertheless, the second aim of these model implementations was to examined how the models perform across a whole year. In this case, Newing et al. (2015) examined the four store subset performance across the whole year, identifying that non-seasonal demand would influence the performance of individual stores throughout the year, while Waddington et al. (2019) was only able to implement their scaled up model on a single week. In this case therefore it is worth examining whether the modelling errors that are seen in the above analysis are the result of the model being implemented on a single week or whether these errors are consistent throughout the year. For this, each model in each region is calibrated using the weekly loyalty card flows for that individual week and is then scaled up in the same way as presented above. This therefore leaves open two main questions: how do the loyalty card errors vary over the year and how do the total revenue errors vary over the year?

[8] Further exploration of this concept is achieved in Chapter 7 Section 7.3

5.6.1) Loyalty Card Spend Annual Error

Firstly the average loyalty card store errors varying across the year for each region and each model are presented in Figure 16 below. What can be seen from this is that, despite some outliers in each region, most average errors are consistent throughout the year and are within a range of 0-2%. This would therefore suggest that behaviour and the ability of the models to predict total store loyalty card revenue is consistent with little seasonal variation. In this case the week which is analysed in the individual weekly results presented above is highlighted, along with two key seasonal periods of the Easter holidays and the school summer break as well. Within this, most weeks where there is a deviation from the underlying trend is from the respective base models where there were issues in calibration for that week due to variances in the underlying weekly flows[9]. For some weeks there is also variation in the disaggregated model performance as well, where a lack of data limited calibration for certain supergroups or there was variation in individual store performance due to local construction works or the closing of a store. However, there are few overall deviations from the trend suggesting there is consistent estimates for the anonymised loyalty card data across all regions and models.

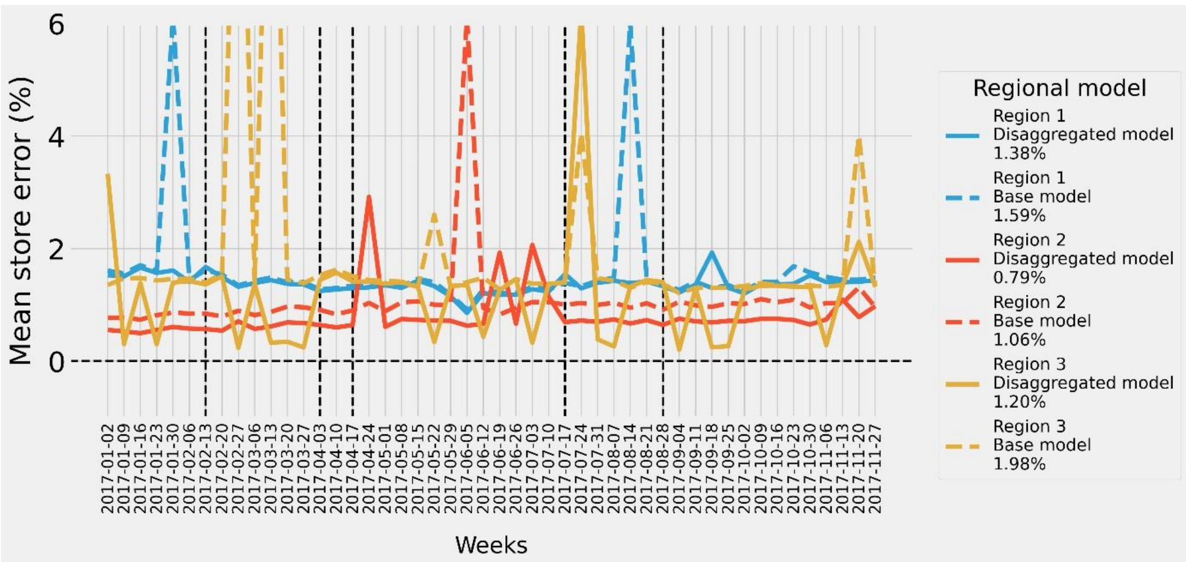

Figure 16 - Mean store error for total loyalty card revenue prediction over the year for each region and model specification

5.6.2) Total Revenue Annual Error

This can then be compared to how the total revenue errors vary across the year in Figure 17 below. As in the previous figure the week chosen for the above analysis, the weeks of the Easter holidays

[9] Calibrated parameters over the entire year can be seen in Appendix A in Figure 47.

and the weeks of summer holidays are highlighted. The main takeaway from this figure is that there is considerably more variation in average errors across the year for the total revenue model than there is for the loyalty card revenue predictions. For example, while in the loyalty card total revenue prediction there is no clear effect of the easter or summer holidays, for all three regions there is a considerable change in the average error during the easter period and a considerable change for Region 1 over the summer period. This is likely to be because of the influence of non-residential demand, such as tourist revenue, where Region 1 and 2 are noted tourist destinations in the UK, the former of which has been highlighted in previous research (Newing, et al., 2015)[10]. This is therefore consistent with the potential influence of non-residential demand which is not accounted for in these model implementations and that variation throughout the year in model prediction are driven by non-loyalty card demand. This figure also shows however that the errors seen in the models presented above are not just the result of a single week, where the week chosen for the analysis is in line with the average errors for each region and model implementation.

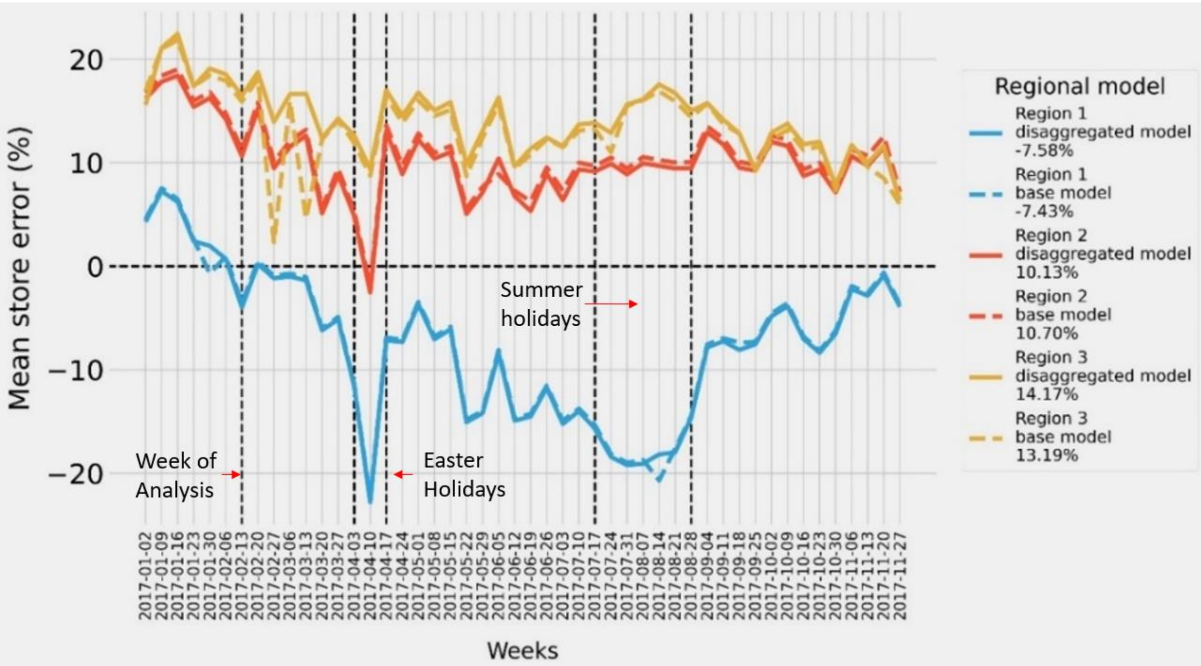

Figure 17 - Mean store total revenue errors for the base and the origin disaggregated model for each region across the year

The same yearly variation in individual store errors can also be examined for Region 2 in response to the addition of the new datasets as presented in section 5.4.2 above, as presented in Figure 18 below. The main takeaway from this is that while the new data brought into the model implementation affects the average store error, this effect is consistent throughout the year. This

[10] Tourist demand data for each region was not available due to GDPR constraints of 5 households per output area.

can clearly be seen in the figure as the separation between each model implementation is consistent and they show the same variation and patterns across the year as the simple disaggregated model which uses distance data. Thus, this is consistent with the previous findings that the addition of new datasets and the alteration of the model formulation does not resolve the issues of model errors.

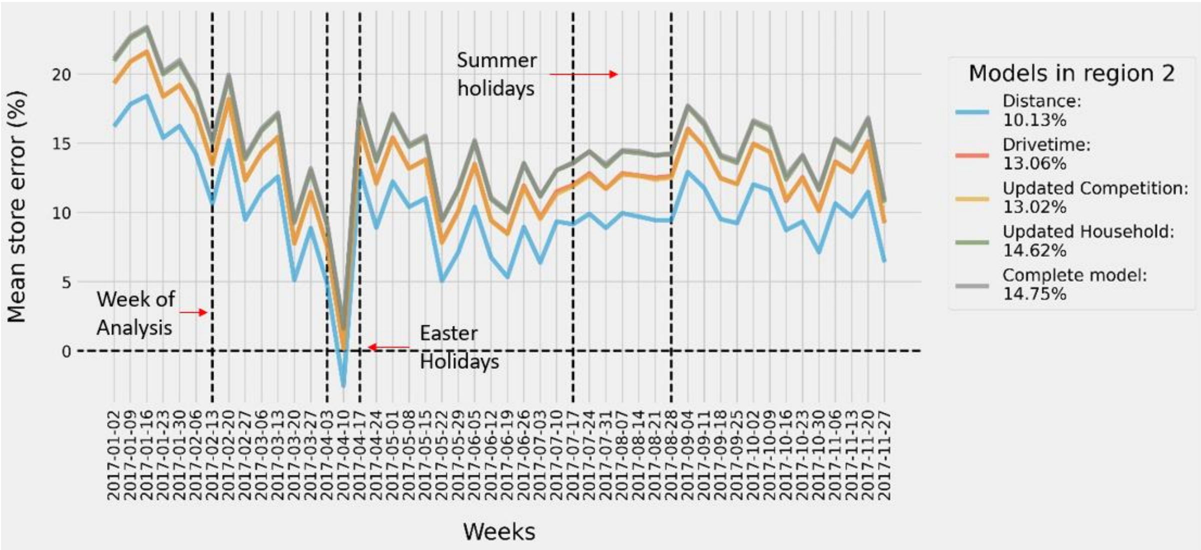

Figure 18 - Weekly average store errors in Region 2 across the different model implementations for the origin-disaggregated model

Therefore, the results that are seen for the single week exploration is not due to factors that affect that single week across each region. Indeed, while there are seasonal variations in each models average accuracy, in relation to predicting total store revenue, the average error from ach week is in line with the results presented above. It can also be seen that while there is consistent errors for the loyalty card revenue prediction, there is considerably more variation in the total store revenue errors across the region for all model implementations. This suggests that the variation in total store revenue prediction is thus driven by changes in non-loyalty card demand. In this sense, Region 1 sees the greater variation in model error across the year, which is likely to be due to changes in seasonal demand levels due to factors such as non-residential tourist demand (Newing, et al., 2015). In contrast there is less variation in the average errors for both Region 2 and 3, which are less likely to be affected by change in tourist demand, although they both see decreases in the average error over the easter holidays. Nevertheless, the consistency in model performance and identification of the influence of non-seasonal demand suggests that we can be confident in the conclusions of poor overall fit of the spatial interaction model at this level, primarily when modelling total store revenue due to non-loyalty card and seasonal demand.

5.7) Cross validation

The previous results show that there is a poor model fit at this scale which is consistent across all three regions, all model implementations across both a single week and the whole year. However, these implementations are based on training, scaling and validating the models all on the same data for each region. This means that the loyalty card flows the model trains on, includes all stores in the respective region, which are then also used in validating the total revenue model. While the wide range of errors and poor model fit suggests that these models are not overfitting on the underlying data, it would however be useful to explore how sensitive these modelling results are to variations in the parameters that are calibrated. This is achieved through cross-validation of the total revenue store errors by applying the parameters from each region to each other in modelling total predicted revenue flows. This is done for all three regions, using as the crow flies distance, for both the base and the disaggregated models. The results of which can be seen in Figure 19 below.

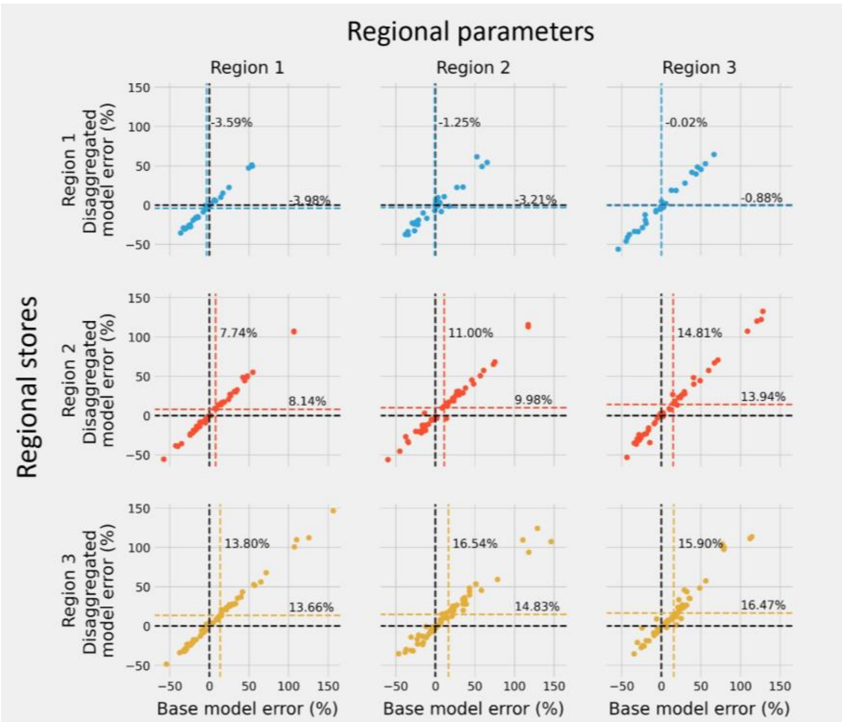

Figure 19 - Regional cross validation results with the regions trained parameters (column) against the regions stores and origins (rows) for the base model (x-axis) and the disaggregated model (y-axis)

This figure shows the results when using the parameters that are calibrated on a single region (column) against the other regions origin and destination flows (rows) for both the base model (x-axis) and the origin-disaggregated model (y-axis) for a single week. There are three main takeaways from this figure. The first is that, as with the weekly results presented in the previous sections, for most parameter-region pairs there is a strong positive correlation between the store errors for both the base and the disaggregated model. This can clearly be seen by the near perfect linear relationship for both models in all subplots and the relatively close alignment of the mean store errors. This is such that the variation between different parameter pairs on the same region, in terms of both the range of errors and the mean store error, is greater than between the base and the disaggregated model. This therefore support the previous conclusion that the origin disaggregated model does not drastically improve performance, at least with the way that it is currently implemented, at this scale, and that system wide parameters are still relevant in this implementation. This is in contrast to the results or analysis presented in the previous literature, which has suggested that the disaggregated model leads to improved model fit (Newing, et al., 2015; Waddington, et al., 2019).

The second key result is that there is greater variation in performance between each region than there is across each parameter pair within each region. This can be seen in the average store errors for both models whereby the variation is greater for the same parameter pair between each region than it is for different parameter pairs within the same region. This is surprising given that each region has different amounts of stores and distances which are applied to determine potential output areas and competition which the parameter are calibrated on. This therefore suggests that the errors for each store and region are driven primarily by the underlying variation in store and surrounding area characteristics, rather than the parameters that are calibrated on the anonymised loyalty card data.

This then leads onto the third key result which is that it appears that the errors in each region are not majorly sensitive to the parameter that are used and that in most cases the parameters that are calibrated in that region produce the most reliable results in terms of individual store errors. This therefore supports the conclusion in the previous paragraph, that the errors for each store in each region are due to the underlying regional characteristics and method of scaling up to predict total revenue, rather than the parameter pairs themselves. It also suggests that for the most part, the parameters calibrated on the region produce the best results for that region within this model. This is because for most implementations presented above, while the most accurate mean error for each region does not necessarily come from the parameters that are calibrated for that region (see region 1), the parameters calibrated on that region produce the most consistent range of errors relative to the other parameter pairs. Therefore supporting the ideas and conclusions presented above that at this scale, the spatial interaction models in their current form are not able to deal with the underlying variance in store conditions to produce reliable and usable estimates of individual store revenue.

5.8) Discussion

These results above suggest that the spatial interaction model, in its current form, with the inputs we have used and standard application, cannot be used at the scale of a whole region and be able to estimate the total revenue of large format grocery stores in the UK. To this end, an initial application of the exponential distance decay form of the non-disaggregated production constrained model showed that an acceptable range of errors and mean store error in modelling total loyalty card revenue across three regions. However, when attempting to scale up the model implementation to predict total store revenue, the mean store error increased, alongside the range of total store errors, which were not in line with the performance seen in previous applications of the model (Newing, et al., 2015; Newing, et al., 2018), and for which could not be reliably used in practice. The increase in the amount of data at the regional level however enabled the application of the origin-disaggregated form of the spatial interaction model. But the results at both the loyalty card and total store revenue scale were not that different from the system wide model results. Further attempts to improve the model fit then came from integrating drivetime, updated households and accounting for the relative attractiveness of different brands to socioeconomic groups, to align the implementation of the model with those seen in the recent literature (Newing, et al., 2015; Newing, et al., 2018). While the addition of these datasets reduced the range of errors for the region to which this was applied, the effect was only marginal and even pushed the average store error away further from the desired target of 0%. This suggested that our retailer located in areas which contained demographics that were attracted to our partner organisation over other competitors and in areas that were more accessible by car than other retailers, but did not resolve the issue of poor modelling fit at this scale.

This conclusion was then explored and tested by firstly evaluating potential factors that could be influencing or related to the individual store error for each region. This included an evaluation of individual store characteristics, store revenue levels and surrounding area conditions. To this end, the only consistent relationship between store performance across all regions and model implementations was seen as the age of the store, suggesting that the older the store was, the less attractive it was. Limited data however meant that this could not be implemented into our existing model framework but instead leaves open the potential for future research. These results and performance were also not seen as the consequence of factors affecting the training or evaluation of the model in a single week. This was because the model performance could be seen as consistent across the entire year, in terms of loyalty card revenue errors, however the average error varied over the course of the year when modelling total store revenue in each region. This therefore suggested that total store revenue error could be influenced by changes in non-loyalty card revenue attribution alongside changes in potential seasonal demand such as due to tourism. This later effect was highlighted by the reasonably large variation in mean store error in Region 1 across the whole year, in line with previous research in this region (Newing, et al., 2015; Newing, et al., 2018). The sensitivity of these results to changes in training conditions and hence parameter values was explored utilising a cross-validation study that used the parameter trained on other regions to model the total revenue flows within each other using the same method and data. These results suggested that the performance was not overly sensitive to changes in parameter values, whereby there was greater variation in store errors between each region than within a region, suggesting that the performance was due to the conditions in each region and the method of scaling up to predict total revenue, rather than the individual parameter pairs. Therefore supporting the conclusion of the poor ability of the models to predict total store revenue at this scale.

These results are therefore in contrast to those presented in the previous literature (Newing, et al., 2015; Waddington, et al., 2018) and would thus indicate that at this scale the spatial interaction model in its current form and data is not appropriate to predict total store revenue. Indeed, the range of errors seen for the stores in each region would mean that such models are unlikely to be used in practice due to unreliable estimates of revenue (Newing, et al., 2015). Therefore, the aim of the next chapter is to explore how the previous literature was able to achieve their results while attempting to resolve any differences between the presented implementation above and that seen in previous papers using data that we have access to.

Chapter 6

Modelling Scenarios Replication

6.1) Overview

The previous chapter applied the Wilsonian form of the spatial interaction model to a regional dataset across three regions in the UK. The results from that chapter suggested that the current form and implementation of the model was unable to account for the heterogeneity of store conditions at this scale, leading to poor overall model fit with high variance in individual store errors. This chapter builds on this work by exploring the differences between models presented in the literature and the current implementation. This includes a summary and critique of the most recent literature to implement these models, followed by attempts to reproduce their results by examining smaller groups of stores and using an iterative calibration method. The main aim of this chapter therefore is to examine whether previous results are replicable with our own data.

6.2) Introduction

The development and application of spatial interaction models in retailing, and grocery retailing in particular, was fueled by the desire to find models that would accurately and reliably be able to estimate store revenue at a variety of different scales. In grocery retailing this was due to increased levels of competition within the industry leading to increasing costs of getting store location wrong (Birkin, et al., 2017). The advantage of spatial interaction models, as opposed to other retail location methods, was that they were able to account for the influence of geography on the movement of people, goods and money better than the methods that were used at that time[11]. This therefore led to their adoption across a variety of retailers (Reynolds & Wood, 2010), with claims of high levels of accuracy to within 10% of actual store revenue (Mendes & Themido, 2004). However, despite considerable history of spatial interaction models broadly in both academic literature and industry, it has often been difficult to verify claims of this level of accuracy and to reliably determine their effectiveness across a wide range of times and scales.

While the Wilsonian and Huffian forms of the spatial interaction models which are commonly used today were developed in the 1960s and early 1970s they were not commonly used in practice in grocery retailing until the early 2000s (Reynolds & Wood, 2010; Clarke & Birkin, 2018). This is because at the time, issues with data and limited computing power made it difficult to operationalise these models in any meaningful format (Haynes & Fotheringham, 1985). The aim of

[11] For a more complete discussion on this see section 3.4)

the models were to try to estimate unknown information, such as new store revenue or where a consumer came from, but to ensure that these estimates were correct there needed to be some form of existing dataset with which to compare results (Haynes & Fotheringham, 1985). Survey data was available but getting access to this data or verifying its usefulness for different contexts was often difficult due to low levels of sampling, selection bias and often inaccurate information (Huff, 1964). Thus, while spatial interaction models claimed to have improved performance relative to existing methods of store location, these claims could not be accurately or consistent verified. This lead to a lack of trust in their conclusions (Newing, et al., 2018) and inconsistent usage in practice early on (Guy, 1992).

It was not until the development of loyalty card data, creating a reliable and accurate dataset of origin to destination flows, that spatial interaction models could be consistently utilised in practice (Clarke, 1998). Early adopters of both loyalty card schemes and their integration with spatial interaction models were able to reap considerable rewards in terms of favourable store locations, increased margins and ultimately increased market share (Newing, et al., 2020). But these implementations, along with their subsequent advancements, were mostly confined to industry applications with limited explorations available in the wider academic literature (Khawaldah, et al., 2012). This has often meant that academia has been left behind to develop incremental tweaks in model implementations that haven't been able to be properly evaluated in this domain as to their relative improvements or merits.

It was not until recently that academic researchers have been allowed to explore how these models are used in practice with loyalty card data. This has been highlighted by a series of papers developed by the Leeds School of Geography and Sainsburys, that have shown how the models can take advantage of loyalty card data, how they need to be adapted to provide accurate estimations and ultimately what levels of accuracy can be expected. These papers were introduced in previous chapters but given the poor modelling results from Chapter 5, it is worth examining their methods and outputs in more detail. This is because the results from the previous chapter and changes in consumer behaviour identified, with the increase in convenience shopping, multi-purpose trips and e-commerce, suggest that while the revenue for some stores may be able to be predicted to within 10%, as per the results in the literature, this is unlikely to be consistently replicated or achieved at scale. Thus, it is suggested that while we may be able to replicate the performance from the previous literature on a few limited scenarios, it is expected that these will not be repeatable consistently across different groups of stores.

The evaluation of this hypothesis therefore achieved in this chapter by firstly reintroducing the key papers in the literature, identifying their contributions and highlight key differences in their model implementation relative those presented so far in this thesis. These differences are then explored in this chapter by examining the influence of scale, in terms of the number of stores, on the model implementation and outcomes, alongside what affect an iterative calibration procedure has on modelling performance at the regional scale. What is found is that the results from the previous cannot be consistently replicated, suggesting that the claims of spatial interaction modelling performance in relation to grocery retailing have been overstated, and supporting the conclusions of the influence of scale on outcomes in seen in the previous chapter.

6.3) Recent Papers

The recent string of papers that have applied the spatial interaction model, in its Wilsonian form, to a practical data driven application of grocery retailing in the UK started in early 2010s with the contribution by Newing et al. (2013). This series of papers have originated from within the Leeds School of Geography, a department with a history of developing and applying spatial interaction models to real world applications, and was enabled by a collaboration with a major grocery retailer in the UK (University of Leeds, 2021). Arguably, this string of papers is the first to have been able to consistently develop and apply a spatial interaction model using anonymised loyalty card data in the grocery retailing sector. Thus, they have been able to provide an insight into how these models have been developed and applied in practice. Therefore, it is worth highlighting and discussing their contributions, including how their model differs in its implementation to that presented in the previous chapter.

6.3.1) Tourism Demand

This series of papers could be seen to begin with the contribution of Newing et al. (2013) who emphasised that while spatial interaction models are used in store site location analysis for grocery retailing, there are often factors beyond the basic residential model that analysts need to adjust for. The example they then present is that of non-residential tourist demand which if not properly accounted for could lead to underprediction of actual store revenue. Typically, analysts would adjust for this using ad-hoc scaling parameters, but this can often fail to reflect the true increase in revenue due to a variety of factors such as weather and the amount and variety of local tourist accommodation. This therefore led to a series of papers that identified the extent of tourist demand in Cornwall in grocery retailing including the creation of a dataset of estimated tourist demand across the region (Newing, et al., 2013; Newing, et al., 2013; Newing, et al., 2014).

The creation of this dataset then led to the way for their 2015 paper which applied a Wilsonian form of the spatial interaction model to four large format grocery stores in the UK. Thus, this is one of the first papers available in the literature that provided an example of a calibrated spatial interaction model using an anonymised loyalty card dataset from a major grocery retailer. The application of this however was also enabled by access to data from previous research that showed the preferences for output area classification supergroups to different retailers, third party data on the size and location of competitors stores and their tourist revenue dataset. Thus, showing that the data requirements for implementation of these models goes beyond solely loyalty card data from a retailer. A main contribution of this work then was to use these diverse datasets to implement a spatial interaction model and to comment on its resulting accuracy (Newing, et al., 2015).

In their application they extend the basic production constrained spatial interaction model by developing an origin disaggregated form of the model. This contributes to the literature by separating the model implementation by socioeconomic group, allowing for different distance decay, β, and attractiveness, γ, parameters. This was achieved applying a different distance decay parameter across three different income groups, and separating the attractiveness parameter by output area supergroup and brand. The model is then also extended by adding in another demand layer for tourist demand based on their own created dataset (Newing, et al., 2014). The calibration of parameters is then achieved using anonymised loyalty card data from their partner retailer for the four stores in Cornwall, achieving an average error per store over the year to within +/-5% of actual revenue. Overall, they suggest that throughout the year the predicted revenue for each store varied within a bound of +/-15%, with the majority of errors within a 10% error band consistently. They then validated their model by using the trained parameters to predict the revenue for another store, achieving a similar margin of error, and also for two further stores for a different retailer. Thus, their results suggested that spatial interaction models could be expected to consistently estimate store revenue to within +/-10% of actual revenue across an entire year.

The results from the previous chapter however saw error ranges far exceed the suggested 10% bounds, even when only modelling loyalty card revenue. To this extent, there are three main differences in model implementation which could be affecting the results. The first is that of tourist demand which is integrated as a new demand layer in their model, but data was not available for our implementation. They suggest that the incorporation of this dataset reduced the overall amount of underprediction for the modelled stores across the year, in one case reducing underprediction of 50% to almost near perfect estimation of actual revenue. This factor could potentially be influencing the performance in Region 1 over the year, where there is evidence of consistent underprediction for which the extent varies over the year, but the overprediction seen in Region 2 and 3 suggest that this is unlikely to be solely driving the modelling results.

The second difference is the fact that they train and validate their model on only four stores in Cornwall, where 29, 47 and 70 stores are modelled respectively for the three regions evaluated in this thesis. It is thus expected that for four stores, especially when they are close to each other, the behaviour observed, underlying conditions and subsequent parameter estimates are likely to be similar. However, when scaling up to a whole region of stores, there is likely to be variation in both behaviour and geographical conditions that would lead to varying performance of the model at the individual store level. This is expected to be so even for an origin disaggregated model applied across eight different socioeconomic groups. Thus, opening up our model implementation to potentially be explored on a smaller subset of data to see how the models compare.

The final difference then is the method of calibration. Newing et al. (2015) firstly calibrate the distance decay parameter across three different income groups using an iterative procedure to align the average trip distance metric with the underlying anonymised loyalty card data. In doing so they follow the logic first suggested by Batty and Mackie 1972) that replicating observed trip making behaviour in terms of the average trip distance metric is likely to lead to a good representation of the overall flow pattern. Following on from this they then adjust the attractiveness parameters by output area supergroup classification according to consumer survey data to reflect their relative preferences (Thompson, et al., 2012). In contrast in this thesis the distance decay and attractiveness parameters are calibrated simultaneously by taking advantage of the iteratively re-weighted least squares calibration routine of Poisson Regression based on the underlying loyalty card flows. This then also open up the potential exploration of how using an iterative search method based on average trip distance would compare to our own results at the regional scale.

6.3.2) Daytime Population Demand

The second main contribution of this series of papers was that of Waddington et al. (2019) who built on the previous work to incorporate new demand layers of daytime population into the model. This was in response to the acknowledge failure of traditional spatial interaction model to accurately model convenience store revenue (Waddington, et al., 2019). This is because smaller format stores (convenience and high street stores) shopping behaviour difference from that exhibited at large format stores (Waddington, et al., 2018). While large format stores benefit from residential demand that is regular, with a large basket and travel by car, small format store shopping is often ad-hoc, infrequent and influenced by changing local daytime population (Wood & Browne, 2007). Indeed, Waddington et al. (2018) used temporal variation in revenue to identify different clusters of stores based on when revenue peaked and what types of goods were bought. These results were inline with the previous work of Hood et al. (2016) who identified clusters of convenience stores that targeted different types of consumers.

In this contribution they used loyalty card and total revenue data but only for a single week in West Yorkshire in 2014. This dataset contained information on 48 total stores, split between 16 large format and 32 small format stores. Ten stores were removed from the model when calibrating the parameter but they used a similar iterative procedure as in Newing et al. (2015). The main difference in this regard was that, due to the greater number of stores, the distance decay parameter could be calibrated across the output area supergroup classifications, as in this thesis, thereby disaggregating the model implementation further.

Their initial model was implemented with only residential demand layers and achieved an average prediction of 90% of large format stores total revenue and 55% of convenience stores revenue. This therefore suggested a much larger average error and range than in the previous paper, emphasised by the poor fit on convenience stores, which is more in line with the results from the previous chapter. In their figure 2 on page 432 (as reproduced in Figure 20 below) for the large format stores the error around each individual store varied greater than the mean error of 10%, and for some, such as S7, appeared to reach up to 30% error. If this was the case, then these results would be much closer to those presented in the previous chapter than those presented by Newing et al. (2015), and would support the theory that these models begin to break down when applied to larger numbers of stores.

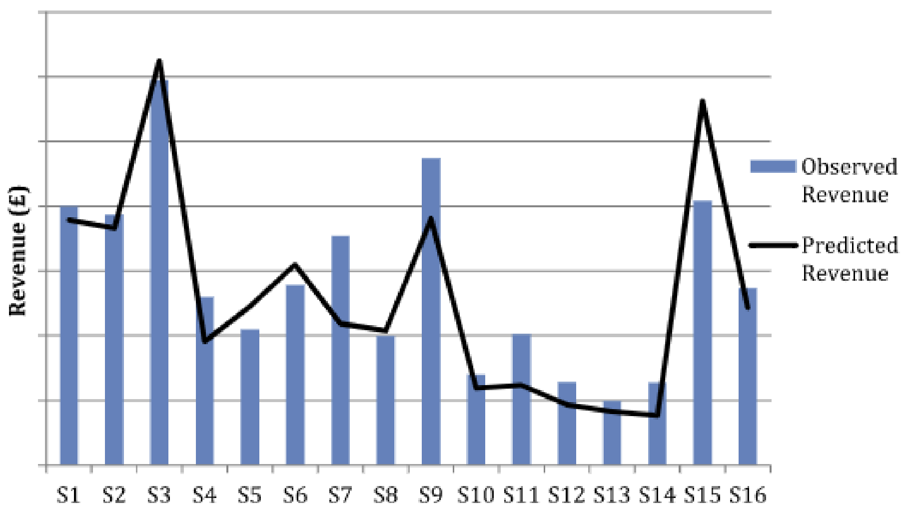

Figure 20 - Supermarket store errors from the first model implementation (Waddington et al. (2019))

The main contribution of their paper however was integrated new demand layers to improve the accuracy of the model. This included the addition of workplace, second school and university students as demand layers in a similar way to the integration of tourist demand in Cornwall. Incorporating these demand layers improved the accuracy of store revenue prediction, in some cases by up to 30% of actual revenue. The average revenue for large stores improved to 103% and for convenience stores to 83% for the single week, thus representing a greater improvement for convenience stores than large format stores. However, looking at their figure 7 on page 439 (as reproduced here in Figure 21 below), the range of errors appears considerable still. Notably for S15 there appears to be overprediction by at least 30% while for stores S1 and S11 there appears to be underprediction by a similar percentage. Thus, while they suggest an average error range of +/-10% would be acceptable for retailers, a wide range of errors may affect confidence in the model implementation and its ability to be used in practice.

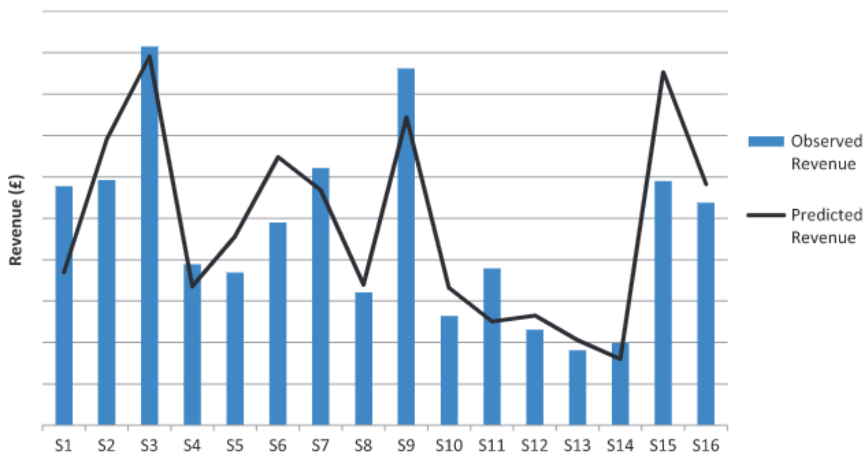

Figure 7. Observed and predicted revenue for supermarkets with the final model (S1-S16).

Figure 21 - Observed and predicted supermarket store revenue from the complete model (Waddington et al. (2019). pg 439)

Thus, while there are similar differences between this paper and the model implementation in this thesis as there were within Newing et al. (2015), notably both scale (16 large format stores) and the calibration method, the results above highlight some of the issues that are part of the spatial interaction model in relation to grocery retailing. The difficulty of modelling convenience stores was addressed to some extent through the integration of new demand layers, notably data that could not be brought into this thesis, but the variance in store errors are closer to those presented in this thesis so far compared to Newing et al. (2015). This therefore lends support to the argument that as the number of stores increases, then the range of errors increases, even with a more disaggregated model. Further analysis of this however was limited as their model was only able to be applied to a single week. Thus, along with the influence of scale and calibration method, opens up a further opportunity for exploration below in terms of expanding a sixteen store subset model over a whole year.

6.3.3) E-Commerce

The latest paper in this series is that of Beckers et al. (2021) who attempt to build on both of the previous papers and also the burgeoning literature on the geography of grocery e-commerce sales (Clarke, et al., 2015; Hamad & Schmitz, 2019; Davies, et al., 2019; Kirby-Hawkins, et al., 2019; Hood, et al., 2020). In doing so, they state that with e-commerce demand growing within the sector consumer demand is becoming more complex due to the variety of potential interactions. Nevertheless, they argued that spatial interaction models were still relevant in grocery retailing to model e-commerce demand because they have a proven track record, they consider the whole system and have scaling potential for regional and national level applications. They thus develop an attraction constrained spatial interaction model to send online revenue from stores to output areas using a modified distance decay function, and then add in a new demand layer to the disaggregated model to account for online revenue spend. While there was no loyalty card data to calibrate the model, they compared the distribution of revenue to previously published results, suggesting that their model could be used to estimate both online and offline revenue together. This therefore left open the validation of the estimates and actual implementation, but their main contribution is to suggest that grocery e-commerce revenue could still be integrated, albeit in an adjusted form, within existing spatial interaction models. This is because of not only difference in consumer behaviour, but also because of how retailers assign revenue to stores through online sales (Davies, et al., 2019), thereby altering the distance decay relationship.

6.3.4) Significance of These Developments

The influence of this series of papers is in showing the ways in which spatial interaction models can be implemented using anonymised loyalty card data and identifying ways in which they have to be modified to estimate total store revenue. These have thus extended the literature through the development of the disaggregated form of the spatial interaction model and the introduction of new demand layers in grocery retailing applications. The main differences between these papers and the application in the current thesis however is the scale at which the models are applied, the calibration method and the introduction of non-residential demand datasets. While we were limited on bringing in new datasets due to regulation or the impossibility to replicate the datasets, the influence of both scale and calibration method can be explored in their influence on modelling results below. Specifically, in light of the results presented in the previous chapter, the analysis below focuses on whether the results from these papers can be replicated by examining the influence of the number of stores modelled and the calibration method used at the regional scale. It is expected that while there will be some scenarios that the results from the literature can be replicated, this is unlikely to achieved consistently across multiple modelling implementations.

6.4) Subset Analysis

The first question to explore then is the potential influence of scale on the modelling results and whether the performance from the papers mentioned above can be replicated. Specifically, this focuses on the number of stores that are modelled and how consistent the results across different number of stores within modelled groups. For this, Newing et al. (2015) trained and evaluated a spatial interaction model on four large format stores in the region of Cornwall while Waddington et al. (2019) examined a total of 48 stores in the West Yorkshire region, 16 of which were supermarket format stores. The scale of this analysis then is in contrast to that presented in the previous chapter where Region 1, 2 and 3 contain 29, 47 and 60 large format stores respectively. From these results it is thus expected that as the number of stores modelled increases, so does the range of errors, but also that it will be difficult to replicate the performance seen in the previous papers even at the same scale. Thus, while one aim of spatial interaction modelling application in grocery retailing is to develop a national level model (Beckers, et al., 2021), this section aims to explore the influence of the number of stores modelled and whether the results from previous papers can be replicated.

6.4.1) Model Formulation

The size of the subsets chosen for this analysis includes groups of 4, 5, 7, 10 and 16 stores. This is to provide a range of group sizes in between the 4 and 16 store group sizes from the previous papers to be able to understand how the performance varies across and in between the range of stores. Individual groups were then identified in Region 2 based on the assumption that the previous applications used stores that were closest to each in terms of distance. Thus a spatial weights matrix was constructed using as-the-crow-flies distance between each store in the regional dataset. This dataset was then used to identify the closest 3, 4, 5, 9 and 15 stores for each individual store in the region to make groups of 4, 5, 7, 10 and 16 stores respectively. An example of this can be seen in Figure 22 below which shows how the group of different sizes were constructed for an example store. This meant that there were 47 different groups of stores for each subset size, one for each store in the region, which could then be used to evaluate whether the results from the previous papers could be consistently replicated.

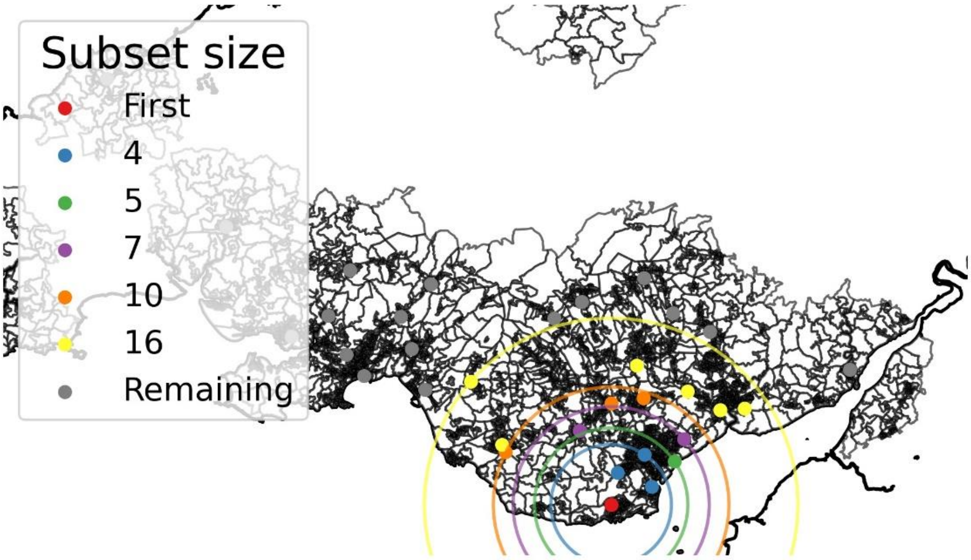

Figure 22 - An example of store group subset construction based on distance for a single store

The disaggregated form of the spatial interaction model was chosen to be used in this analysis, thus following the implementation of the equation below:

$$T_{ij}^{kn} = A_i^k O_i^k \mathrm{W}_\mathrm{j}^{\alpha^{\mathrm{kn}}} e^{-\beta^k d_{ij}} \quad \textit{Eq. 54}$$

Where:

$$A_i^k = \frac{1}{\sum_j w_j^{\alpha^{kn}} e^{-\beta^k d_{ij}}} \quad \textit{Eq. 55}$$

And:

$$O_i^{kt} = e^{kt} n_i^{k2011} \quad \textit{Eq. 56}$$

This therefore follows the final model implementation from the previous chapter whereby the model is disaggregated by the output area supergroup classification *k*, and also by brand *n*. This also follows the model implementations used by both Newing et al. (2015) and Waddington et al. (2019), so as to ensure that there is consistency in model application to allow for a clear comparison[12]. This takes advantage of both the drivetime data that was available for the analysis in Region 2 and the brand attractiveness factors for each output area supergroup classification. This disaggregation however creates implementation issues for the smaller subsets of 4, 5, and 7 stores due to data

[12] The inverse power decay form of the model was also used in this analysis. However since the results do not differ considerably from the exponential decay form of the model the results are presented in Appendix C.

limitations where some output area supergroup classifications do not have enough data to accurately calibrate parameters. Nevertheless, as with the application in the previous chapter, where there is not enough data for a supergroup to be calibrated, then the parameters are replaced with those from the system wide model for that group of stores.

However, there are two main differences in model implementation between this model form and those in the previous papers. The first difference is the lack of transient population data in the form of either tourist or daytime population data, which was available to Newing et al. (2015), or daytime population which was available for Waddington et al. (2019). To this end the effect of tourist demand should be minimised as the application focuses on Region 2, previously identified as being less influenced by tourist demand than region 1, and the analysis is performed on a single week where there is expected to be no or little tourist demand (13th February 2017). Furthermore, daytime demand was seen to have the largest influence on convenience stores which are not examined here(Waddington, et al., 2019). Thus, transient population effects are expected to be minimal, and multiple groups are stores are analysed so as to minimise any potential influence of these effects. The second difference then is that Poisson regression is used to calibrate the model parameters, in contrast to the iterative calibration procedure used by both Newing et al. (2015) and Waddington et al. (2019). While the iterative calibration method allows only for the distance decay parameter to be calibrated based on loyalty card data, we take advantage of the iteratively re-weighted least squares calibration routine in the Poisson Regression formulation to allow for multiple parameters to be calibrated simultaneously. These model differences therefore may affect the results, however if the model is as robust as suggested in the previous papers, then we should be able to replicate the performance for at least a few instance of the subset implementations.

6.4.2) Subset Range Performance

The results from the application of this model can be seen in Figure 23 below which shows the distribution of mean store errors from each group within each subset size. Thus, each violin represents the mean total revenue error from 47 different groups of stores. What is interesting to see here is that as the subset size increases then the range of mean errors decreases. This suggests that the modelling results become more stable and consistent as the number of stores within each subset increases. To an extent this is to be expected because as the size of the subset increases then there is a greater chance of overlapping stores and hence consistency in behaviour. However, the range of errors from the four and five store subsets suggest that even when training on a small group of stores and then modelling their total revenue, there are issues in implementation and underlying conditions that still lead to poor model fits and inconsistent results. Thus, while there is consistency in application for larger groups of stores, it suggests that there is a lack of consistency in model performance when examining smaller group of stores. This is then likely to undermine the results seen in Newing et al. (2015) if they fail to be replicated consistently, where consistency increases with the number of stores modelled.

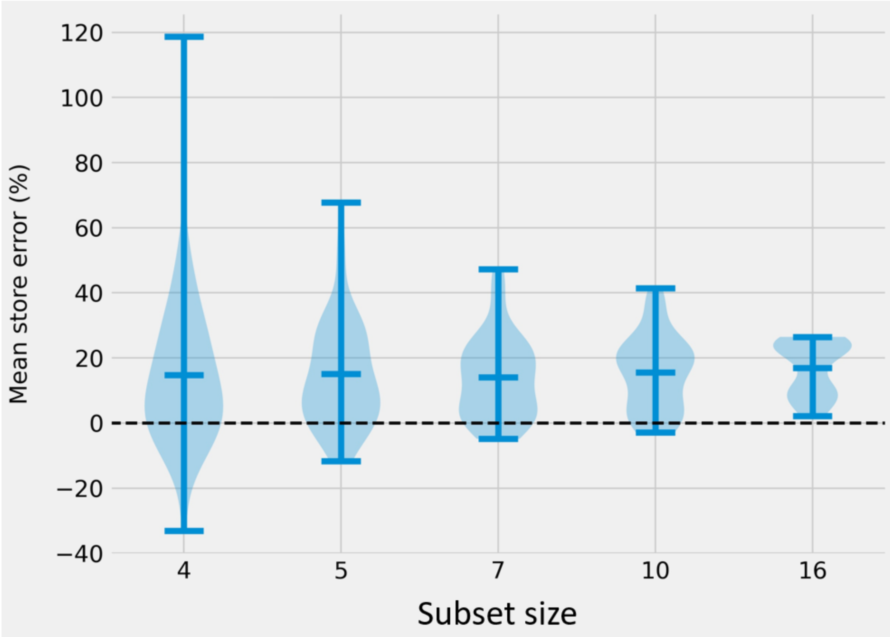

Figure 23 - A violin plot showing the ranges of mean percentage error for each group for each subset size in Region 2

Nevertheless, it can also be seen from this figure that for all subset sizes there should at least be one group of stores for which mean store error would be close to the optimal value of 0%, or at least within the 5% mean error range suggested by Newing et al. (2015). In this case, it is only the 16 store subset groups that do not appear to have any group of stores that would produce an estimate that be at the 0% ideal, but there still appears to be groups within a 10% error range seen in the first model implementation from Waddington et al. (2019). Therefore this suggests that at least one group of stores from each subset size should replicate the results seen in the previous papers, but the overall distribution suggests that this is unlikely to be replicated. Thus, it becomes improbable that other researchers may be able to replicate the results from the literature at this scale using this form of the model and anonymised loyalty card data. To this end therefore we can take a deeper look at the variation in store performance for the groups within both subset size four and sixteen, so as to directly relate their performance to previous results.

6.4.3) Four Store Subset Performance

The first subset size to examine is the four store subset because it is the same size of store group that Newing et al. (2015) applied their model on. In their application they were able to achieve a mean error range of +/-5% across the whole year for each store with the majority of store errors within a 10% error band individual weeks. Our results, as applied to a single week, can be seen in

Figure 24 below which shows the error range across the 47 different groups of four stores in the region. If we take the range of +/10% of actual revenue as the target range, with mean store error within 5% of actual revenue, then out of all 47 potential groups of stores, only two were within these bounds. For most other groups only 10 had a mean error within +/-5% of total error and only one other group had a total error range of 20%. This therefore suggests that the results seen in Newing et al. (2015) are not consistently replicable if only two groups were found to have errors within their suggested bounds. Furthermore, these two groups represent the same four stores, which further reduces the confidence in the ability to replicate Newing et al.'s (2015) analysis.

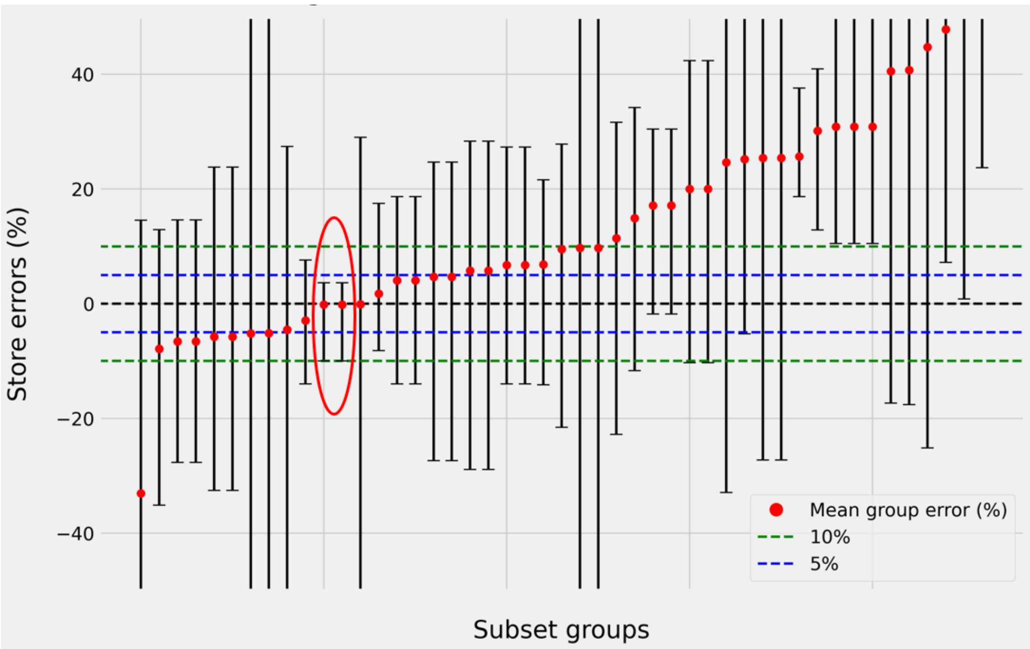

Figure 24 - Error range for groups of four stores including the mean error in comparison to the error ranges suggested by Newing et al. (2015)

These results, however, were only for models that were calibrated on a single week. In Newing et al. (2015) the model was applied on every week over a whole year, which is where these error bounds come from. This is because they acknowledged that while the mean error for each store was within +/-5% of actual revenue, the error could vary over the year. Indeed, they suggested that for a single week the error range for an individual store could be within +/-15%, but that the majority of errors were within +/-10%. This can therefore be repeated by taking the highlighted group of stores from the figure above and apply the spatial interaction model over every week in a complete year to see how the error range varies. The results of this application can be seen in Figure 25 below which shows the range of errors for each individual store that were modelled in the group. This is alongside the error range of other stores across the year which had a mean error within a 10% error bound across the whole year when they was modelled using the parameters calibrated from the four store subset. This is because Newing et al. (2015) validate their analysis by applying their model on a further store to ensure that their model is generalisable.

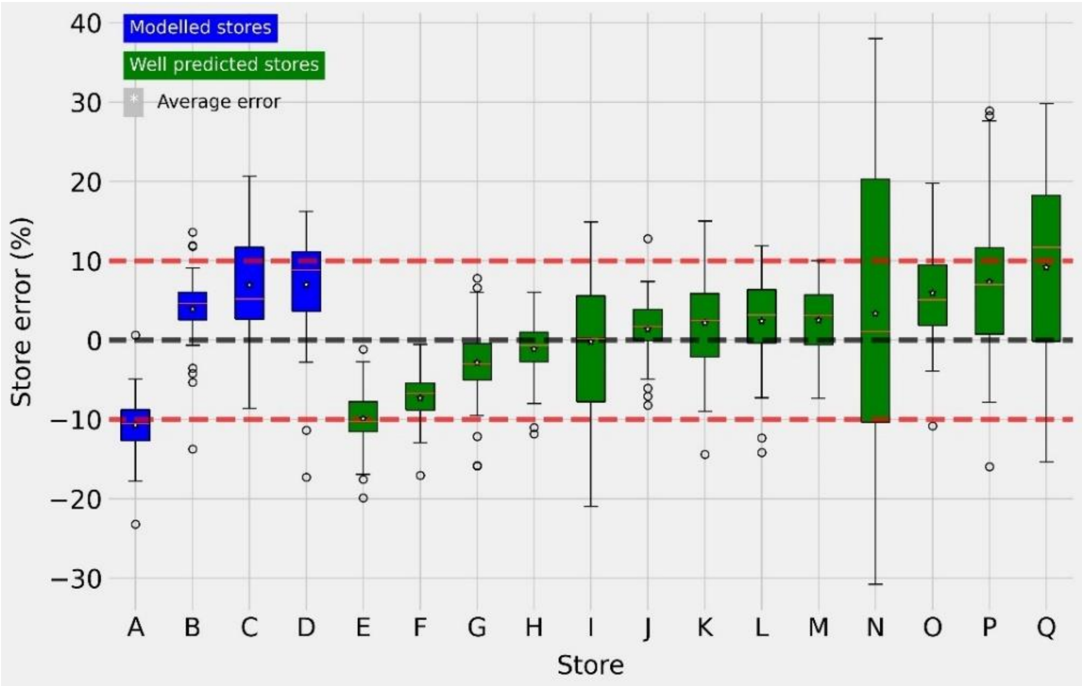

Figure 25 - The range of weekly store errors for each store in the modelled store group of four stores and other "well-predicted" stores in the region

The results above shows the distribution of errors for each individual store for every week in 2017 including the mean and median store error. This shows that the errors for the four modelled stores vary over the year, with errors ranging from below -20% to over 20%, and that there are different ranges of errors for each store across the year. Although for the single week analysed in the previous figure the four store errors were within +/-10%, when the model is extended over the full year the error range for this group increases to within +/-25% of actual store revenue. These errors are therefore beyond the +/-15% total range suggested by Newing et al. (2015), and thus do not support the replicability of their results. Such range of errors could exceed the results suggested by Newing et al. (2015) potentially due to non-residential demand not being integrated into the model, however if this was the case then all four stores would be consistently underpredicted over the year. Thus, the fact that only a single store is consistently underpredicted with the other three are consistent over predicted suggests that this is not likely to be the case. Nevertheless, for this group of stores, all four stores mean errors are within +/-11% over the full year, suggesting that in this case the spatial interaction may potentially be used to examine a small subset of stores consistently over the year. However this group of four stores is only two out of the potential 47 identified in Figure 24. Thus, if the range of errors seen in Figure 25 for the whole year are larger than for the individual week, this is also likely to be the case for other 45 groups of four stores. This therefore supports the suggestion that the modelling performance seen in Newing et al. (2015) are not able to be consistently replicated.

When extending the model implementation by utilising the parameters trained on the four store subset to model all other stores in the region, it can be seen that 13 other stores in Figure 25, out of the remaining 42 in the region, have mean yearly errors that fall within a 10% error band. These stores are identified on the basis that Newing et al. (2015) validated their model on another store to show the generalisability of their model. Out of the further 13 well-performing stores however only three, H, J and M, have errors that are consistently within the +/-10% error band. Thus, since this is out of a possible 43 stores across the region, this model is not generalisable across the majority of stores within the region. In summary, out of 47 possible combinations of groups of four stores in the region, only two groups showed error ranges that were within those suggested by Newing et al. (2015). When then extending this model for this group of stores over an entire year, our error range exceeds that of Newing et al. (2015), with further evidence of poor regional generalisability. Thus, at this scale, we are unable to replicate the results from previous research.

6.4.4) Sixteen Store Subset Performance

Given the results in Figure 23 and those from four store subset groups, it is therefore worth examining the results from the 16 store subset groups. This is because it is the same number of large format stores that Waddington et al. (2019) applied their spatial interaction model on. In their original model format, without additional non-residential demand, they applied the origin disaggregated spatial interaction model to 48 total stores within West Yorkshire, 16 of which were large format stores. This scale of application allowed them to apply the origin-disaggregated model format across all seven 2001 output area classification supergroups, similar to the eight different 2011 output area classification supergroups as evaluated in this scenario. In this application they were able to achieve an average store error for the large format stores of within 10% of actual store revenue with a range of errors around +/-30%. For this, the replication of this scale of model across all 47 possible groups of 16 stores within region 2 can be seen in Figure 26 below which shows the range of errors for each group and their mean error.

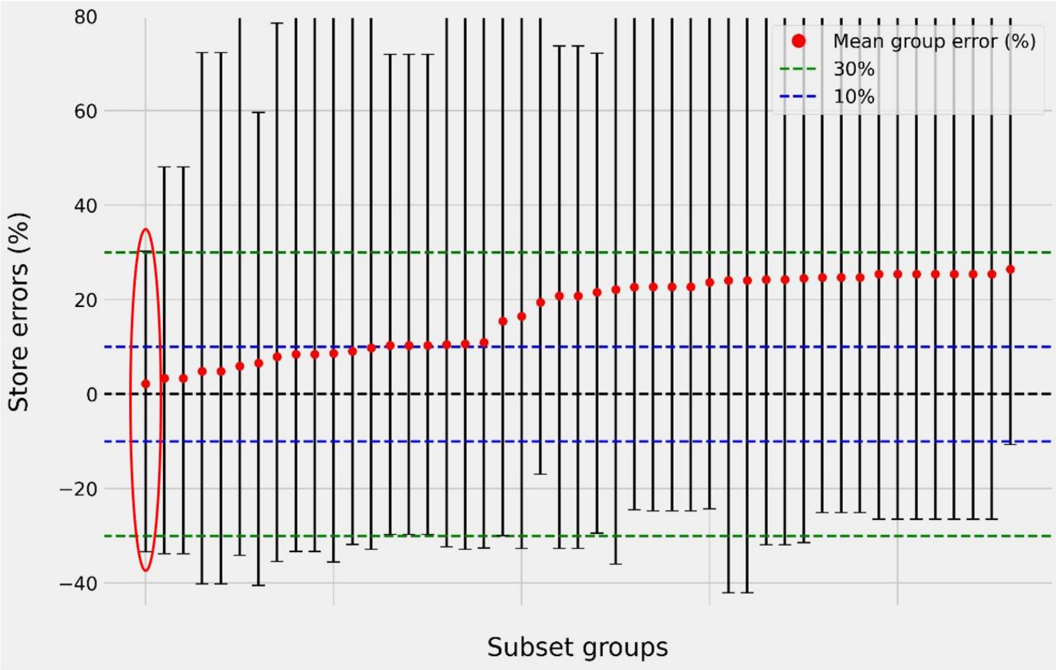

Figure 26 - Range of errors for groups of 16 stores along with the groups mean error. The horizontal bars indicate the error performance achieved by Waddington et al. (2019)

From this figure it can be seen that while the range of mean errors across each group is more consistent than for the four store subset results in Figure 24, the range of each individual group of stores is consistently larger. This is therefore consistent with the results presented in Figure 23 in terms of the range of mean errors and the idea that as the number of stores increases in the model then so does the range of errors. This therefore supports the conclusion of the previous chapter that the regional scale is not appropriate for the spatial interaction model implementation due to the range of behaviours and store conditions influencing the range of errors. Nevertheless, taking the error bounds suggested by Waddington et al. (2019) as a mean error within +/-10% of actual revenue and an individual store error range of +/-30%, only one group out of forty-seven comes close to these levels of performance. The group that achieves level of performance, or close to this, is the first group that is highlighted in the figure above. This group has an upper error bound of 31% overprediction and a lower bound of 35% error prediction with a mean error of 2.22%. Thus, out of 47 different groups of 16 stores, only one group is able to replicate results seen in the residential only origin-disaggregated model implemented by Waddington et al. (2019), suggesting that these results cannot be consistently replicated.

A limitation of the Waddington et al. (2019) analysis, relative to that of Newing et al. (2015) however is that they were only able to develop and apply their model for a single week in West Yorkshire. Thus, their derived range of errors and mean error could be greater, as it was in the case of Newing et al. (2015) and in the four store subset analysis presented above. Therefore, as with the four store subset model, we can take the group identified in the above, which roughly fits within the error bounds suggested by Waddington et al. (2019), and extend the model implementation over an entire year. The results of this can be seen in Figure 27 below which shows how the minimum, maximum, mean and inter-quartile range of individual store errors for this selected group varies over the year.

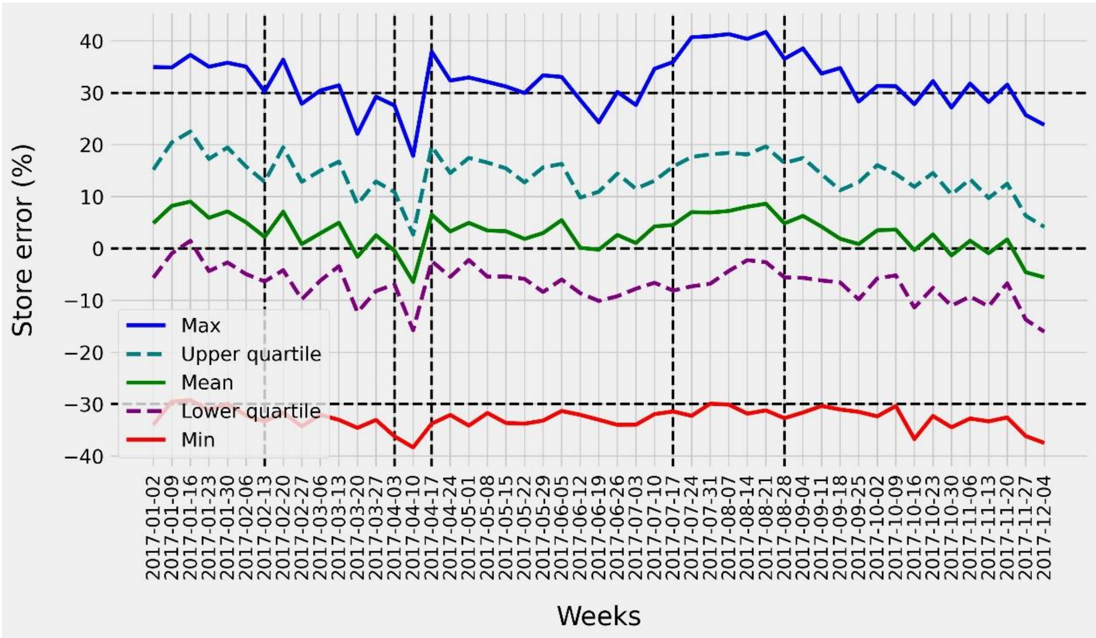

Figure 27 - Sixteen store subset group applied over the entire year, including the maximum, upper quartile, mean, lower quartile and minimum store error from within the group

What can be seen from this figure is that, as with the results seen in Newing et al. (2015) and the four store subset analysis above, the range of errors over an entire year is greater than for the single week. In this figure it can be seen that despite the single week errors ranging from -35% to 31%, this increases to a minimum error of -38% and a maximum error of 42%. This therefore suggests that if Waddington et al. (2019) were able to extend their analysis over a whole year, the range of errors is likely to increase. Alongside this however it can be seen that the mean error within the group varies between a range of 10% overprediction and 6% underprediction. This suggests a degree of consistency across the year but it is mainly extended by seasonal variation within the group such as clearly identified changes at the beginning of the year, during the Easter holidays and the summer school holidays. These changes suggest that during the Easter holidays more revenue is generated by the stores than predicted by the model, thereby suggesting the influence of tourist demand in the region. In contrast in the summer holidays there is a tendency towards overprediction, suggesting that there is a reduction in residential demand which is potentially due to local residents going on summer holiday abroad or elsewhere in the UK. These changes can also be seen in the upper and lower quartile errors for the group, where most errors appear within a range of -10% and 20%, suggesting that the total range of errors is driven consistently by only a few poorly modelled stores in the group. Thus, while the majority of the stores can be consistently well predicted, there are a few that are consistently poorly predicted across the whole year.

6.4.5) Subset Summary

What can be seen from these results presented in the sections above is that firstly the larger the subset of stores analysed is then the more consistent the mean store errors are. This is likely to be because of the potential overlap between conditions, as there are only 47 total stores within the region, but also because of a degree of consistency in there being some well modelled stores and others that are consistently under or overpredicted. Secondly, the larger the subset is, the larger the individual range of errors and the more consistent this range of errors is as well. This therefore supports the conclusion from the previous chapter whereby the greater the number of stores in the model, the greater than range of errors, and thus the spatial interaction model in its current form may no longer be appropriate when applied to a full regional scale. Finally, in terms of being able to replicate the results from previous papers, at both the four and sixteen size subsets, only one group out of 47 were able to come close to the suggested performance. Thus, these results were not able to replicated consistently, therefore suggesting that the spatial interaction model not even be consistent at this scale. Furthermore, when the “ideal” groups for each size of subset were then modelled over the whole year, the range of errors exceeded the suggested threshold consistently. This therefore further supports the idea of lack of consistency in application or replicability of previous results and hence therefore suggests that spatial interaction models cannot be used consistently in their application to grocery retailing in the UK.

6.5) Iterative Calibration

The results presented above therefore suggests that the performance seen in Newing et al. (2015) and Waddington et al. (2019) cannot be replicated consistently even when applying the spatial interaction model to smaller groups of stores within the same region. The results also support the argument that as the number of stores increase, then so do the potential behaviour and store conditions that the model needs to represent, which at the regional scale the origin disaggregated model is not able to account for these variances within the data and current model formulation that is used. However, it must be acknowledged that the calibration methodology used in this thesis is different to that used by Newing et al. (2015), Waddington et al. (2019) and Beckers et al. (2022). In this thesis a Poisson regression methodology is used to model and predict loyalty card flows, with the parameters obtained from the model then being used to model total store revenue. In these papers however an iterative calibration procedure is used to align the average trip distance from the total revenue predictions to that seen in the anonymised loyalty card data. Furthermore, while we take advantage of the iteratively re-weighted least squares calibration routine of the Poisson Regression model to calibrate both the attractiveness and distance decay parameter simultaneously, these papers calibrate the distance decay parameter using the anonymised loyalty card data and then adjust the attractiveness parameters afterwards based on brand attractiveness for out output area supergroup classification. Thus, there is a difference in the calibration methods used between this thesis and the previous papers in terms of both the method used and the parameters that are calibrated.

6.5.1) Model Implementation

While it is argued above that the main reason the spatial interaction model in its current form has failed to perform as expected is because of the scale at which the models are applied, it could also be suggested that the calibration method used is not suited to the application of a production constrained spatial interaction model in a grocery retailing environment. This is because, while the Poisson regression formulation is a well known and used calibration tool in spatial interaction modelling (Flowerdew & Aitkin, 1982; Tiefelsdorf & Boots, 1995), it is trained on individual flows rather than systems behaviour. This means that it could be argued that the estimated parameters do not reflect the overall system behaviour (Batty & Mackie, 1972), but rather a subset of the potential behaviours. Thus, the model may not be able to accurately estimate total store revenue consistently. This is despite the discussion presented in Chapter 4.1 on calibration techniques, primarily because loyalty card datasets are biased samples of the overall system behaviour (Birkin, et al., 2017; Rains & Longley, 2021). Furthermore, it could also be suggested that the calibration method is over-fitting and over-parameterising the model with the use of regression to calibrate two different parameters (Newing, et al., 2015). Therefore, an iterative procedure can also be evaluated on the regional datasets to examine how the performance of the model changes, if at all, in response to an alternative calibration technique with a different calibration objective.

The metric then chosen to calibrate the model is that of average trip distance (ATD). This is because it is the metric that Newing et al. (2015) and Waddington et al. (2019) use to calibrate their model, therefore attempting to replicate their model implementation. Furthermore, it has been previously argued that calibrating a spatial interaction model using this metric would generate predictions in line with the observed data by modelling the overall system behaviour (Batty & Mackie, 1972). However, rather than use an iterative procedure to search for the "optimal" parameter values that would generate predictions of total revenue with an ATD value of 1, a grid was used to explore the potential parameter space. The benefit of this, as opposed to an iterative search procedure, is twofold. Firstly, a search procedure may get stuck in a local minima, which could potentially be constrained by the acceptance and search criteria, resulting in non-optimal parameter pairings. Secondly, the use of a grid search allows for an exploration of the full parameter space, thereby facilitating an exploration of any changes in behaviour across the potential parameter space alongside the identification of any potentially interesting results.

Due to the aim of two parameters to be calibrated, both the attraction and distance decay, the model formulation chosen to was that of the base, non-disaggregated, spatial interaction model. This was chosen in light of the results from the previous chapter that suggested that there was only a marginal performance improvements from the use of the origin-disaggregated model, and that exploration of the full parameter space would mean that only one instance of the model needs to be explored rather than the influence of several simultaneously. The use of this model therefore limits the potential for overlapping or inconclusive results that would have taken considerable time and efforts to explore with no clear or significant benefits. Thus, the model formulation utilised takes the form of:

$$T_{ij} = A_i O_i \mathrm{W}_{\mathrm{j}}^{\gamma} e^{\beta d_{ij}} \qquad \textit{Eq. 57}$$

Where:

$$A_i = \frac{1}{\sum_j w_j^{\gamma} e^{\beta d_{ij}}} \qquad \textit{Eq. 58}$$

6.5.2) Grid Search Results

For this, the potential parameter space to be explored was identified by reference to the parameters calibrated form the Poisson regression implementation of this model of γ (as destination attractiveness parameter) equal to 1.36 and a β (as distance deterrence parameter) equal to -0.00604. This is because the ATD value from this implementation was already close to 1 (1.057) and thus suggests that parameters surrounding these values would produce an ATD equal to 1. Initial exploration of potential parameter pairs surrounding this thus suggested that optimal values could be expected within the region of γ from 6 to 0.1 and β from -0.015 to -0.005. Intervals were thus set to 0.01 for the γ parameter and 0.0001 for the β parameter to generate 100,000 observations within this parameter space, the results of which can be seen in Figure 28 below.

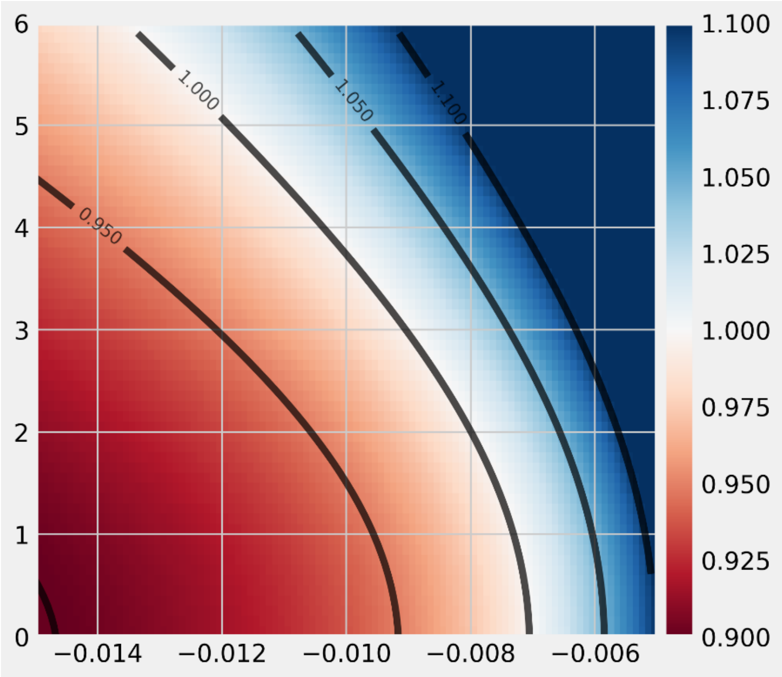

Figure 28 - Average Trip Distance (ATD) for a grid search of parameter values as implemented for a whole region

What can be seen from this figure is that there are a range of parameter pairs which produce optimal values for ATD. Indeed, while the previous literature suggested that a single optimal value could be found when calibrating only a single parameter, this figure shows that when calibrating two parameters simultaneously then there are a range of potential parameter pairs that produce total revenue estimates for which ATD value is equal to 1. This is clear from both the contour line on the figure, alongside the fact that there is a crossover point from which ATD values go from greater than 1 to less than 1 consistently. This therefore raises the question as to how do the total revenue store estimates vary across the potential parameter range and whether any of these parameter pairs produce a model that outperforms the one achieved through Poisson regression calibration. Such improvements would be measured by the range of errors seen within the modelling results alongside the average error of total store revenue predictions.

6.5.3) Parameters Pairs Exploration

Since there is a potential range of parameter pairs along the contour line that would produce an ATD value equal to 1, rather than implement a model using all of them, a subset of potential parameters were picked to be further explored. These values were selected by identifying 100 parameter pairs that produce values closest to ATD equal to 1 but that were roughly evenly spread across the line of potential parameter pairs. The parameter pairs selected based on this criteria can then be seen in Table 9 below relative to those calibrated by Poisson Regression.

Table 9 - Parameter pairs derived from the grid search which produce ATD values closest to 1

Pair	Attractiveness (γ)	Distance decay (β)
1	1.14	-0.00738
2	2.56	-0.00854
3	3.98	-0.01034
4	4.82	-0.01160
5	5.96	-0.01346
Poisson Regression	1.36	-0.00604

These sets of parameters were then used to estimate the flow of total revenue from origins to each store within the region. These flows were then aggregated at the store level in order to determine how week they were able to predict total store revenue. The predictions then generated could then be compared to actual store revenue in order to determine the error within the model, as with previous implementations of the models presented in this thesis. To this end, the range of store errors for each parameter pairing can seen in Figure 29 below as compared to those produced with Poisson regression.

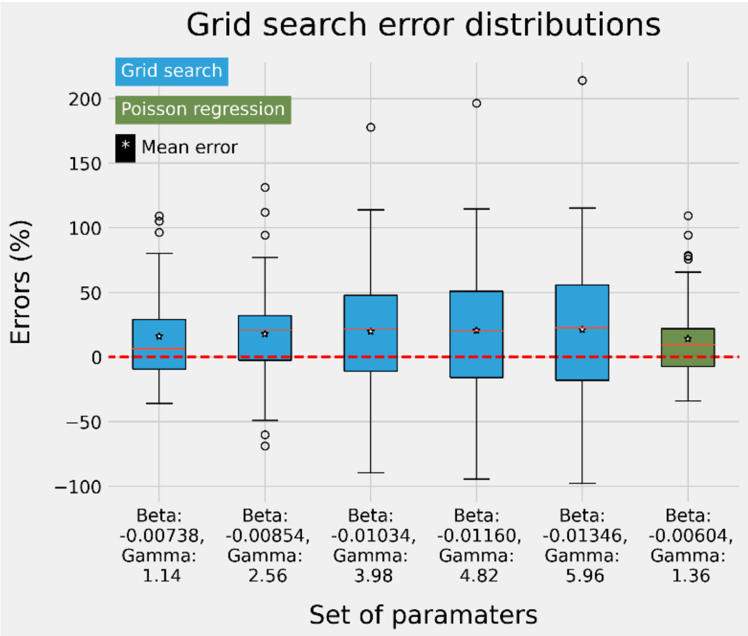

Figure 29 - Error ranges for each set of parameter pairs derived from the grid search for which the parameters generate total revenue predictions whose ATD value is close to 1

This plot shows that none of the selected parameter pairs from the grid search implementation produce results that were able to resolve the issue of large variance of store errors that were seen in the Poisson regression calibrated model. Indeed, while the first pair of parameter pairs from the grid search model appears to have the lowest range of errors for individual stores, this range is not considerably different than those produced by the Poisson regression model and the inter-quartile range of errors is larger. This is such that first group of parameter pairs errors range from -36.00% to 109.17%, with an inter-quartile range of 38.27%, while the Poisson regression model has an error range from -34.03% to 109.30% with an interquartile range of 29.33. Thus arguably the Poisson regression performs slightly better while producing less variance in individual store errors. It is also notable in this sense that all five pairs of parameters implemented in this model show large ranges of errors for stores, with the majority showing at least one error above 100% overprediction and one error below 50% underprediction. Thus, this suggests that calibrating the model using an iterative calibration method or grid search with the aim of producing an ATD value close to 1 has not resolved the problem of large variance in errors across the region and hence not produced a more accurate model. Furthermore, that a sole target of ATD is not enough to accurately and reliably calibrate a model which has two parameters.

What is interesting however is to explore how the results for individual stores vary across the five different selected parameter pairs. These responses, in terms of the individual store errors, can be seen in Figure 30 below. For this, the colour of the dot in the scatter plot on the left, showing the individual store error for that parameter pair, corresponds to the colour of the dot in scatter plot on the right, which shows the parameter values in that parameter pair. What can be seen in this figure is that the first parameter pair (represented by the blue dot) shows stores that are both under and over estimated, and the individual store error ranges from around 40% underprediction to 100% overprediction. As with the Poisson calibrated modelling results, there is also a tendency towards overprediction within the model as well, as indicated by the greater number of stores that are overpredicted in the figure. What is interesting however is that as the parameters change from pair one to pair five (blue to grey dots) each store responds differently. In this case, some stores see more revenue assigned to them which leads to increases in the percentage error, as highlighted by stores in the blue circle, while others see a decrease in revenue assigned to them, leading to a reduction in the percentage error as highlighted by those in the red circle. For this, both the direction and strength of the change differs between stores, with some seeing strong positive responses in terms of both revenue and percentage error, while others react weakly positive, and similarly with negative reactions as well. Furthermore, there are also some stores were there is almost no change in response to different parameters. Importantly however there is no evidence of a convergence of all stores towards a zero error for any parameter pairing which would suggest the existence of an optimal parameter pair. Indeed, while some stores that were underpredicted shift towards overprediction across these parameter sets, others that were underpredicted become even more so. Thus, there is no consistency in the response of individual stores that would suggest a convergence.

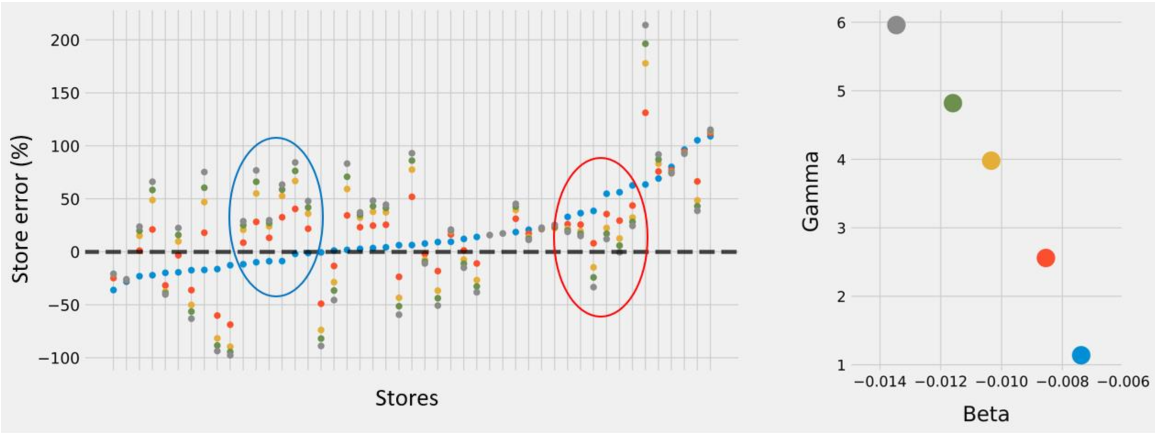

Figure 30 - Individual store error in terms of total revenue for each parameter pair derived from the grid search model

These results are also interesting as they suggest that although the evaluated parameter pairs may be able to produce results that represent the overall system behaviour (Batty & Mackie, 1972), there is more going on within the model implementation that affects modelling performance than simply the average trip distance of revenue. Indeed, from this exploration there appears to be no individual set of parameters that are able to deal with the heterogeneity in behaviour and store conditions that would lead to an optimal model. Indeed, even the best performing set of parameters identified from the grid search performs slightly worse in terms of both mean and range of errors than the parameters derived from the Poisson regression model. This therefore suggests the Poisson regression model results are accurate, in terms of range of error and mean error, relative to those produced through calibration targeted on the ATD value. Nevertheless, the results also suggest that there may be something missing from the model implementation, whether that is data or model formulation, that could potentially lead towards a convergence towards a zero error within the total revenue estimation, or that the spatial interaction model cannot be accurately calibrated or assessed on this scale.

To this end, the factor that appeared to be most related to the change in error value for each store across the parameter pairs was that of store size. The effective response, in terms of the change in error value, can be seen in Figure 31 below. While the relationship is not perfect, with some variation around a positive linear relationship, the results appear to suggest that larger stores gain value while the smaller stores lose value when moving from parameter pair 1 to parameter pair 5. This would therefore suggest that there is a trade-off inherent in the “optimal” parameter pairing line between store attractiveness and distance in modelling total revenue to stores. Here, this suggests that the increase in the attractiveness parameter outweighs the increase in the distance decay parameter such that the larger stores for our partner organisation become relatively more attractive than smaller stores. This is all while maintaining an ATD value equal to 1 which suggests that the average trip distance of revenue is still the same across each of these parameters pairs as observed in the anonymised loyalty card data. The imperfectly linear relationship however suggest that this is not a perfect trade-off and that there are still other factors influencing this effect.

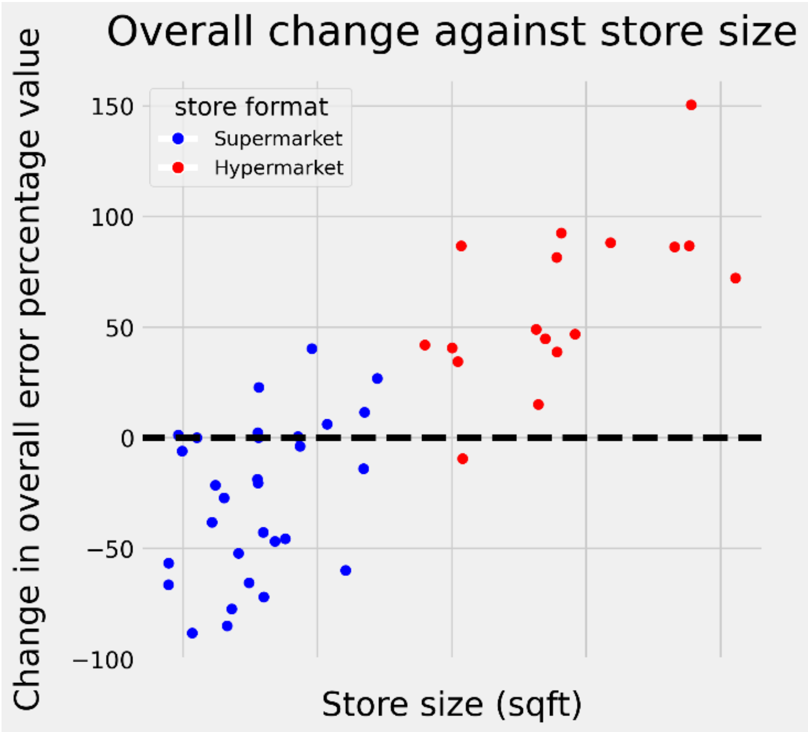

Figure 31 - Absolute change in percentage error in response to moving form parameter pair 1 to parameter pair 5 from the grid search model

6.5.4) Iterative Calibration Summary

The conclusion that can be drawn from these sets of results then is that even when using an iterative calibration procedure based on the ATD metric, the modelled total revenues are still not able to account for the heterogeneity in store conditions and shopping behaviour across a whole region. Indeed, even the best performing parameter selection from the chosen parameter pairs is only able to perform slightly worse than the Poisson regression model when used to calibrate the base model specification. This therefore supports the conclusion that it is the scale at which the model is applied to that leads to the results from the previous chapter, not the effect of the chosen calibration procedure. Furthermore the grid search shows that if an iterative calibration or search procedure were to be applied in this case, based on finding an optimal ATD value, it is likely that they would not be able to the “optimal” parameter pair due the range of parameter pairs that could in theory satisfy this condition. Therefore, future model implementations should utilise more than just the ATD metric when calibrating the model using two parameters. Furthermore, future examination of this could include the implementation of the origin disaggregated form of the model, to see how each individual output area supergroup is able to be calibrated. However, given the results above it is likely that a range of "optimal" parameter pairs would also be found, which would complicate the analysis and implementation further.

6.6) Conclusion

Spatial interaction models were originally developed in the early 1960s and 70s with the aim of creating a model that could accurately replicate geographical flows between origins and destinations. Early implementations of these models in retailing struggled due to the limitations of survey data and computing power at that time. This led to a lack of trust in their results due to the inability to accurately evaluate or test their performance. It was thus not until the development of loyalty card datasets that they could be consistent and accurately calibrated and evaluated. Early adopters of this technology were then able to reap the locational rewards in terms of increase profit and market share, but the spillover of knowledge to academia in this regard was initially limited due to confidentiality and competitive pressures. It was not until recently that a series of papers originating from the Leeds School of Geography were able to show how spatial interaction models could be applied in practice in a grocery retailing environment, alongside how these models can be adapted to improve their performance. This series of papers suggested that spatial interaction models could be used to accurately estimate grocery store revenue, even to the level of within a 5% error band consistently throughout the year.

The results from the previous chapter, however, were unable to replicate this level of performance at a regional scale. Thus, this chapter aimed at attempting to replicate the performance by reproducing the conditions under which their models were applied. This was achieved by exploring both the influence of scale and calibration method on modelling performance. The results from the scale analysis, when implementing the model on groups of four and sixteen store subsets, suggested that the performance from the previous papers could not be consistently replicated. This was such that only one out of 47 possible groups for each subset size were able to replicate the error ranges suggested, and that even these groups were unable to maintain this error level consistently throughout the year. The evidence presented here suggests that the spatial interaction model implementation employed by Newing et al. and Waddington et al. is not robust enough to deal with heterogenous conditions consistently at any scale in grocery retailing with the current data and model formulation, even when minimising the potential influence of non-residential demand. This conclusion was supported by the iterative calibration analysis which used a grid search to explore the modelling response across a range of potential parameter pairs in terms of the average trip distance metric. The results from this showed that there was not a single optimal parameter pair that could be replicate the system behaviour, but rather a range of potential parameters. Exploration of the individual store errors across this range showed that this calibration method was unable to improve on the Poisson regression implementation, therefore further supporting the idea that these models could not be reliably employed at a regional scale.

Overall these results therefore support the conclusion from the previous chapter that the spatial interaction model is not appropriate to be applied at this scale of analysis in its current form. Furthermore, they even suggest that the spatial interaction model is not consistent enough even at a small scale to be confidently able to utilise it in practice. This therefore opens up further questions as to whether the model or data could be adapted to resolve these issues and improve overall model performance. These questions are therefore examined in the next two chapters which present the results from a competing destination model, an adaptation of the existing model formulation, the limitations of the current research and how they may be affecting model performance, and the identification of potential future directions that could be explored for spatial interaction models in grocery retailing.

Chapter 7

Alternative Model Implementations

7.1) Overview

The previous chapter discussed how the recent literature implemented spatial interaction models in a grocery retailing context, focusing on a series of papers originating from the Leeds School of Geography. The Chapter then attempted to replicate the results in these papers by adapting the current model implementation in terms of both scale and calibration method used. The results suggested that the with the data available and the current model form, we were unable to consistently replicate the results from previous papers. This therefore supported the idea that the spatial interaction model in its current form and application was unable to account for heterogeneity in behavior and store conditions at a regional scale. It was therefore suggested that alternative forms of models or data could have the potential to resolve some of the issues highlighted. This chapter therefore builds on this by developing and applying a competing destinations model to examine whether the influence of competition or agglomeration may be biasing the results, the implementation of a model that includes store age as an attractiveness factor, and the development of a model focused on large basket shopping behaviour.

7.2) Competing Destinations Model

The previous two chapters showed poor overall fits for the production constrained spatial interaction model when applied at the regional scale, on subsets of stores within a region and when calibrated using the average trip distance metric. These results therefore suggest that the spatial interaction model in its current form and with the data available is unable to accurately estimate total store grocery revenue. Thus, we turn to alternative forms of the model that could improve performance. To this end, a potential factor that could be influencing the results could be competition or agglomeration effects that would affect how much revenue would be attracted to individual stores (Li, 2012). This influence could be explored by adapting the current model implementation to incorporate store accessibility using a form of the competing destinations model that was originally proposed by Fotheringham in 1983. This accessibility measure identifies the accessibility of stores relative to all other potential destinations within a given distance. A parameter can then be calibrated to determine the strength and direction of the influence of either competition or agglomeration in revenue estimation whilst also examining the effects that this has on overall modelling performance.

7.2.1) Model Implementation

The competing destination model was originally developed by Fotheringham in 1983 in response to perceived issues with the existing spatial interaction model. The argument was that the original model implied that the destination chosen by individuals was the result of a single decision making process, for which the final destination would be picked by comparison with all other potential destinations. Fotheringham however suggested that many types of interaction based decisions, including retail destination choice, could be considered the result of a two-stage decision making process instead. This would take the form of an individual firstly selecting a broad region of destinations and then choosing a specific destination within that subset (Fotheringham, 1983). It was believed that such a decision making process would increase the efficiency of individual decision making as not all alternatives would need to be evaluated simultaneously (Fotheringham, 1988). For grocery retailing this would take the form of an individual choosing a broad destination set, such as a group of stores to the North West of their home, and then choosing a single destination from within that subset, rather than selecting from all possible destinations (Guy, 1987). This decision making idea, of a hierarchy of decision making levels, was also seen to extend to other store attributes as well such as format and brand (Recker & Schuler, 1981). Decisions made this way would then influence spatial interaction behaviour because the volume of interactions terminating a single destination would be determined by how many other destinations were in the same broad destination set. Thus, "*the more accessible a destination is to all other destinations in a spatial system, the less likely it is that the destination is terminating point for interaction from any given origin, ceteris paribus"* (Fotheringham, 1983, p. 20).

The result of this two stage decision making process is that as the accessibility of a destination to all other destinations increases, then the volume of interaction termination at that destination would decrease. However, this is based on the assumption that these destinations are competing with each other. In reality destinations could alternatively draw revenue to each other which would create agglomeration effects, increasing the revenue terminating at a single destination (Li, 2012). Either way, if the two stage decision making process did exist in consumer behaviour, then this would suggest that the current gravity model was misspecified because it does not include a variable that explicitly measures the relationship between competing destinations (Fotheringham, 1983; Kerkman, et al., 2017). Thus, Fotheringham adapted a variable that would account for the accessibility of a destination to all other destinations based on the idea of Hansen accessibility (Hu & Pooler, 2002). This was constructed such that the influence of a destination would be proportional to its size, or attractiveness, and inversely proportional to the distance between them. Thus showing a similar relationship to the one assumed by the original gravity model but between destinations.

The first implementation of the competing destinations model took the form of an origin specific production constrained competing destination model as represented by:

$$T_{ij} = Z_i O_i w_j A_{ij}^{\delta_i} \, d_{ij}^{\beta_i} \qquad \text{Eq. 59}$$

Where:

$$Z_i = \frac{1}{\sum_{j=1}^{n} w_j A_{ij}^{\delta_i} \, d_{ij}^{\beta_i}} \qquad \text{Eq. 60}$$

In this specification, A_{ij} represents the accessibility of the destination, j, to all other possible destinations available to origin, i, as perceived by the residents of origin i. This is defined as:

$$A_{ij} = \sum_{k=1}^{w} w_k d_{jk}^{\sigma_i} \qquad \text{Eq. 61}$$

Where σ_i represents the importance of distance in determining the perception of accessibility of the destination to all other destinations. In this scenario, *w* and *n* are not necessarily equal sets as *w* represents the potential competing destination for each destination while *n* represents the potential destinations that each origin may visit. In the original formulation of the model, as devised by Fotheringham, *w* would represent the total number of destinations available to origin *i*, whether or not they were included in *n*. Thus, in the original implementation $w \geq n$.

This model implementation however is not suitable for the application of grocery retail shopping as has been presented in this thesis so far. Firstly, an origin specific model cannot be calibrated using loyalty card data because all the potential origins that could be expected to visit an individual store are not necessarily represented, as loyalty card data is only a subset of the total system behaviour (O'Kelly, 2011). Thus system wide values for σ, δ and β can be calibrated and used. Furthermore. In the spatial interaction model formulations introduced in previous chapters, an attractiveness parameter (γ) is used to account for the influence of size on store attractiveness and hence on the flows from origins to destinations. Thus, this parameter can be added into the above formulation of the competing destination model as well and applied at a system wide scale (Guy, 1987). This model would thus take the form:

$$T_{ij} = Z_i O_i w_j^{\gamma} A_{ij}^{\delta} d_{ij}^{\beta}$$ Eq. 62

Where:

$$Z_i = \frac{1}{\sum_{j=1}^{n} w_j^{\gamma} A_{ij}^{\delta} d_{ij}^{\beta}}$$ Eq. 63

In this specification, A_{ij} represents the accessibility of the destination, j, to all other possible destinations available to origin, i, as perceived by the residents of origin i. This is defined as:

$$A_{ij} = \sum_{k=1}^{w} m_k^{\phi} d_{jk}^{\sigma}$$ Eq. 64

In attempting to calibrate a model of this form, rather than using a calibration method to estimate the parameters of the model, a grid search can be used so as to illuminate the behaviour of the model across the potential parameter space. This was chosen due to the poor performance of the models in previous chapters, suggesting that calibrating a competing destinations model may lead to local optima calibration rather than global optima results (Hu & Pooler, 2002). Furthermore, as with the average trip distance calibration in the previous chapter, it allows for the identification of how modelling behaviour changes across different parameter values. To simplify this exploration, the already calibrated distance decay, β, and attractiveness, γ, can be used to represent the attractiveness (ϕ) and distance decay parameters (σ) in the accessibility measure (Guy, 1987). While individual parameters could be iteratively calibrated, the effect of this is likely to be smaller than switching from the traditional model to the competing destinations model (Thorsen & Gitlesen, 2002). Furthermore it can also be expected that the accessibility interpretation of both attractiveness and distance decay is likely to have the same behaviour as the initial destination choice (Thorsen & Gitlesen, 2002). This therefore reduces the number of dimensions that have to be explored in the model implementation, reducing the overall complexity, while also unlikely to have a considerable effect on model performance compared to a fully calibrated model. These adjustments therefore result in the model as:

$$T_{ij} = Z_i O_i w_j^{\gamma} A_{ij}^{\delta} d_{ij}^{\beta}$$ Eq. 65

Where:

$$Z_i = \frac{1}{\sum_{j=1}^{n} w_j^{\gamma} A_{ij}^{\delta} d_{ij}^{\beta}}$$ Eq. 66

And

$$A_{ij} = \sum_{k=1}^{w} w_k^{\gamma} d_{jk}^{\beta}$$ Eq. 67

Within this application stores within the accessibility set, *w*, are likely to represent the range of destinations that might reasonably be assumed to act with competition or agglomeration forces around each destination. Previous research in this regard has suggested that agglomerative forces can occur between 0.3-0.4 miles and competitive forces between 1-4 miles for grocery retail (Li, 2012). This is based on the idea that agglomerative forces occur when stores are clustered close together, allowing for convenient multi-purpose shopping trips, while competitive forces occur at a greater distance when clustered within the same general shopping area (Li, 2012). Thus, a range of distance thresholds was defined to be explored from 1km to 10km around each destination to cover the complete range of potential agglomeration or competition influences. Furthermore, since there is the potential for either competitive or agglomerative forces, both negative and positive δ values were deemed relevant to explore (Hu & Pooler, 2002). In this case, a positive value would represent agglomerative forces while a negative value would indicate competitive forces.

7.2.2) Model Application

The results of this model application can be seen in Figure 32 below which shows the variation in the performance of the model across the range of distances and δ values. This performance is measured in terms of the mean store error, the standard deviation of store errors and the average trip distance metric. These measures were chosen to represent the modelled system because the aim is to develop a model with an average store error close to zero, with low standard deviation of errors and representing the overall system behaviours (Batty & Mackie, 1972). Each of these aims are thus represented by the mean store error, the standard deviation of store errors and the average trip distance metric respectively. What this figure shows is that the best performing models in terms of mean store error do not necessarily align with the best performing models as measured in standard deviation or average trip distance. In these results, there are two regions where the mean store error crosses a zero value, the first with an accessibility parameter (δ, delta) between 1 and 2 for distance distances greater than 2,000m, and secondly with strongly negative deltas where the distance is less than 3,000m. The existence of these regions suggest that an accurate model shows strong competitive forces at small distances or agglomerative forces acting strongly across a larger range of distances. These results are therefore in contrast to the results seen in the previous literature of agglomerative forces between 0.3-0.4 miles and competitive forces between 1-4 miles (Li, 2012). Importantly however both the lowest standard deviation and the average trip distance metrics occur when the accessibility parameter is close to zero across all ranges of distance. Thus within this model specification there appears a trade-off between accuracy and consistency in terms of store errors, creating two regions of parameters to be explored further. These regions are highlighted by both the red and green boxes in the figure below.

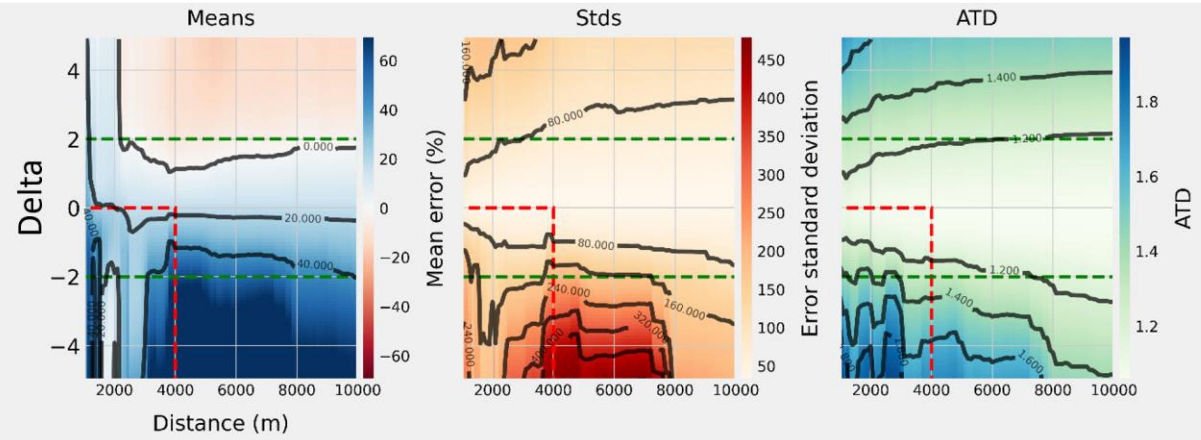

Figure 32 - Competing destination model results across a range of distances and delta parameter values as measured by the mean store error, the standard deviation of store error and the average trip distance metric

The first subset of the model to be explored in more detail is that within the red bounding box. Within this parameter space it can be seen that a mean error close to zero is achieved between delta values -2 to -5 and a distance range of 1,500 to 2,000 metres. While this is suggestive of a potentially accurately fitted model, it can be seen that the standard deviation of model errors and the average trip distance are high for this region (160 and 1.6 respectively). Thus, while these model parameters produce an accurate model in terms of mean store error, they are unlikely to represent a good overall spatial interaction model. This is because the high standard deviation value indicates that there will be a large range of store errors in terms of both large underpredictions and overpredictions, alongside a poor overall fit in terms of overall system behaviour. This subset of modelling results are therefore unlikely to resolve the modelling issues present in the previous chapters.

Therefore, the second subset of models that can be explored is between a delta of -2 and 2 within a distance of 1,000 to 10,000 metres highlighted by the presence of low standard deviation and average trip distance values. What can be seen in this parameter region is that while there are some models that have a mean store error close to zero, as with the previous region analysed, these models also appear to have high standard deviation and average trip distance metrics. Therefore highlighting similar trade-offs in terms of mean store error relative to the range of errors and the ability to replicate system behaviour. What is interesting then however is the region in the centre of the delta values that is consistent across the range of distances which appear to have the lowest standard deviation and average trip distance values. This appears to suggest that the best performing models, in terms of standard deviation and average trip distance metrics, occurs when the delta value is close to zero. In this case a delta value of zero would mean that the accessibility value in the model formulation would become one and thus represent the base spatial interaction model without the influence of competing destination. Such a result would suggest that this form of the competing destination model does not add any explanatory power over and above the base origin disaggregated spatial interaction model.

7.2.3) Individual Model Results

These results therefore, within the green bounding box, can be further examined by selecting the parameter pairings, in terms of distance and accessibility parameters, that produce models with the mean error closest to zero, the lowest standard deviation of errors and an average trip distance metric closets to one. The individual store errors for each of these “optimal” models can be seen in Figure 33 below. From this the mean error model, which has the mean error closest to zero, occurs at a distance of 5.6km and an accessibility parameter of 1.4, suggesting the influence of agglomeration. However it can be seen from the individual store errors and their distribution that there is considerable and consistent evidence of both under and overprediction of store revenue where some stores are assigned almost no revenue. This range of errors therefore suggests that while this is a good model in terms of average store error, it would not be useful in practice and would not represent the total system. In contrast, both the lowest standard deviation model and the lowest average trip distance model have errors that are mostly distributed within a range of -50% to 50%, with a few outliers. This therefore suggests better model fit and consistent modelling errors. These models occur at distances of 7.7km and 8.1km, along with delta values of 0.3 and -0.4 respectively. This therefore suggests that to produce the lowest standard deviation model there is evidence of agglomerative forces up to a distance of 7.7km, while the lowest average trip distance value suggest evidence of competition at distances of 8.1km. Importantly both models perform better at the individual store level than the mean error model.

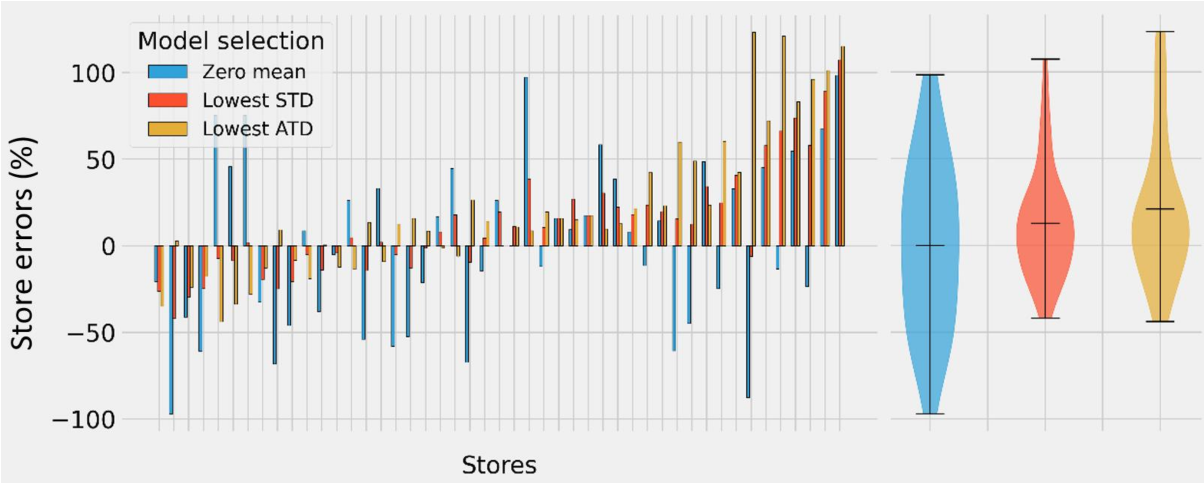

Figure 33 - Individual store errors from the competing destination model implementation with the results from the parameter pairs that produce the model with the store mean error closest to zero, the lowest standard deviation and the mean trip distance closest to 1, alongside their error distributions

Therefore, the lowest standard deviation and lowest average trip distance model can be compared to the baseline model implementation that does not include accessibility. This comparison, in terms of individual store errors, can be seen in Figure 34 below. This figure shows that, compared to the base model, the effect between the lowest standard deviation model and the lowest average trip distance model varies by store and that in some cases these effects are in opposite directions to each other. This is because the lowest standard deviation model derives from agglomeration forces while the lowest average trip distance model derives from competitive forces. For some stores, the difference between the two is minimal, however there are other stores that show large differences in effects. This can be seen by the stores highlighted by the circles in the figure below where the size of the effect varies. For example, for store D the effect of competitive forces increases the absolute percentage error by over 75% while agglomerative forces decreases the error by 50%. In contrast, for Store C the effects are 5% and 25% respectively. This therefore shows that the effect of either agglomeration or competition varies between stores at this scale. While for store A and C agglomeration increases the revenue assigned to these stores, for B and D agglomeration leads to reduced revenue, and vice versa for competition effects. This therefore shows that the influence of competition or agglomeration effects can vary by store, with no convergence towards a zero error. Thus, competition or agglomerative influences are no consistent across the region.

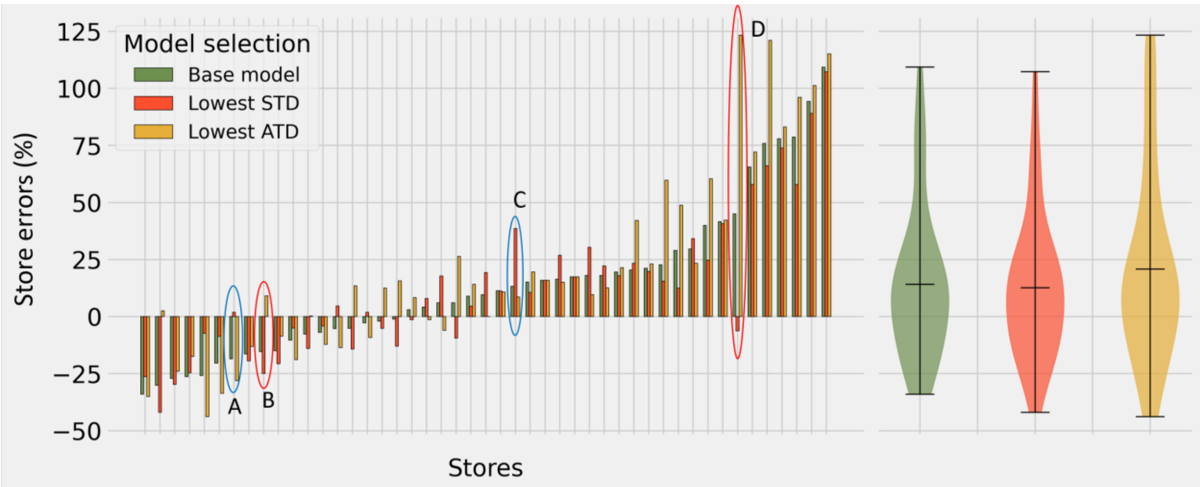

Figure 34 - Base model performance per store compared to the competing destination model implementations with the lowest standard deviation of store errors and average trip distance value closest to 1

Nevertheless, while both models behave differently, neither leads to considerable improvements over and above the existing base model. This is highlighted by both the distribution shown in the figures above and the results in Table 10 below. Interestingly, from these results it can be seen that the lowest standard deviation model shows improvements over the base model as given by both the mean store error and the standard deviation of errors. This may therefore suggest that agglomeration may be influencing modelling results and that further exploration may be required. However, these improvements are relatively small compared to the base model and still do not resolve the issue of large variances of error at the regional scale. Furthermore, the distance of agglomerative forces appear over much larger distances than suggested in the previous literature (Li, 2012). Thus, this does not suggest that the competing destinations model, in its current form, would resolve the issues inherent in the spatial interaction model at the regional scale, primarily because of the diverging effects across stores at this scale.

Table 10 - Model performance metrics for the competing destination models

Model	Mean store error (%)	ATD	Standard deviation (%)
Base spatial interaction model	14.12	1.06	33.25
Zero mean model	0.00	1.15	48.90
Lowest Standard deviation model	12.68	1.07	31.28
Lowest ATD model	20.86	1.06	41.51

7.2.4) Conclusion

Models presented in the previous chapters had shown mean store errors that deviate from zero and have large variances in individual store performances. This meant that such models would be unlikely to be used in practice at this scale. The range of these errors however suggested that a factor may potentially be missing from the current modelling specification. Thus, in this section a competing destinations model was implemented using a production constrained competing destinations model with system wide parameters (Guy, 1987; Thorsen & Gitlesen, 2002). For this, a grid search was used to determine the optimal parameter pairs in terms of both distance (m) and the accessibility parameter (δ). Across this parameter space it was identified that some pairs produced a model that had a zero mean store error, however such models were associated with high standard deviations and average trip distance metrics. Other models that produced that produced the lowest standard deviation or average trip distance metric, also did not lead to considerable improvements over the existing spatial interaction model implementation. It was shown that this was because the effect of competitive of agglomerative parameters had diverging effects on stores, whereby some stores that where already under or overpredicted became more so. Therefore, with either competition or agglomeration there was not seen to be any convergence towards a zero store error. These results therefore suggested that the underlying variation in store conditions at the regional scale could not be resolved through the application of a competing destinations model. While the model may be improved however with simultaneous calibration of all parameters, previous research has suggested that doing so is unlikely to lead to any greater improvements than those already seen (Thorsen & Gitlesen, 2002). Thus, this was not explored further. Nevertheless, future research may look to build on the results presented here by calibrating all parameters together (Thorsen & Gitlesen, 2002), or developing an origin specific model (Guy, 1987) which may lead to further incremental improvements in model performance.

7.3) Store Age Model

From the above analysis it becomes clear that adding in an accessibility value and parameter is unlikely to resolve the issues of model performance highlighted in previous chapters. It therefore becomes necessary to examine whether alternative modelling formulations could instead perform better. Since we are using a production constrained spatial interaction model, additional factors could focus on either the travel cost or store attractiveness values. To this end, it has been previously acknowledged that store attractiveness is likely to be influenced by a variety of factors beyond store size already integrated into the model (Newing, et al., 2020). This could include factors such as distance to the street, store frontage, other services located in the store and the age of the store (Birkin, et al., 2017). It was identified in chapter 5 Section 5.5. that store age was seen to be consistently negatively correlated with individual store errors across all model specifications and regions. This therefore led to the suggestion that a younger store may be less attractive than its size would suggest and vice versa for an older store. Thus, it was argued that this could potentially be due to consumers having more information about an older store in terms of price, selection and travel time. Thereby drawing parallels with the influence of size in the original modelling application. This is such that size is equated with consumers being able to access all goods they are interested in (Newing, et al., 2018), while the older a store is then the more information consumers may have about whether they can purchase all their items from the destination.

7.3.1) Model Implementation

Computational limits initially restricted the integration of store age into a regional scale model, however this could be explored through a model on a subset of stores. This would reduce both the incidence of missing data, in terms of stores for which data is not available on their opening, and allows a model to be calibrated that includes store age as an attractiveness factor. The subset identified for this implementation is based on the 16 store well performing subset that was highlighted in Chapter 6 Section 6.4.4 and Appendix C. The number of 16 stores was chosen as it allows for the model to be consistently trained cross the year whilst previous results in Chapter 6 have also shown that the errors, in terms of the range and average error, are consistent at this scale. The specific stores chosen for this evaluation are identified in Figure 35 below which shows their location in Region 2.

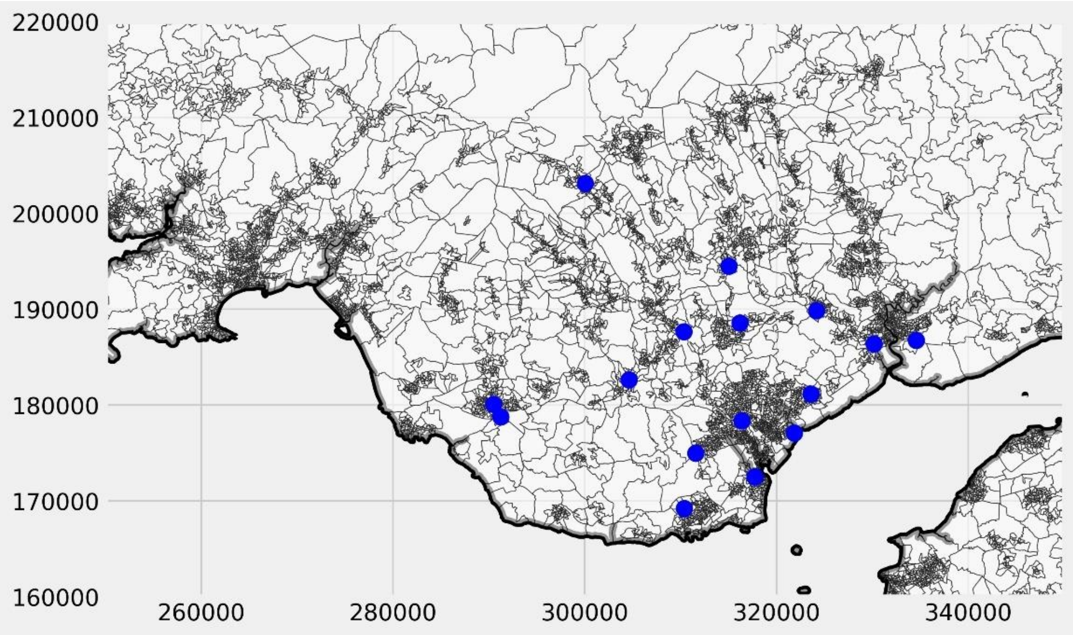

Figure 35 - Sixteen subset for store age model analysis

Due to the correlation results presented in Figure 15 and the modelling comparisons in Chapter 5 more broadly, the non-disaggregated spatial interaction model was chosen to be implemented. Of this modelling specification, the inverse power decay model was used due to computational resources limitations, whereby the exponential distance decay model could not be consistently applied across the whole year. Thus, the original non-disaggregated model takes the form:

$$T_{ij} = A_i O_i W_j^{\gamma} d_{ij}^{-\beta} \qquad \text{Eq. 68}$$

Where:

$$A_i = \frac{1}{\sum_j W_j^{\gamma} d_{ij}^{-\beta}} \qquad \text{Eq. 69}$$

As introduced in Chapter 4, Section 4.5.2. However, the above formulation does not account for the age of the store (in months). Therefore we can add a factor G_j that represents the age. This changes the model to take the form:

$$T_{ij} = A_i O_i W_j^{\gamma} G_j^{\tau} d_{ij}^{-\beta} \qquad \text{Eq. 70}$$

Where:

$$A_i = \frac{1}{\sum_j W_j^{\gamma} G_j^{\tau} d_{ij}^{-\beta}} \quad \text{Eq. 71}$$

As G represents the age of the store in months, while τ represents the attractiveness of store age for consumers based on loyalty card data. Both models can therefore be implemented on the sixteen store subset identified above[13].

7.3.2) Loyalty Card Performance

The first stage of model implementation is to see how well the model is able to replicate total store loyalty card sales. This is achieved by aggregating the loyalty card sales predicted by the model to each store and then comparing them to the total loyalty card revenue. The results of this can be seen in Figure 36 below. This figure shows the Age model implementation against the Original model in terms of individual store errors, along with a line of equality. From this figure it can be seen that there is a strong positive correlation between the Original and Age model errors, as indicated by the 0.98 Pearson correlation value and with most points lying close to the line of equality. Although there appears to be two potential outliers, the strength of this correlation suggests that including the age of the store does not considerably alter the performance of the model in being able to replicate total loyalty card revenue.

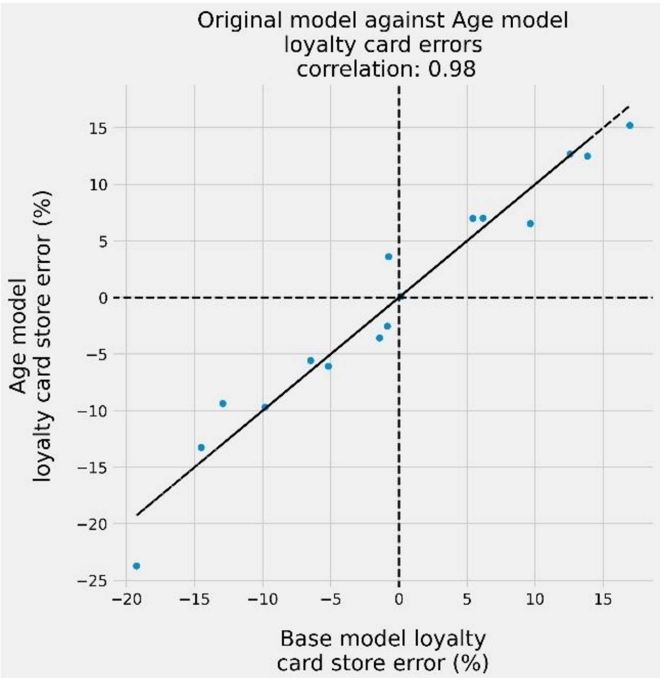

Figure 36 - Base model and Age model loyalty card store errors

[13] The original model was implemented with the SpInt module (Oshan, 2016) while the store age model was implemented using Poisson Regression from Statsmodels.api.

7.3.3) Total Revenue Performance

As with previous model implementations however the main concern is how do these model implementations scale up to predict total store revenue. In doing so the Original model implementation identified 87 total stores (including the 16 modelled) that the surrounding output areas could reasonably be expected to visit. However, of these 87 stores, data was only available on the opening date for 81 stores. Therefore, two versions of the Original model are evaluated: the Original model estimated with the full 87 store subset of potential competitors, and a New model which uses only 81 stores for which have a store opening date. Both of these results can then be compared to the Age model which is estimated with the 81 stores. The results of this can then be seen in Figure 37 below which shows all three modelling implementations in terms of their loyalty card store and the total store revenue error across the 16 stores form our partner organisation.

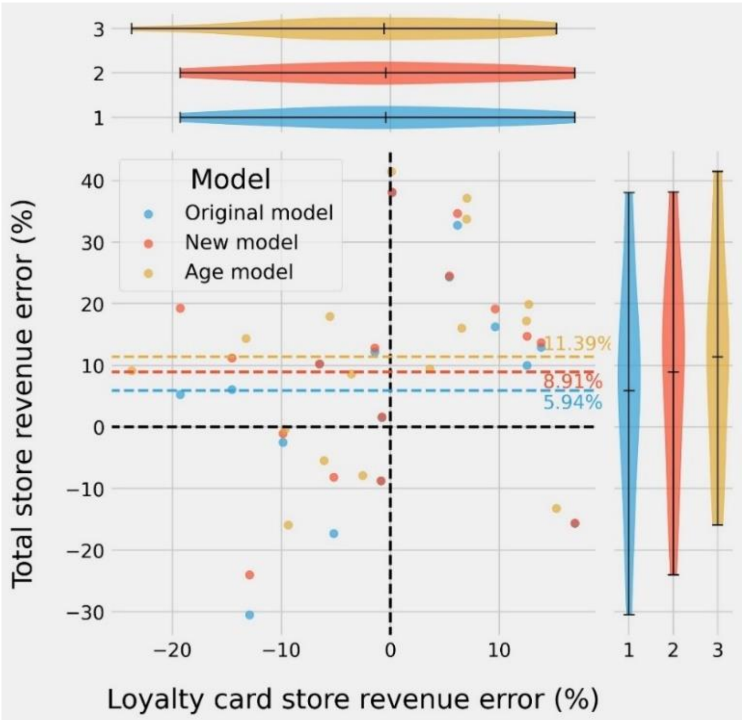

Figure 37 - Age store modelling results for the original model, original model with limited competition and the age model

Firstly, the figure above shows that the Age model loyalty card errors have a similar distribution to the Original model implementation. This therefore supports the strong correlation that was shown in Figure 36 above. The main difference between these distributions however is that the Age model distribution is shifted slightly to the left, driven primarily by a reduction in the maximum error and minimum error. However, the variance in mean error is small suggesting that the effect of these changes is minimal on the overall distribution. The main difference between the model implementations appears when estimating the total store revenue. Firstly, comparing the Original model and the New model shows the effects of removing six potential competitors from the dataset (those that are used calculate A_i in Eq. 69). From this it can be seen that as a result of removing potential competitor stores, the mean store error increases from 5.94% to 8.91%. This change is driven by an increase in the store errors from the stores that were previously underpredicted. This is in line with what would be expected by reducing the amount of potential competition within the model, increasing the revenue assigned to our partner stores. The increase in mean error away from zero however suggests that this does not improve modelling performance and that the original model is more in line with an ideal implementation. When then comparing the New model to the Age model, it can be seen that adding in store age as a variable further increases the average error away from zero and that this is driven by a shift upwards of the entire distribution relative to the New model. Therefore, with age included in the model more revenue appears to be assigned to our partner retailer than the competitors, but that this pushes the modelling results further away from the ideal distribution of mean zero error and low variance. Thus, the inclusion of this factor in the regression specification has not lead to considerable improvements in model performance. Indeed, arguably it had led to worse performance than the original model due to both lack of data on store's age and how the Age model has distributed revenue to our partner organization.

7.3.4) Store Age Relationship

It thus becomes necessary to examine how this affects the store age relationship with individual store errors. Figure 38 below addresses this question by showing the total revenue error for each model against store age, along with their lines of best fit and strength of the correlation. Surprisingly this figure shows that, in contrast to the results seen in Chapter 5 Section 5.5, the original model shows a positive correlation between total revenue error and store age. This therefore suggests that sub-models implemented within each region may have different correlations and relationships than the complete regional scale model. This is because models calibrated on subsets are likely to have different parameters that when scaled up affect the performances at the individual revenue scale, as seen in Chapter 6. In this case, these correlations suggest that this model may be overpredicting an older store and underpredicting a younger store. The direction and strength of this correlation is consistent across model specifications with the Age model showing a similar strength to the original model. This suggests that adding in an age parameter to the model strengthens the correlation as opposed to removing it as a potential factor affecting modelling performance.

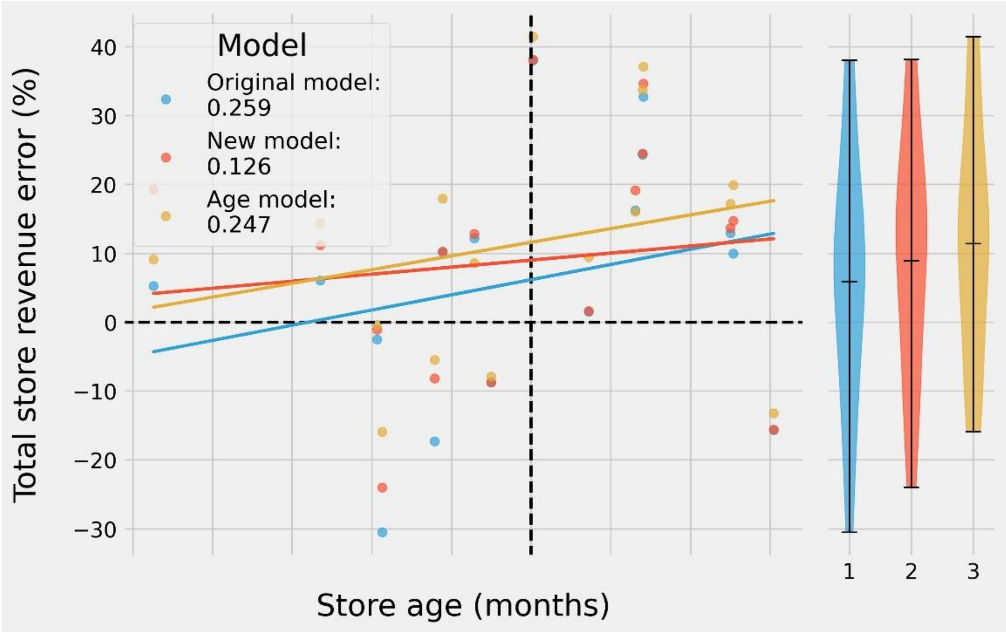

Figure 38 - Store age against total revenue error for all three models

A reason for this could be seen in Table 11 below which shows the parameters for both the Original model and the Age model specifications. From this it can be seen that in the Age model the Tau parameter (τ) is positive which suggests that an older store is more attractive than a younger store within this modelling specification. Thus, it would lead to more revenue assigned to an older store, *ceteris paribus*. Therefore, a positive correlation with store age has been strengthened by the inclusion of this factor in the model specification. Furthermore, the size of the attractiveness parameter is larger in the Age model. This therefore suggests that the increase in revenue assigned to our partner stores in the Age model is due to our partner organization having older and larger stores than their competitors. These conclusions however are likely to be affected by the sample that is examined here whereby it can be seen there is considerable variation around the line of best fit for all model specifications in the figure above. Thus to verify these claims it would be beneficial to repeat the model over the year and also to potential repeat it on further subsets of 16 stores.

Table 11 - Inverse power decay subset model parameters

Parameter	Original model	Age model
Beta (β)	-4.526	-4.520
Gamma (γ)	1.869	1.946
Tau (τ)	N/A	0.161

7.3.5) Yearly Application

The results from the model presented above can be extended by implementing the model across every week in the year. The outcome of this, for all three model implementations, can be seen in Figure 39 below. This figure shows the total revenue errors across the whole year in terms of the range of store errors and the mean store error. What this shows is that the results seen and discussed above are consistent across the whole year. Specifically, that the increase in mean store error and the upwards shift in distribution from the Original Model to the New Model and then finally to Age model is consistent across all weeks of the year. It can be seen that the majority of the increase in mean error across these models is driven by an increase in the minimum error, although for the Age Model there is also an increase in the maximum error as well. Therefore, the results from above are not just the result of a model trained on a single week, but are consistent across the year for each model implementation.

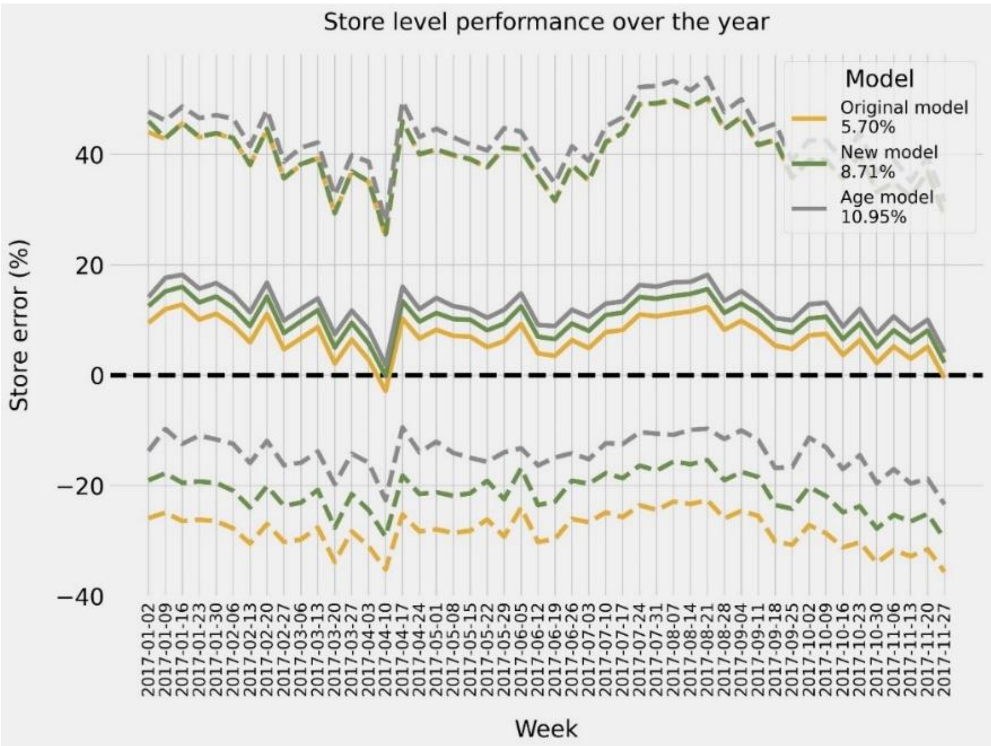

Figure 39 - Total store revenue errors across the whole year for the original model, new model and age model

7.3.6) Model Replication

The model implementations were also replicated across all 47 possible groups of 16 geographically related stores within the region[14]. The results from these model implementations can be seen in Figure 40 below with the group presented in the previous results highlighted. This figure shows the distribution of mean store errors across all three model implementations and each group of stores, including the distribution of those mean errors and the associated store age parameters (Tau). From this it can be seen that the general influence of including store age as a parameter within the model is consistent across most of thegroups, whereby there is a positive store age parameter which leads to increases in revenue assigned to our partner organisation's stores. This can be seen by the greater mean errors of most of the Age models as opposed to both the Original Model and the New Model for each group. This is supported by the distribution of mean errors Figure 40d that shows that the distribution of Age Model mean errors is greater than either the Original Model or the New Model. These results are due to the store age attractiveness parameter values that are seen in Figure 40a which shows that out of 47 groups, only 5 had negative age parameter values. This therefore suggests that in most groups, the older a store was the more attractive it was to consumers within the anonymised loyalty card data. The size of this effect increases with the size of the parameter whereby greater increases in mean store errors from the New Model to the Age Model can be seen on the right hand side of Figure 40c. Furthermore, the increase in revenue assigned to stores of our partner organisation as a result of these parameter values suggests that stores from our partner organisation are older than those of the competitors.

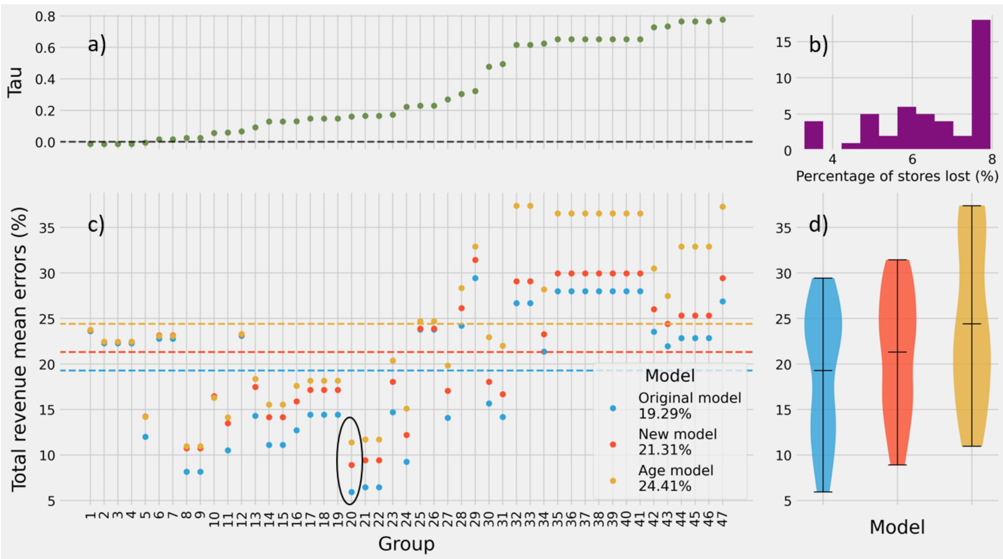

Figure 40 - Store age model results across 47 different groups of 16 stores in Region 2. a) Tau parameter values for each group, b) percentage of stores without store opening information, c) mean store error from the groups of stores across all three models, d) distributions of mean store errors across each model

[14] The identification of these groups of sixteen stores is explained in Chapter 6, Section 6.4.1

The increase in mean errors from the New to the Age model is consistent over and above the increase in mean errors from the Original to the New model as seen in Figure 40c. The first increase in mean errors, from the Original to the New model, however is due to the loss of data on competing stores, thus not being included in the estimation of revenue calculation. The distribution of the percentage of stores without store opening information from each group can be seen in Figure 40b. It can thus be seen that the number of stores lost as a result of the New model implementation ranges between 3% and 8%, therefore leading to increases in revenue assigned to our partner organisation stores across all New model implementations. The almost uniform further increase in mean store error from the New model to the Age model across the groups of stores, however, suggests that it is not solely the reduction of competing stores that is driving the shift in mean store error and the increase in revenue assigned to stores of our partner organisation. These results therefore support the conclusions drawn from the analysis of the single group of sixteen stores presented above.

7.3.7) Conclusion

These results therefore show the implementation of a spatial interaction model that integrates store age as a measure of store attractiveness. This implementation was based on previous results introduced in Chapter 5 that showed a clear and consistent correlation between store age and store revenue estimation errros across all three regions and model specifications. The idea was that the integration of this factor could potentially reduce the strength of this correlation and lead to improved modelling results in terms of a mean error close to zero and reduced store error variance. The implementation of this model however was limited by both computational resources, thereby restricting the scale of the analysis and model used, alongside data limitations on store opening dates. Nevertheless, the results showed that for this subset of 16 stores, integrating store age did not alter the original models relationship between store errors and age, and pushed the store errors even further away from the ideal distribution expected. This was driven by a positive store age attractiveness parameter which suggested that the older a store was, the more attractive it would be, potentially linking to the idea of more information by consumers (Newing, et al., 2018). Furthermore, the results suggested that stores from our partner organization, at least in this regional subset, were older and larger than their competitors due to increase revenue assignment in the Age model. These results were supported by the replication analysis that extended the model across the whole year and re-implemented the models across 47 different groups of sixteen stores within the region.

However, this exploration could be argued to be limited by the data available and computational resources. This means that future evaluation and exploration could extend this analysis through adding in new data on store opening times, developing an exponential distance decay form of the model and adding extending the scale of the model application. The aim of this would be to extend the research presented above and to verify the conclusions presented. Nevertheless, this initial exploration suggests that the influence of store age is not driving the overall poor modelling performance seen in previous chapters and thus further exploration is beyond the scope of this thesis.

7.4) Large Shopping Basket Analysis

The results presented above and in previous chapters have thus highlighted the poor performance of current spatial interaction model specifications in terms of modelling total store revenue for all large stores from our partner organisation. It could therefore potentially be argued that changes in behaviour by consumers has meant that the gravity model, in its current form, is no longer relevant. This includes changes in consumer lifestyles such as working more hours per week, travelling further and longer for work, and changes in work location that have necessitated changes away from the weekly regular shop that characterised behaviour up until the early 2000s (East, et al., 1994; Popkowski Leszczyc, et al., 2004). Instead, new shopping behaviours have arisen including increased convenience shopping (Buckley, et al., 2007; Hallsworth, et al., 2010), characterised by smaller travel times, active modes of transport, shopping more than three times per week and multi-purpose shipping trips (Elms, et al., 2010). This could thus be seen as a departure from the original assumptions of the gravity model of regular, defined purposed shopping made up of large baskets and travel by car (Waddington, et al., 2018). Retailers themselves noticed these trends and responded by building smaller format stores, with convenience stores growing at a greater rate than any other format from 2003 to 2012 (Hood, et al., 2015). For which, it has been previously acknowledged that gravity models in have found it difficult to accurately estimate revenue for these types of stores (Waddington, et al., 2019).

The rise of convenience shopping has also coincided with the increasing influence of non-residential populations (Birkin, et al., 2017; Waddington, et al., 2019), multi-purpose and multi-destination shopping (Brown, 1992; Arentze, et al., 2005), and of e- and m-commerce (Kirby-Hawkins, et al., 2019) on shopping habits and revenue distribution. Therefore also showing a shift away from the behaviours and habits that underlie the implementation of spatial interaction models used in this thesis. Thus, despite consistent use of spatial interaction models in the early 2000s and 2010s (Mendes & Themido, 2004; Reynolds & Wood, 2010), they may no longer be able to accurately model these new consumer behaviours. One potential avenue to explore however is whether a subset of the population can be identified from the anonymised loyalty card data that still behave in line with the assumptions of the spatial interaction model. That is, identifying consumers who shop regularly, travel by car and have large baskets of shopping for which there is likely to be a large penetration of loyalty cards. It is to this subset of behaviour that the following section aims to attempt to model.

7.4.1) Large Basket Data

The basis for this analysis is to be able to identify consumers that behave in line with the assumptions of the spatial interaction model. This is such that consumers undertake single-purpose trips that originate from home, while travelling by car and purchase regular weekly large baskets in a single shopping trip. These consumers could be identified by assuming that large shopping baskets in terms of both value and number of items are likely to represent consumers that are undertaking a shopping trip in line with expected behaviours. This would thus be compounded by high levels of loyalty card penetration would also indicate regular customers that are loyal to the brand or store (Newing, et al., 2013; Newing, et al., 2014). This subgroup of population was identified from anonymous loyalty card data as those with a single basket of value greater than £40 and more than 10 items, based on analysis performed by our partner organisation. This subset of behaviours relative to the total anonymised loyalty card data for Region 2 can be seen in Figure 41 below.

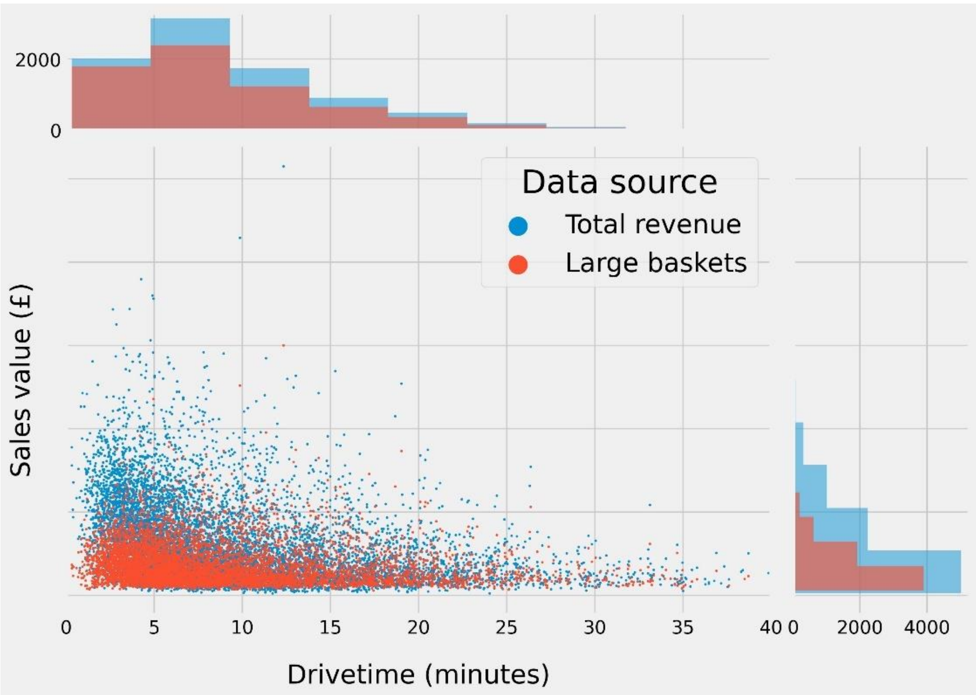

Figure 41 - Sales value over distance in Region 2 for the total loyalty card data and those representing large baskets

What this figure shows is the sales value per output area for a single week to each store over drivetime for the anonymised loyalty card data. From this figure it can be seen that while the large basket data shows a similar distribution to the total revenue over drivetime, there is a notable difference in the distribution of sales value per output area .Importantly, while there are fewer output areas in this dataset due to data regulations, each output area is also associated with less revenue. This is because the large basket data represents only a subset of the total flow data, as only sales that have large baskets and value. This can also be seen to affect the distance decay relationship whereby the peak in sales value relative to all other flows appears smaller and occurs at a greater travel time than the total sales data. This then translates into a gentler slope of sales value over drivetime, potentially meaning that consumers who have large baskets may be willing to travel further to shop. On the other hand it may also mean that those that travel shorter distances are those that are characterised by convenience shopping with smaller baskets.

7.4.2) Model Implementation

The potential differences in behaviour between the total anonymised loyalty card data and the large basket data can be identified through the application of the production constrained spatial interaction model with exponential distance decay. For this, both the system wide and origin-disaggregated models, as introduced in Chapters 4 and 5 respectively, can be trained on the large basket loyalty card data and used to predict total store revenue in terms of large basket sales only. This means that the total store revenue in this model represents only that which derived from large basket sales and that the estimated revenue available has to be scaled down to estimate only spending on large baskets. The latter is achieved through calculating the percentage of total grocery sales in the region represented by sales at large stores with large baskets, then multiplying estimated revenue available by this percentage. This is such that:

$$O_i^{kt} = e^{kt} n_i^{k2011} l \qquad \text{Eq. 72}$$

Where:

$$l = \frac{Total\ revenue_{region}}{Large\ basket\ revenue_{region}} \qquad \text{Eq. 73}$$

Such that *l* represents the percentage of total sales in the region that come from large basket spend at large stores. An alternative methodology, however, is rather than estimating the revenue available from each origin, is to scale the loyalty card predictions to estimate total revenue. This is such that:

$$\hat{R}_{store} = \frac{\widehat{LC}_{store}}{P} \qquad \text{Eq. 74}$$

Where $\hat{R}_{store}$ is a stores total predicted large basket revenue, $\widehat{LC}_{store}$ is the estimated large basket loyalty card revenue and P is the loyalty card penetration for large basket sales in the region. This alternative therefore scales up the loyalty card flow predictions to model total store revenue.

7.4.3) Modelling Results

The results of this model implementation can be seen in Figure 42 below for both the system wide and origin-disaggregated implementations of the model. This figure shows the distribution of store errors in the region and the percentage of stores within a +/-15% error bound. From this figure three main insights can be drawn. Firstly, in line with the results presented in Chapter 5, there are no clear discernible differences between the system wide implementation of the model and the origin disaggregated model in terms of the distribution of errors or the mean store error. This suggests that the added complexity of the origin disaggregated model has not resulted in respective increases in performance and thus the system wide model implementation can be focused in for further analysis.

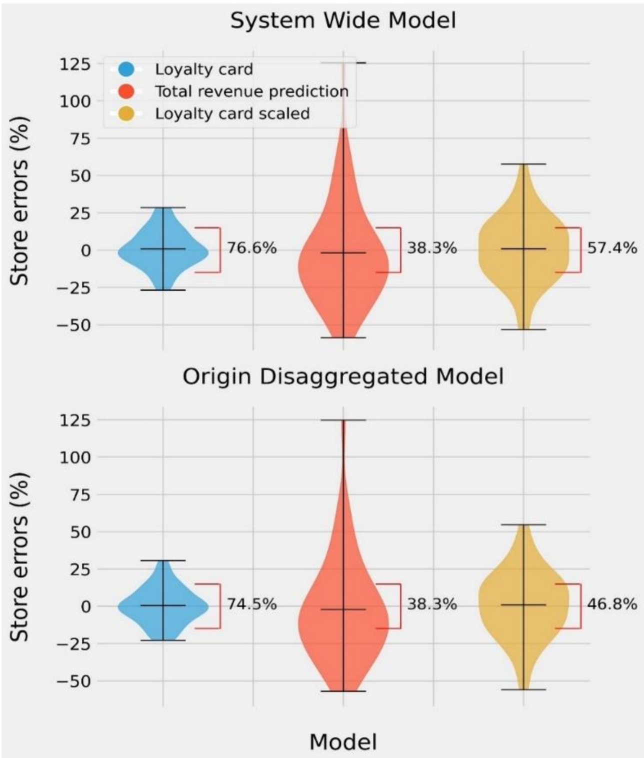

Figure 42 - Store error distribution for 1) The prediction of total store loyalty card revenue, 2) The total store revenue predicted by estimating the total revenue available from each origin, 3) Total store revenue predicted by scaling up the loyalty card flow predictions

The second insight is that both the range of loyalty card and total revenue errors do not show considerable improvements in performance relative to the results presented in the previous chapters. Notably, the range of errors for modelling both loyalty card and total revenue prediction is similar, if not greater, than the models that predicted total store grocery revenue seen in Chapters 5 and 6. This therefore suggests that even when attempting to model a subset of behaviour that aligns with the assumptions of the spatial interaction modelling implementation, the model is unable to replicate this behaviour consistently at the regional scale. Therefore, supporting the results and conclusions drawn from previous analysis in this thesis.

Thirdly however, it can be seen that the predictions based on scaling up loyalty card revenue to predict total store revenue performs better than the predictions based on estimated total revenue available. This is shown by the higher percentage of stores within a +/-15% error bound and a smaller range of errors. Notably, the scaling up methods performance is closer to the accepted standards of our partner retailer. This therefore suggests that there are issues in the traditional methodology of predicting total store revenue using the expected distribution of available revenue, especially for large format stores and modelling large baskets. This could be due to the usage of ONS household estimate expenditures, the scaling of revenue to large basket revenue not accounting for store distribution or the usage of data from our partner retailer to determine the revenue scaling percentage[15]. Thus, suggesting that the modelling performance from previous chapters could potentially be improved by more accurately estimating the total revenue available to spend in the region. However, this methodology of scaling up loyalty card spend cannot be consistently used in practice to estimate store revenue as it requires existing loyalty card data; data that is not available for new large stores or estimating the future.

7.4.4) Yearly Application

Nevertheless, the application of this model can be extend over the whole year to ensure that these results are robust across all weeks in the year. In this regard, the first result to examine is how the modelled behaviour for the large basket data is different to the total loyalty card data in terms of the calibrated parameters. This can be seen in Figure 43 below which shows both the distance decay and attractiveness parameter values from the non-disaggregated large basket data and non-disaggregated total revenue data. What can be seen from this figure is that, apart from a few anomalous data points, the large basket data shows a larger absolute distance decay value and a larger store attractiveness value. This therefore suggests that distance is more of a deterrence to large basket shoppers than in the whole regional dataset, but that large stores are more attractive to large basket shoppers. While the former result is surprising in that it would be expected that large basket shoppers would be expected to be willing to travel further for their shopping, the latter is not as larger stores are more likely to have all the products needed for a large basket shop, thus being more attractive.

[15] The influence of these factors are discussed in more detail in Appendix D, Revenue Estimation

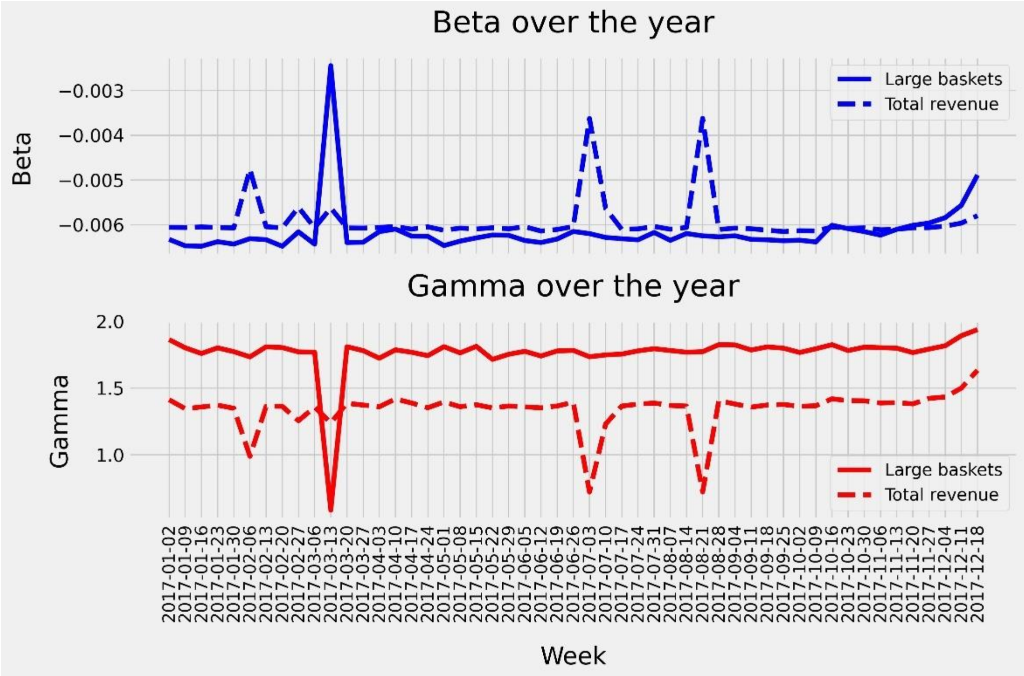

Figure 43 - System wide parameter values from the large basket and total revenue loyalty card data over the whole year

This behaviour then translates to the average store errors for both the system wide and origin disaggregated models in terms of the ability to predict total revenue when the revenue is predicted based on estimated revenue available or when the loyalty card revenue is scaled up at the individual store level. The results of this can be seen in Figure 44 below which shows the average store errors. From this there are two important results. The first is that, unlike modelling total store grocery revenue in Chapter 5, the estimation of total revenue based on estimated expenditure in each region shows consistent underprediction throughout the year. This suggests that the estimated available revenue for large basket spend is underpredicted throughout the year or that our partner retailer is overperforming relative to the behaviour predicted by the model. The former is more likely which showcases the difficulty of accurately estimating the total revenue available to a subset of the population.

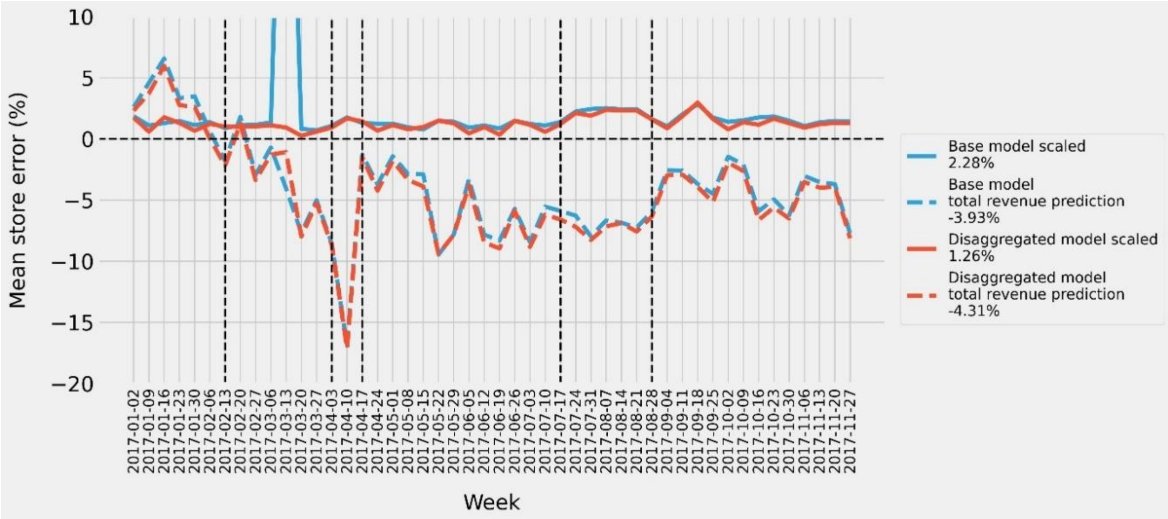

Figure 44 - Base and origin-disaggregated model average store error over the years for predicting total large basket store revenue when scaled up and when estimated using distributed revenue

The second main result is that while there is variation throughout the year for the total revenue prediction models, there is relatively little variation in the average store error for the scaling of the loyalty card predictions. This is because the scaling percentage for the loyalty card data is recalculated every week and therefore is able to account for the variance in the level of loyalty card penetration throughout the year. In contrast, the underlying estimated available revenue does not vary because of the same ONS estimates being used each week. This therefore suggests that the variance in model performance for the total revenue prediction model is due to variances in total basket revenue throughout the year, which indicates evidence of seasonality or non-residential demand changes throughout the year in terms of loyalty card penetration. An example of this can be clearly seen during the Easter holidays where there is a considerable dip in model performance, indicating that expenditure on large basket increases over this period relative to the rest of the year from non-loyalty card sales. This therefore aligns with the results presented in Chapter 5, Section 5.6, and highlights the issue of using only estimated residential demand in estimating total store revenue.

7.4.5) Conclusion

This analysis has therefore attempted to identify whether the performance of the production constrained model at a regional scale could be improved by developing a model on a subset of loyalty card data. This subset was expected to be able to better represent the behaviour that underly the assumptions of the spatial interaction model by identifying large basket sales based on a basket of more than £40 and 10 items. It was expected that this consumer subset would represent consumers who travel to large format stores by car, regularly, for a single purpose grocery shop.

What these results show however is the complexity that underlies the model implementation and attempting to estimate the revenue available to assign to stores in the region. Notably, the implementation of this model form has not led to improvements in model performance at either loyalty card sales or total revenue scale in terms of the distribution of store errors. This therefore shows that the current model formulation, methodology and data is still unable to accurately predict revenue on this scale, even based on a subset of the population that is expected to better align within the spatial interaction modelling assumptions. Nevertheless, what is also shown is that modelling performance is likely to be affected by the scaling methodology used to predict total store revenue. This is because simply scaling up the loyalty card predictions shows a considerable improvement over predicting total store revenue based on estimated available expenditure. Thus, highlighting that the model performance is likely affected by the way in which the store revenue is predicted based on the revenue available at each origin. Future research could this build on these results by identifying the sensitivity of the modelling outcomes to estimation of total revenue available in each output area and region.

7.5) Conclusions

The models presented in this chapter have attempted to resolve some of the performance issues highlighted in the previous three chapters in terms of the spatial interaction model on the data that is available. This has been undertaken though the development and application of a competing destinations model, a model that integrates store age into the model formulation and a model that was applied to a subset of loyalty card data based on large basket sales. The chapter began with the adaptation of a competing destinations model which aimed to identify whether the influences of competition or agglomeration were affecting model performance. The results however showed only marginal improvements in the model performance and that the effect of either competition or agglomeration at this scale resulted in diverging store errors away from a zero error. This therefore further highlighted the inability of spatial interaction models to account for variances in store conditions at this scale, and that as such as more local model may be more appropriate to estimate total store revenue.

The second model therefore developed was one that included store age as a variable based on the correlations identified in Chapter 5, Section 5.5. This model was applied to a subset of sixteen stores in Region 2. The results showed that, for the subset of sixteen stores chosen, integrating store age did not alter the original models relationship between individual store errors and age at this scale. Furthermore, the distribution of errors was pushed further away from the ideal distribution centred around a zero mean error. These results therefore suggested that our partner organisations stores in this group were both older and larger than competitors stores within this sub region. These results were therefore driven by a positive store age parameter which suggested that the older a store was, the more attractive it was to consumers. This could thus suggest that the older a store is, the longer consumers have to gather information about that store or to increase loyalty to the brand. Further exploration could expand the model implementation at a greater scale to see how this relationships changes.

Finally, the chapter proceeded on the suggestion that grocery shopping revenue is made up of a variety of different consumer behaviours and that these behaviours have been moving away from the traditional assumptions of the gravity model in recent years. It was thus suggested that the gravity model may not be able to accurately represent this shift in consumer behaviour and activities. On this basis, a model was applied to a subset of loyalty card data for large baskets. The aim of which was to see whether a subset of consumers could be identified and modelled that aligned with the original assumptions of the spatial interaction model in the form of regular weekly shops undertaken at large stores and travel that was done by car from the home. The results however did not show improvements in modelling performance, but did highlight potential for future research to explore the influence of origin revenue estimation and how sensitive modelling outcomes were to output area revenue assignment.

This chapter therefore identified and implemented potential alterations to the underlying spatial interaction model that were believed to potentially lead to improvements in performance. What the results showed however that when accounting for the influence of competition, agglomeration, store age or consumer behaviour the spatial interaction model in its current form was still unable to consistently or accurately estimate total store grocery revenue at the regional scale. These results therefore support the conclusions from previous chapters that the spatial interaction model in its current form, data and methodology may no longer appropriate. Instead, alternative methods and data may be sought to improve modelling performance. The identification of potential way forwards for the literature therefore are discussed in the next chapter.

Chapter 8

Future Modelling Implementations

8.1) Overview

A recent paper by Rowe et al. (2022) suggested that spatial interaction models are a core tool in spatial data modelling and thus argue for more investment in both teaching and developing the models. The results of the previous four chapters show that spatial interaction models, in their current format and data, are limited in their ability to accurately model grocery retailing in the UK. The question thus becomes where the investment should be aimed in order to improve the performance of spatial interaction models in the future. This chapter therefore seeks to discuss this question, identifying avenues of future exploration in terms of model formulation, data, evaluation, open source implementation and even models other than spatial interaction models.

8.2) The future of Spatial Interaction Models

Based on the results from the previous three chapters, the question then becomes where does the implementation of spatial interaction models, especially in relation to grocery retailing, go from here? There are many potential avenues of research that could be explored which could in theory lead to improvements in accuracy and reliability of spatial interaction models. The important task is identifying which of these are most likely to lead to these improvements and hence where resources should be directed. These avenues include the development of new modelling techniques, formulations and adaptations of existing models from the last 20 years. Such new implementations aim to resolve some of the issues of using the Wilsonian form of the spatial interaction as implemented so far. However, one of the main issues of these new implementations is often how to operationalise them, especially in relation to retailing, with a lack of available code, resources or data to implement them. This issue is often raised alongside the lack of consistent evaluation of their relative performance, which has so far been unable to clearly identify which models are most suitable under different conditions. Secondly, there is also the development of "big data" sources, including the use of existing loyalty card data, credit cards, mobile phone, footfall and open source image datasets that could be built into the spatial interaction model to improve performance. This encompasses data on flows, attractiveness and emissiveness, especially in relation to grocery retail stores, which could more accurately reflect these underlying attributes that current data. However, issues of confidentiality, accessibility and privacy often presents a barrier to their implementation and adoption in practice which can limit their evaluation.

Alongside the development of new models and data sources however, future research must also tackle two key challenges of current model implementations. The first is the development of a consistent method of evaluating spatial interaction model. This is such that there should be a consistent and clear method for identifying how new datasets or model formulations improve spatial interaction model performance and under what conditions these improvements occur. This would facilitate the advancement of the literature with a clear understanding of which models are most accurate and practical. Such research would also be facilitated by the development of new open source tools and data sources, with particular focuses on both model calibration and evaluation. Availability of these tools would facilitate the implementation of spatial interaction models in practice, across a variety of disciplines and applications, which would reduce the barrier for entry for researchers and allow for consistent evaluation and replication of performance across a variety of conditions. This would therefore allow for the use of existing models in new areas of research or the utilisation of models on new and existing datasets for the purpose of replication and robustness checks. Thus, In order to facilitate both the adoption of new data and models, the literature has also to embrace research into methods of consistent evaluation and the development of open source tools that enable that research.

8.2.1) New Models

With interest in spatial interaction modelling coming from diverse domains of research, such as commuting, trade, patenting, migration, healthcare, travel and retail (Wang, et al., 2016; Wu, et al., 2021; De Mello-Sampayo, 2016; Cullinan & Duggan, 2016; Lata, et al., 2018; Blum & Goldfarb, 2006), there has been the development of a variety of different model formulations and implementations that have been adapted to different scenarios. This includes the adaptation and usage of data science methods such as neural networks and decision tree methods, the modification of existing model formulations such as the geo-lagged implementation, the development of new modelling formats such as the radiation model and the adoption of new tools such as agent-based modelling. The majority of these new implementations however have come from domains other than retail analysis, meaning that often a key difference compared to the model formulation in this thesis is that they have been developed on “complete” data sources that represent whole systems. This is often in contrast to the choice-based sample of loyalty card data that needs to be scaled up to model total revenue flows of stores in retail analysis (O'Kelly, 2011). It has also created a tangle of different model implementations that are often hard to replicate or evaluate relative to existing model formulations. Thus, it becomes necessary to discuss how these models may lead to improvements, alongside their benefits, drawbacks and how they may fit in with the development of spatial interaction modelling in relation to grocery retailing.

8.2.1.1) Data Science Methods

One line of modelling development has been the integration of data science methods, such as tree based algorithms or neural networks, into applications modelling spatial flows. This is primarily because such methods have the ability to behave like traditional statistical methods as they can predict or model an outcome by minimising an error term, while making no assumptions as to the underlying form of relationship or distribution of the data (Black, 1995). This is in contrast to traditional spatial interaction model implementations that follow a clearly defined expected relationship, with the calibration method chosen often making assumptions about the underlying data distribution. Thus, this means that the data science methods can be more flexible in modelling the underlying data and be able to capture different relationships that a Wilsonian form of the model couldn't (Karlaftis & Vlahogianni, 2011). These methods were thus sought as a potential solution to a general lack of accuracy of existing spatial interaction models (Openshaw, 1998; Fischer, et al., 2003), leading to the development of the "geocomputation" approach to spatial interaction modelling (Fischer & Reggiani, 2004).

Neural networks were one of the first data science techniques to be applied to the spatial interaction modelling domains with initial applications focused on commodity flows, trade and migration (Fischer & Gopal, 1994; Black, 1995; Openshaw, 1998). These studies suggested that neural networks could improve on the accuracy seen from the Wilsonian model (Fischer & Gopal, 1994; Black, 1995), but where often only compared to the unconstrained OLS calibrated model thereby leaving open questions as to their true benefit (Mozolin, et al., 2000). In contrast, tree based algorithms have been adapted to spatial flow modelling only recently with examples of their application provided for by Pourebrahim et al. (2019) and Spadon et al. (2019). In the former, a random forest regression model was compared to the Wilsonian form on daily commuting in New York, finding that the random forest model performed better as measured through the R^2 and RMSE values. It was concluded that this was because they could better reflect the relationships of both small and large scale flows better than the restrictive Wilsonian model formulation. In contrast, Spadon et al. (2019) split the flow modelling problem into a classification and regression problem, similar to that of a Zero-inflated Poisson or Binomial Regression model (Krisztin & Fischer, 2015). In this, they concluded that an XGBoost Random Forest Model outperformed an unconstrained OLS Wilsonian model and a radiation model. Therefore suggesting that such models could accurately reflect geographical flows.

However, there are several limitations of these implementations that have yet to be resolved such that they could be used consistently. The first issue is the development of constrained versions of these models, whereby post-hoc implementation of constraints for neural networks showed relatively poor fits compared to the Wilsonian models (Openshaw, 1998; Fischer & Reismann, 2002; Fischer & Reismann, 2002). It was only when the constraints were implemented in the model itself that results improved (Fischer, et al., 2003), but consistent application or implementation in this regards is not clear for either neural networks or random forest algorithms. Furthermore, there is critique of the "black box" nature of these algorithms, with some arguing that their usage comes from their greater predictive accuracy as opposed to their ability to explain the phenomenon they are modelling (Karlaftis & Vlahogianni, 2011). Although there are techniques to identify important features in the model, such as feature importance or Shapely Additive explanation (SHAP) values which identify the size and direction of influence, they can still be difficult to understand and interpret (Spadon, et al., 2019). This is of particular concern for retail applications where an understanding of why some stores perform better than others is often an important reason for the application of spatial interaction models. Thus, there is a potential trade off in terms of greater accuracy and interpretability of the modelling outcomes.

Furthermore, although there are several open source frameworks such as PyTorch, Keras and SciKit Learn, application of such models have often been limited in relation to spatial interaction implementations. This is often attributed to the computational resources and data requirements, alongside often long training times that are required for often relatively small improvements in accuracy (Fischer & Reismann, 2002). This is especially difficult for large datasets, such as loyalty card data, where even simple calibration methods can be limited by computational capacity. Thus, potentially limiting the size of their application. Finally, it is also worth highlighting that these models were often applied in domains where the training data could be said to be "complete". This means that the behaviour and scale of the data is comparable to the final modelling outcome. This is in contrast to the application of grocery retailing where flow data is often based on loyalty cards which represent only a subset of the total system behaviour and the scale of flows (Waddington, et al., 2018). This can therefore make training, testing and validation of these models complicated compared to regression or iterative calibration. Thus, implementation of neural networks or tree based algorithms in grocery retailing would thus require the development of clear, relevant and unambiguous loss functions to estimate total store revenue. The pitfalls of this are highlighted in the iterative calibration section which uses average trip distance.

Therefore, while data science methods may be more flexible in their applications, they are often limited by a lack of clear implementations, especially in relation to the constrained models, their black box nature, and their requirements of large amounts of complete data and computational resources. Thus, while they may represent one of the biggest opportunities for model performance improvements, these issues need to be resolved before they can be routinely used in a grocery retailing sector.

8.2.1.2) Agent Based Models

Another alternative modelling framework that could potentially improve the performance of spatial interactions models would be Agent-Based-Models (ABMs). These models are used to understand emergent outcomes by modelling individual agents, their relationships and interactions, and their environment (Macal & North, 2014). Thus, while spatial interaction models replicate the overall system from a top-down perspective, ABMs are used to represent and understand individual decisions and preferences and how these decisions then aggregate to create emergent outcomes (Clarke, 2014). In general, these models are enabled by new and large datasets that are becoming available that allow for the identification of individual behaviour, thus having the potential to model the expected interactions in more detail than a spatial interaction model (Sturley, et al., 2018). A key benefit of these could be in illuminating the dynamic nature of retailing, an already acknowledged limitation of the Wilsonian form of the model (Wilson, 2010), and modelling multi-purpose shopping trips or destination chaining behaviour.

Examples of their implementation in relation to retail include working by Birkin and Heppenstall (2011), Dearden and Wilson (2011) and Sturley et al. (2018). The first two papers attempt to utilise Agent Based Models to examine the dynamic nature of retailing through the adaptation of the Harris-Wilson (1978) dynamic spatial interaction model form. In doing so they create an ABM which accounts for the dynamic relationship between consumers and retail outlets in terms of price, location and floorspace, allowing for the location and size of retail stores to change over time. Both of these use macro-level data to calibrate the model and replicate the dynamic nature and in doing so highlight that micro-level data would be preferable to calibrate both retailer and consumer behaviour. Nevertheless, they highlight that such data is often difficult to obtain (Dearden & Wilson, 2011). In contrast, Sturley et al. (2018) utilise loyalty card data to distinguish different consumer behaviours while replicating the outcomes of a Wilsonian spatial interaction model. The results were deemed to capture the observed choice behaviour of different consumers groups thus highlighting the potential of these forms of models to understand individual behaviour. They nevertheless suggested that improvements could be made to incorporate a more realistic underlying geography, capturing the effects of competition and brand attractiveness, and also to model non-residential demand. Such improvements were suggested to be able to be obtained if there was access to more detailing individual level behaviour and greater computation resources to apply the models at scale.

These papers therefore highlight the feasibility of Agent-Based-Models in relation to spatial interaction modelling but that developing a detailed and calibrated model can be difficult in practice. This is because of limitations in terms of both computational resources and data. Firstly, considerably computation resources are often required to run and calibrate an ABM, with Dearden and Wilson (2011) only able to model 10% of the total interactions in Yorkshire. This is a known issue with ABMs because of the amount of decisions that agents make and their interaction, especially when applied at scale (Heppenstall, et al., 2020). Furthermore, while model calibration and validation are crucial steps in model building there are often considerable challenges that need to be overcome for ABMs because of the complexity of model implementation and the amount of parameters (Sturley, et al., 2018). Notably it is often difficult to find or extract data that represents individual interactions at a scale that would allow for reliable calibration and validation (Heppenstall, et al., 2020). Thus, while such models allow for a better understanding of individual behaviour and how it effects system outcomes, they can be hard to implement due to computational and data limitations. To allow them to be useful, retailers need to collect data beyond loyalty cards to understand the complex behaviour that consumers exhibit. This may allow them to model not just behaviour captured in the Wilsonian model but also the potential to replicate multi-purpose, non-residential and online shopping that is growing in the grocery retail market, which could improve modelling performance on convenience and high-street stores, filling in an acknowledged gap in the literature.

8.2.1.3) New Modelling Formulas

Other modelling developments have either attempted to derive an alternative formula, to move away from the Wilsonian model, or to adapt and improve existing ones. In terms of the former, the radiation model was developed by Simini et al. (2012) in response to some of the perceived issues of the Wilsonian model including: the use of a variety of different distance deterrence functions without clear theoretical guidance, required data to train the model, systematic errors in predictions across a variety of domains, increases in flows without limits and the static nature of the model (Simini, et al., 2012). The radiation model was thus developed through the analysis of daily commuting patterns in New York, creating a parameter free model. The aim of this was to allow the model to be applied to a wide variety of scenarios without having to be recalibrated, thereby resolving the requirement of previous data to train the model (Simini, et al., 2012). Initial results in replicating commuting suggest that the radiation model performed better across a wide range of time scales whilst also capturing diverse underlying processes that the Wilsonian formulation could not. Notably, it was seen to better represent large commuting flows, a point which has been acknowledged as a limitation of the Wilsonian form of the model in the previous literature (Fischer &

Gopal, 1994). Thus, it was argued that the new model could potentially replace the application of the Wilsonian formulation, at least in regards to commuting.

Papers that followed up the analysis, applying the radiation model to a variety of different datasets, noted however that in most scenarios that the traditional gravity model still performed better (Masucci, et al., 2013; Lenormand, et al., 2016; Hilton, et al., 2020). This was suggested to be due to the parameter free nature of the model, meaning that it was unable to be adapted to different scenarios, scales and datasets unlike the Wilsonian model (Stefanouli & Polyzos, 2017). Subsequent papers have then since sought to develop the radiation model further, including the addition of at least one calibratable parameter which while removing one of the key benefits of the model allowed it to be more generalisable (Stefanouli & Polyzos, 2017). Nevertheless, even with these adaptations it has been that suggested that one type of model, out of the radiation, intervening opportunities, or a gravity model, do not consistently outperform the others at different scales or application (McCulloch, et al., 2021). Therefore, it has been suggested that rather than a single model, an ensemble method could be utilised to combat the relative weaknesses of each or that future research should further explore under what scenarios and applications each model formulation performs better to help with future applications (McCulloch, et al., 2021).

Therefore the radiation model, at least in its original formulation, may be of limited use for retail location analysis. This is primarily because the model was created for a defined scale and scenario in terms of commuting, which is likely to exhibit different behaviours relative to retailing shopping in terms of frequency, motivation and scale. Nevertheless, it could be explored with more recent adaptations in terms of whether they are able to be calibrated to a grocery retailing scenario and what may thus be lacking from the current formulation. Furthermore, it also highlights the potential of new model formulations, if not the radiation model, to be developed that resolve some of the issues of the current model. Thus, future model adaptations or formulations must also be adapted and compared. Nevertheless, this highlights the difficulty in developing new model formulations that are adaptability to different scenarios, especially when they are developed in relation to a specific scale and purpose. Developing a model in this was can make it difficult for the model to be adapted or implemented in new scenarios, limiting their generalisability.

8.2.1.4) Adaptation of Existing Models

Finally, there is also the development and adaptation of existing model formulations to account for limitations of the existing models. An example of this was presented in Chapter 7 with the application of the competing destination model. While this however failed to resolve some of the issues presented in the rest of the thesis it is worth discussing other alternative adaptations. One of which is the geolagged family of models which were developed in response to the Wilsonian model assumption of independent of errors after controlling for distance. This is because it is suggested that in reality there is expected to be a some spatial dependence of flows (Lee & Pace, 2005; Krisztin & Fischer, 2015). The traditional model formulations were seen to omit variables such as accessibility, visibility of signage or retail demand externalities that are expected to lead to spatial dependence of variables, thereby affecting modelling performance. In the case of grocery retailing this could be because two neighbourhoods have similar levels of flows to a store due to similar levels of accessibility, store visibility, common transportation options and the sharing of information between neighbours (Lee & Pace, 2005). These groups of models then were developed to integrate the spatial dependence of flows into the model formulation and calibration for both the origin and destination (Griffith, et al., 2017).

Results from the initial implementation of the models suggested that accounting for spatial dependence led to reduced bias in parameter estimates and more accurate models, thereby suggesting the importance of these factors in model calibration and estimation (Lee & Pace, 2005; Krisztin & Fischer, 2015). This therefore meant that including the potential influence of spatial dependence would allow for more accurate estimation of flows from origins and destinations, including the estimation of both small and large flows more consistently (Lee & Pace, 2005; Kerkman, et al., 2017). The importance of these relationships have also been identified in other domains of geographical analysis with recent papers integrating spatial overflows in random forest regression methods to account for the influence of spatial dependence (Georganos, et al., 2019; Barzin, et al., 2022). This therefore suggests the importance of accounting for these factors in geographical modelling and hence spatial interaction modelling.

However it is also acknowledged that implementing such models is difficult and requires a large amount of data to accurately estimate the lagged relationships, alongside the potential for multiple different model implementations (Lee & Pace, 2005; Kerkman, et al., 2017; Yeghikyan, et al., 2020). This has often meant that practical implementation of these models has been limited or constrained, with the model implemented in Lee and Pace (2005) only being able to estimate an unconstrained model in relation to retailing. This therefore raises a similar issue as the data science implementations as to how a constrained version of the model could be implemented and calibrated, or whether such model formulas lead to relative improvements anyway. This issue, as already mentioned, is especially difficult for retail location analysis because loyalty card data isn’t representative of the total population behaviour and thus limits the final application. Therefore further examination of these models for grocery retailing would need a method for scaling up the estimates for total revenue whilst also being able to account for the relative influence of competition. This could be done with access to new datasets, such as credit card or loyalty card data, that could be used to identify total revenue flows and the influence of store brands. Nevertheless, across boundary flows in grocery retailing are likely to influence each other and thus future research would need to identify how such models could be implemented in retailing.

8.2.1.5) Other Models

Mentioning such developments over the last 20 years is also not to ignore previous modelling adaptations and implementations. This includes the creation of the competing destinations model (Fotheringham, 1983), the dynamic model (Harris & Wilson, 1978), the intervening opportunities model (Harris, 1964) and many more which have each aimed at overcoming some of the limitations identified on previous model forms. However, this also highlights the wide variety of literature that discusses and models spatial flows, coming from different domains, with different data sources, and with different areas of expertise. This therefore often makes it difficult to select and evaluate each model implementation under different scenarios, especially when the implementation has not been made clear or code is not available to replicate existing results. What is needed then, alongside the exploration of the models above, is a comprehensive overview of all the current model implementations, their advantages, disadvantages and under what scenarios and applications they can be used. Without this, arguably the literature is likely to continue to be impenetrable to new researchers, especially with spatial interaction not being taught consistent at an undergraduate level (Rowe, et al., 2022), thereby acting as a barrier to new developments in the field.

8.2.2) New Data

Alongside the development of new models for spatial interaction modelling, new datasets are also becoming available which could improve the performance, especially in relation to grocery retailing. The use of these new datasets can be split across the part of the model they support, either flows from origins to destinations, the attractiveness of the destination, or the emissiveness of the origin. Such new datasets include mobile phone, mobile application, social media or credit card data that could be used to show flows from origins to destinations, measures of footfall, transit stops, and open source images that could be used to determine store attractiveness, and micro-simulated population estimates, credit card information and surveys that can be used to estimate the outflow and emissiveness from origins. The exploration and integration of these datasets into both existing and new model formulations would aim to improve the accuracy of the models, especially in relation to retailing (Aversa, et al., 2018), in light of the results presented in this thesis so far. However, it must also be acknowledged that the exploration and integration of such datasets into the spatial interaction modelling framework is not likely to be easy given concerns over confidentiality, privacy and security. However, an exploration of the benefits and drawbacks of each type of data could highlight the most fruitful areas for future research to explore and build on.

8.2.2.1) Flow Data

In order to calibrate any form of spatial interaction model, data on actual flows from origins to destinations is needed. This is because the purpose of a spatial interaction model is to estimate unknown patterns from known information, but in order to generate accurate and reliable estimates of these unknown patterns an interaction dataset is needed to compare the results with (Haynes & Fotheringham, 1985). Initial spatial interaction models in retailing were developed using survey data from which information about the flow of people and revenue from homes to stores could be derived (Huff, 1964). The issue with this however was that this data was often fragmented, difficult to collect and hard to interpret, thereby limiting early implementations of spatial interaction models and their accuracy (Clarke & Birkin, 2018). The development of loyalty cards was thus a key catalyst for the general use of spatial interaction models in retail location departments (Clarke, 1998). This is because they provided consistent identification of flows from geographically linked origins and destinations, including the number of households, baskets, size of baskets, items bought and the time of the shop (Clarke, 1998). The issue with this however is that the usage of loyalty card data from a single retailer is a choice based sample that can introduce bias into the model estimates (O'Kelly, 2011). This is not to mention that it is often difficult to acquire, access or even develop such datasets for both retailers and academics (Newing, et al., 2015). Therefore such data could be supplemented, or even replaced, with data from other sources that could potentially resolve some of the issues seen so far in this thesis.

One dataset that could potentially resolve some of the issues of loyalty card data is that of credit card data or loyalty cards that extend beyond a single retailer (such as Nectar cards). These datasets could be used to augment and expand loyalty card data by showing sales from origins to destinations across a variety of retail outlets along with timestamps. Access to this data would thus provide access to a wider, potentially less biased, set of flows from origins to destinations, alongside clear examples of revealed preference for different retailers (Suhara, et al., 2019). This could also help in identifying how much different types of households spend in regions across the country, purchase categories and time periods. Furthermore, this may also support improved analysis and accuracy of origin emissiviness and the potential identification of multi-purpose shopping trips. However, access to such data is likely to be as limited and sensitive, if not more so, than access to loyalty card data from a single retailer. This is highlighted by the relative lack of usage of such datasets within the existing retailing literature (Suhara, et al., 2019). Furthermore, the data will still link back to residential locations, thereby only extending the data from loyalty card data rather than fundamentally altering the information that can be extracted from them. However, it would represent actual purchase to quantify the amount of revenue flow between origins and destinations.

This data could therefore potentially be augmented and supported by data generated by mobile phones, such as GPS data, mobile phone tracking or revealed location from social media. The main benefit if this is their potential to show how and when consumers shop, including where their trips originate from. This would therefore help to extend the spatial interaction model application to include non-residential population influences, building on the existing work of Hood et al. (2016) and Waddington et al. (2018, 2019). Furthermore, it could also provide micro-level behavioural data for individual based models such as ABMs, which could inform modelling of multi-purpose trips and workplace shopping (Heppenstall, et al., 2020). These datasets would also be expected to be more representative of the population with the widespread adoption of smart phones, especially in the developed world (Li, et al., 2021). This is why such datasets have already been used in retailing analysis such as examining how different consumer groups reacted to the COVID-19 pandemic (Smith, et al., 2022). However, usage of such datasets can be limited due to large amounts of data and its complex spatio-temporal nature requiring considerable cleaning before it can be operationalised (Hu, et al., 2021; Bonnetain, et al., 2021; Willberg, et al., 2021). Furthermore, there are also ethical issues of using this data due to the high level of granularity alongside potential for monitoring and identifying individuals (Li, et al., 2021; Kishore, et al., 2020). Similar concerns can also extend to the usage of social media data, including issues of infrequent and biased posting, although they are more openly available (Hu, et al., 2021; Li, et al., 2021). Finally, while they may represent flows of individuals, this will also not always translate into expenditure values and amounts which adds additional complexity into total revenue estimation. To this end, initial exploration of integrating such data into spatial interaction model has suggested that they cannot replace traditional survey or census data, but could be used in conjunction with them to improve their accuracy (Pourebrahim, et al., 2019).

Therefore, these datasets could either replace or supplement traditional sources of flow data. Their main benefit would be the wider population that could be examined, allowing for a more representative sample of individuals. Furthermore, they could be used to inform the relative attractiveness of different retailers by representing broader shopping habits, including the identifications of multi-purposing shopping trips and non-residential origins. Thus, utilisation of these datasets could influence the form of the spatial interaction model that would be used. This would be especially so if these data suggest that the modern consumer is no longer behaviour consistently in line with the assumptions of the Wilsonian form of the model. However, accessing such data is often more difficult, alongside more complicated to work with, relative to loyalty card data, and they are not without their own drawbacks in terms of representation (Li, et al., 2021). This can become especially difficult if researchers were to attempt to combine and integrate datasets from multiple sources due to security and commercial concerns (Newing, et al., 2015).

8.2.2.2) Attractiveness Data

The use of new datasets for spatial interaction modelling for grocery retailing can also extend to how we measure the attractiveness of an individual store. Typically, store size as measured by square footage is used to represent an individual stores attractiveness. This is based on the assumption that a larger store will offer more products and thus a consumer would be more likely to be able to purchase all the goods they want and hence have more confidence that they will not need to perform a further shop at another store (De Beule, et al., 2014). Thus, the larger store being more attractive. This use of size as a proxy for attractiveness even dates back to the original development of the gravity model for retailing by Reilly in 1929 when the size of the town was used as a measure of attractiveness. However, this has been criticised for its lack of dynamic perspective, assuming single purpose shopping trips (Trasberg, et al., 2018), the fact that total floorspace does not always represent the product category the model focuses on (Newing, et al., 2015), and often that floorspace measurements are not readily available. Therefore, there are several different datasets that could be used to complement or extend the store size attribute, including footfall data, points of interest data, transit data and store image data.

The first of these, footfall data, can be used to show the actual amount of people that visit or pass by a store in a given time frame (Wood & Browne, 2007). The idea is that this would therefore be seen as a proxy for the level of activity that a store generates and in the local area, such as the influence of workplace, residential or transit populations (Trasberg, et al., 2018). Such data could be derived from a variety of sources such as social media data, phone GPS data, or even traditional survey data sources, which could then be used to indicate a potential stores attractiveness beyond its size. An example of this is Trasberg et al. (2018) who derived footfall data from a street network analysis in Liverpool and correlate this with traditional and non-traditional measures of store attractiveness. Thus, laying the foundation for future usage in retail location analysis. Such a measure could be particularly beneficial for estimating the attractiveness of convenience retail outlets that target non-residential populations (Hood, et al., 2015). In particular, this metric could be reflective of true "convenience" shopping where consumers shop while completing other activities such as their journey home, visiting attractiveness or if the store is near to their workplace. The issue with this data however is where the data comes from, whether it is derived or actual, and how it correlates with actual revenue (Trasberg, et al., 2018). This is because passing footfall may be converted into sales at different rates depending on brand, advertising, sector and relative location. Thus, such data will not necessary mean that a store is more attractive, even if it locates in an area with high levels of footfall. To test this there are commercially and openly available data such as the Sprinboard data collection in the UK (Sprinboard, 2022), or SafeGraph in the US (Safegraph, 2022) which could be used to examine whether this data does actually improve model fit.

Further attractiveness datasets could include point of interest data, transit data, visibility, use generated content or actual expenditure (Wood & Browne, 2007). While transit data and actual expenditure values are data sources that are already available, their use and implementation in relation to retail spatial interaction models have often been limited. This is because, with transit data the relationship with store attractiveness is not necessarily clear due to different purposes such as car parking and public transport nearby stores (Themido, et al., 1998; Hood, et al., 2015), while the use of total revenue would often mean that there would be a circularity within the model implementation with total revenue being the aim of the model. Point of Interest (POI) data is also likely to have a complex relationship with store attractiveness, being related to the influence of multipurpose shopping. This would require information not only about the location of the POI, but also the type of POI and how many visit the attraction in order to relate this to the potential foot traffic in the area. This would then have similar complexities in terms of translating POI foot fall into actual sales (Trasberg, et al., 2018). Finally then, user generated content, or google street view images, could be used to identify store attractiveness through images, tagging or reviews (Wood & Browne, 2007). While ratings of store attractiveness based on the age of façade, street direction, and accessibility were suggested to lead to improved model performance by our organisation, the collection and analysis of this data would be a considerable undertaking on a UK wide scale. Thus, there may be potential to utilise advancements in image classification technologies that could automate this process in the future and feed into store attractiveness measures (Balali, et al., 2015; Kang, et al., 2018; Hu, et al., 2020). However, this may be complicated and difficult to routinely collect.

The benefit of these datasets then is the potential to improve the measure of store attractiveness beyond simply a measure of store size. While store age was already identified as potentially influencing the modelling results in Chapter 5, consistent data in this regard along with other datasets could thus potentially lead to more accurate estimation of how consumers view individual stores. Specifically, this could be of considerable benefit to the estimation of convenience stores attractiveness where store size of often not the main driver of store performance but rathe the convenient nature of shopping, with relations to footfall, POI and transit data. However, for most of the datasets there is likely to require considerable effort and exploration to operationalise and understand their relationship with individual store revenue, especially in relation to data derived from primary or secondary sources such as footfall from mobile phones or attractiveness based on open source images. This would then be further complicated by their integration with existing and new forms of the spatial interaction model.

8.2.2.3) Emissiveness Data

On the other side of the spatial interaction model is the emissiveness or the amount of revenue available from an origin. This includes not only the amount of revenue but also the type of people who come from each geographical area. In the model presented in this thesis, and in most spatial interaction models in the literature on grocery retailing (Newing, et al., 2015; Waddington, et al., 2019), the amount of revenue available is calculated by multiplying the number of households in an origin from the census data by the average household spend that is estimated form survey data. There are three main issues with this however. The first is that household numbers often lags behind the actual number of households in some output areas due to new construction or social change in the years following the census. The second issue is that even within geographically boundaries as small as output areas, there can be diverse household groups in terms of their sociodemographic thereby affecting their modelled behaviour. Finally, actual spend per household can vary across the country due to factors such as differences in disposable income and variation in prices, as opposed to current estimates that are split across eight different output area classification supergroups. Consequently, such estimates for expenditure may be different to actual levels of expenditure and thus affecting the model. This estimate could therefore potentially benefit from data from other sources and methods to support the estimation.

One method that has been developed and used recently in the literature is that of microsimulation. This method is used to estimating data through the combination and recombination of different small area data sources to be able to identify the true make up of small area populations (Birkin, et al., 2017). The benefit of this is that it allows for the households to be assigned to different social groups at a level lower than most census estimates, thereby reflecting the true underlying variety (Birkin, et al., 2017). In our application this would mean an estimate of household groups within each output area, which would thus allow for a more realistic estimate of the number of households in differe3nt groups and hence potentially more accurate estimates of total revenue. Beyond the level of the census, it could also be extended to individual household estimates which could generate even further disaggregated model implementations as well (Nakaya, et al., 2007), although this may be limited by GDPR cut-offs. The accuracy of this method depends on the extent to which the different datasets can be joined together and hence on their individual accuracy. This is further complicated by a reliance on census data which can still become quickly outdated, even by the time it is published and made available. Therefore, alongside this, open source data such as open street maps could potentially be used to estimate the number of households within each origin. This could be done by identifying address changes or the number of individual units within a geographical estimate. Although the accuracy of this will depend on how often this type of updated and hence will lead to potential regional differences (Langford, 2013).

On the other hand, there are also different ways and datasets that could be used to estimate the total amount spend per household on grocery products. This could potentially include the use of credit or debit card data, including cross retailer loyalty cards, which could show how much is actually spent from each type of origin at different retailers over different time periods. This would thus show the total spend across all possible destinations which could be used to understand the variance in expenditure on groceries across the UK. This could also be used to inform relative attractiveness of different retailers as well (Newing, et al., 2015). Such data could be augmented and/or supported by collecting receipts from customers to identify where and when they shop which would not only identify how much they spend but also which retailer they see as more attractive. An example of this would be data from the HuYu app which shoppers can scan their grocery shopping receipts (dunnhumby, 2022). However, access to debit or credit card data is likely to be restricted, as mentioned above, and collection of receipt level data can be expensive. Alternatives therefore include existing survey data as applied with further disaggregation, such as to the output area classification group level, loyalty card data if it is assumed that all expenditure is spent at the one store, or even social media data (Marchetti, et al., 2016). The latter of these suggests the inventiveness of new potential data sources in retailing data science.

8.2.2.4) New Data Summary

The identification of the way in which new data sources can be added into the spatial interaction models for flows, attraction and emissiveness estimates highlight potential avenues that future research can explore. This includes the potential of mobile phone mobility data, social media, credit card data and potential new methods such as microsimulation. It becomes difficult to identify which data sources may be most beneficial but the main issues that could be overcome in relation to traditional datasets are the range of population that the data covers, the ability to analyse multi-purpose shopping, estimating the attraction of retailers, and reliable estimates of expected expenditure. This could therefore lend itself not only to the extension of existing models but also to new implementations such as ABMs or Data Science methods to better understand and represent the full range of shopping behaviour. The issue with many of the datasets presented so far however include privacy, access and commercial sensitivity of data, especially when combined for research purposes. This, along with the often high costs of purchasing or dealing with “big data” can often mean retailers still rely on traditional sources of data (Aversa, et al., 2018). Thus, retailers and academics could work together to identify data that would improve the model the most, a job which the Consumer Data Research Centre is aiming to currently do (CDRC, 2022).

8.2.3) Model Evaluation

With the potential of new models and datasets to improve the current modelling performance it becomes important to discuss how models are evaluated. This is because for the literature to progress it must be clear when and where improvements occur. This means that the literature needs a reliable measure, or group of measures, that tell us how accurate new model implementations are relative to the data they are applied to and previous model implementations (Wilson, 1976). This involves being able to compare the accuracy of different models applied on the same dataset, the accuracy of the same model across different datasets and whether there are differences between the modelling outputs and the underlying data (Knudsen & Fotheringham, 1986). Without clear and consistent measures to identify when improvements occur, it becomes difficult for the literature to move forward or to identify the most relevant conditions for different model formulations. Arguably this is the state of the current literature where there are a considerable number of different model variations or suggested data integrations with no clear future research direction. This is because of a lack of consistency in methods of evaluation across domains within the literature with a variety of different metrics being used.

In this sense, spatial interaction models produce two key outputs of flow length frequencies and an origin destination matrix of flows which can be used to evaluate model performance (Black & Salter, 1975). The former measure is used to evaluate how well the model replicates the overall system behaviour (Batty & Mackie, 1972), while the latter is used to evaluate how well the model represents individual behaviour and flows (Birkin, et al., 2010). When evaluating the performance of spatial interaction models it is necessary to consider both outputs to ensure that there is a good fit with the underlying data. In particular, focusing on a single metric as an evaluation criteria can often lead to onerous conclusions in modelling performance. To an extent this can be seen in the iterative calibration in Chapter 6.5 whereby the calibration was based on a single metric that led to a range of modelling outcomes. It therefore becomes necessary to identify the ways in which the existing literature evaluates modelling performance through these two criteria whilst suggesting ways in which they could be improved in the future.

Firstly, in regards to the trip frequency of the model, the most commonly used evaluation metric is the average trip distance which is used to evaluate how well the model replicates the overall system behaviour (Batty & Mackie, 1972). This is based on the idea that if the model can replicate a characteristic of consumer behaviour, in terms of how far the average consumer of pound travels, then it is also likely to be able to estimate the spatial patterns within the modelled system (Newing, et al., 2015). The issue with this however is that there are no commonly accepted standards or ranges for the metric which would suggest a “well” performing model or one that would be classed as acceptable. Indeed, it is often only used to compare relative improvements in model implementation. While useful, it would also be beneficial to identify commonly accepted standards in performance of this metric, even if this is a soft standard that does not always have to be met. This would help to guide practice and implementation, with the suggestion that there was a good model fit.

Furthermore, the use of a single value to represent a complete and often complex system of flows means that a lot of the underlying variances and different behaviours are masked. This can create a problem of model evaluation as variance in the performance of the model can be hidden with the use of a single metric. Therefore future research should dive deeper into understanding how this metric behaves across different circumstances, such as different datasets and within the same dataset as well. For example, in the case of retailing it may be expected that consumers who travel further to visit a retail outlet may have larger basket sizes and average spend per trip (Merino & Ramirez-Nafarrate, 2016), thereby having different average trip distance for this group as opposed to those who live closer and shop more regularly. This is not to suggest that the metric should devolve far enough to compare the actual flows, but rather that the performance of the metric may vary across different subsets of the population which should be identified. This would then go beyond the disaggregation already seen in the model implementation to see if the model is replicating all ranges of behaviours well.

The second concern is then how to evaluate how well the predicted origin-destination matrix of flows replicates the original data matrix. The main consideration here is that the two main metrics that are often used in this regards, R^2 and RMSE/SRMSE have been previously suggested to not represent the true underlying performance of the model due to the nature of the spatial interaction flow, thereby leading to incorrect conclusions as to relative performance (Knudsen & Fotheringham, 1986). Their continued use then in the spatial interaction modelling literature is often due to the ease of interpretation, especially for researchers from different domains, alongside their consistent historical use in the literature (Newing, et al., 2015). Thus, Knudsen and Fotheringham (1986) identified different families of metrics including traditional statistical values, distance based metrics and information based statistics which could be used to evaluate spatial interaction models. Whilst metrics from the two first families of metrics have traditionally been used, and continued to be used, more recent model formulations have begun to use information based statistics (Piovani, et al., 2018). It was identified that there were a number of potentially relevant metrics from this family, but exclusive use of such metrics can often fall into similar issues to that of traditional or distance based metrics. Primarily, the use of a single statistic can often be inappropriate for the underlying application or can hide performance issues that would be identified with other metrics. It is therefore argued here that a combination of statistics from each family should be used in future studies. This would allow for the alignment and compared with previous studies using metrics such as R^2 and SRMSE, whilst also taking advantage of the benefits of information based statistics. Therefore, a consistent range of evaluation metrics should be identified that can be used across many different disciplines and applications which would allow for simple model comparison.

Using these metrics goes hand in hand with being able to identifying ways of determining how each new feature and dataset contributes to model performance. This becomes important when new datasets are added to existing ones, either creating indexes from multiple measures or utilising many individual features in model calibration. While the relative improvements in metrics with the removal or addition of features could help illuminate whether the new feature adds new information or improves the model performance. It is thus suggested that the literature could adopt techniques from the data science domain. This could include the use of metrics such as feature importance or permutation importance which can be used to establish improvement thresholds for new data based on existing metrics, or attempts to integrate methods such as the Shapely Additive exPlanations (SHAP) values which would indicate both the importance and direction of influence from new features (Spadon, et al., 2019; Pourebrahim, et al., 2019). Such integrations could be addressed in the future literature, especially as the availability of new datasets becomes more prevalent alongside new model formulations that utilise existing data science methods.

While the identification of specific metrics or ways in which they are implemented is beyond the scope of this thesis, it is argued that it is important for the future of the field to be able to clearly identify ways of evaluating model performance going forward. This is especially in light of the potential model and data improvements that have been identified in the previous two sections. Without a clear way of evaluating improvements in modelling results consistently it will become difficult to identify where the literature should focus future research endeavours. This will include making it difficult, as it arguably currently is, the most relevant models or datasets for researchers and practitioners in an already crowded field.

8.2.4) Model Implementations

With the creation of new modelling forms and the use of big data it is also important to enable quick and easy implementation of spatial interaction models for their continued use and evaluation. This is especially important given the often highly complex and mathematical nature of model formulations, such as the work of Wilson (1969, 1971) on which much of the existing literature rests on, where often an understanding of at least undergraduate level mathematical concepts is required to get started. This can therefore acts as a barrier to entry into the field for new researchers, thereby limiting future developments. Thus, in order for the literature to continue to advance and allow for new models and data there must be a way to easily and quickly implement the models to allow researchers to gain an understand how the models behave in practice. This would thus require the development of open source infrastructures and making code and data available with the release of new papers. These developments would then improve accessibility for researchers that are interested in the area and this expand the potential domains and applications to which these models could be applied (Rowe, et al., 2022).

The first stage of this is the development of open source software and infrastructure that enables the running of spatial interaction models whilst also being accompanied by clear implementation guides. The aim of this would be to allow any researcher or practitioner who is interested in the use of these models to be able to develop and implement a spatial interaction model with relatively little time and resource costs. These tools would thus help individuals get an understanding of how the models should behave in practice across different conditions. This would also help to facilitate a movement towards open available datasets and code which would support reproducibility within the field (Rowe, et al., 2022). This is because if all researchers and practitioners were using the same tools then the methodology would be replicable if the data is made available.

This ecosystem is already starting to develop. For R users there is the simodel (Lovelace & Nowosad, 2022) and gravity (Woelwer, et al., 2022) packages, while for Python users there is SpInt (Oshan, 2016) and the SciKit Mobility (Pappalardo, et al., 2019) which are able to run spatial interaction models. The development of these packages, alongside clear instructions of how to use them and dummy datasets, enables researchers to implement at least a basic form of the gravity model. Thereby acting as foundation from which the spatial interaction modelling literature can build on in the future. The issue with these implementations however is that so far they are limited in the models that can be implemented and how they can be used. For example, in the case of SpInt the flows have to be integer values (Oshan, 2016), often only Poisson regression can be used, or only a few model formulations have been implemented such as only an unconstrained model in the simodel package (Lovelace & Nowosad, 2022). These limitations therefore constrain their usage in practice. For example the unconstrained version of the model has been shown to be of limited us, in domains such as trade or patent flows additional parameters are often added to the model (Blum & Goldfarb, 2006; Picci, 2010), and model calibration can often determine modelling performance (Mozolin, et al., 2000). It is therefore hoped that in the future these packages and further open source tools are developed with resources and support in the future, such as the recent Google Summer of Code for SpInt (Hasova, 2021). This should focus on implementing more model formulations, such as competing destinations or intervening opportunities, allowing for alternative calibration methods, and the integration of multiple evaluation methods. Otherwise, the usage of open source spatial interaction software will be limited.

The development of an open source software ecosystem should also help to facilitate a move towards open source publishing data and code. Making both code and data available with the publishing of a paper, especially one that proposes to create a new model or validate an existing one on a new dataset, would allow the community to both verify the papers claims and build on their advancements. In the case of spatial interaction models this is especially important given the variety of different methods of implementation and evaluation. Foe example, in evaluating the performance of tree based algorithms against a Wilsonian form of the gravity model, Spadon et al. (2019) used an unconstrained gravity model calibrated via OLS. In this case the model should have instead be compared to an origin constrained Poisson regression model. Thus, if the data and code were available then the study could be replicated and the claims validated.

Access to data and code would also benefit the literature by allowing a greater understanding of the implement and data pre-processing steps that are used. This is important because such steps are often critical in the application of spatial interaction models as well as helping to identify which forms of the model are most appropriate for each datatype. For example, Pourebrahim et al. (2019) utilise twitter data in their implementation of a spatial interaction model on daily commuting flows in New York. In this case, access to both data and could would allow future researchers to understand to extract and clean twitter data in spatial interaction model applications, potentially allowing for the validation of the performance of this type of data in other modelling forms and domains such as trade, retail or migration.

However, it must also be acknowledged that in many cases there are likely to be restrictions in publishing the code or data due to confidentiality, commercial or privacy concerns. Many of these issues are likely to arise in relation to some of the proposed new datasets in section 8.2.2 above or in commercially sensitive domains such as retailing. For example, loyalty card information are often closely guarded by retailers in accordance with strict data protection regulation. Access if restricted to anonymised subsets and aggregations of the underlying dataset. This can be seen in the case of this thesis and also in papers such as Newing et al. (2015) or Waddington et al. (2019). This is highlighted by the lack of exploration of loyalty card data in regards to spatial interaction models where results and performances are often not published openly. To this end, while a call for more open access publications, data and code would be beneficial for the wider literature, it is to be acknowledged that this may be limited under certain scenarios.

Nevertheless, it is the hope of this author that the spatial interaction modelling literature continues towards and encourages the trend towards open access code, data and software. This would enable the continued development of model implementations and to bring clarity to the existing literature. This is especially important given the often high levels of mathematical proficiency to understand many of the original papers and subsequent developments, acting as a barrier to both implementation and evaluation. Such developments would allow for lower barrier of entry into the field and a greater collective understanding of the application and benefit of such models. This would then reduce the amount of varied model formulations and implementations that are currently in the vast literature.

8.3) Beyond Gravity Modelling

While the above discussion has presented ways in which the grocery retailing spatial interaction modelling literature could potentially progress it is also worth asking the question as whether grocery retailing location should move beyond the gravity model. The results from the previous chapters highlighted the relatively poor fit of existing model formulations and data to grocery retailing and while new models or data may resolve some of the issues highlighted, it could also be argued that changes in behaviour has meant that gravity models are no longer relevant. These changes include both the shift towards convenience shopping that began in the late 2010s (Buckley, et al., 2007; Hallsworth, et al., 2010) and the rise in online grocery retailing (Kirby-Hawkins, et al., 2019). Such behaviours have already changed the focus of grocery retailers towards the development of convenience stores (Hood, et al., 2015) and online channels (Beckers, et al., 2021), as opposed to large store format stores of the past. This is to the extent that despite consistent use of spatial interaction models in the early 2000s and 2010s (Mendes & Themido, 2004; Reynolds & Wood, 2010), it has been suggested that they are no longer accurate or reliable enough to be used in practice. Thus, although the models could potentially be changed and adapted to improve their current performance, are they applicable to current shopping behaviour?

The first question to answer then is to what extent has grocery retail shopping behaviour moved away from the assumptions of the spatial interaction model? This is because the gravity model was originally applied to retail shopping on the assumption of single-purpose, large basket size shopping trips that originated from the home with the use of a car and were regular weekly shops (Birkin, et al., 2017). However, it was identified that by the mid 2000s consumer behaviour was shifting towards increased convenience shopping characterised by lower travel distances, walking, increase in frequency of shops, smaller basket sizes and multi-purpose trips (Buckley, et al., 2007; Hallsworth, et al., 2010; Elms, et al., 2010). Thus, a departure from the spatial interaction modelling assumptions. Retailers responded by developing convenience and high street store formats in record numbers, accounting for most of the increase in grocery stores in the UK from 2003-2012 (Hood, et al., 2015). This is alongside the more recent rise of grocery e-commerce which, while lagging behind the adoption in other sectors (Van Droogenbroeck & Van Hove, 2017), was accelerated during the COVID-19 pandemic due to limited in in store shopping (Song, 2021). This is expected to affect the distance decay and attractiveness relationship as retailers bear the cost of transporting goods (Kirby-Hawkins, et al., 2019) and determine which store the purchase is attributed to (Davies, et al., 2019). Both of these changes then have considerable effects on the overall grocery market, with e-commerce taking up an estimated 7.3% of the market in 2016 (Kirby-Hawkins, et al., 2019) and convenience stores 22% in 2015 (Hood, et al., 2015), with both market shares expected to grow in the future[16].

These new behaviours and channels of engage with grocery shopping therefore departs from the behaviours expected in the current gravity model formulation. For convenience shopping the trips aren't as regular, baskets are smaller, travel is undertaken through active travel modes and many of the trips are combined with other activities such as work, school or leisure (Waddington, et al., 2019). It could be suggested however that this behaviour could be integrated into existing gravity models by increasing the distance decay parameter thereby accounting for slower modes of transportation and smaller distances, increasing the time scale over which the models are applied to account for the reduced regularity of shopping, and the introduction of new demand layers to represent non-residential populations. Waddington et al. (2019) attempted this by introducing workplace, high school and university student populations as new demand layers, leading to model improvements especially in relation to small format stores. However, there remain considerable discrepancies between production and actual revenue for many stores that remained to be resolved. Furthermore, these changes are also unlikely to capture trip-chaining behaviour or integration with public transport options such as bus or rail. Therefore this is potential to account for convenience shopping but this may be limited in its applicability.

In terms of online shopping it could be suggested that the distance decay relationship expected in the spatial interaction model would be fundamentally changed. This is because consumers don't

[16] This could therefore lead to the suggestion that the gravity model, in its current form, is only modelling a subset of the population, those that perform their regular weekly shop by car at large format stores. An attempt to identify this subset of the population, and thus model their behaviour, can be seen in Chapter 7, Section 4 which explores the implementation of a spatial interaction on a subset of consumers that shop using regular large baskets.

have to travel themselves, instead shifting this costly burden onto retailers (Hood, et al., 2020). While there are theories of the geographical relationship in e-commerce shopping, including the difficult and innovation theory (Hood, et al., 2020), and attempts have been made to alter the functional form of the distance decay relationship to account for this (Beckers, et al., 2021), it could still be argued that the distance decay relationship has been fundamentally altered. Thus, affecting the implementation of the spatial interaction model in its current form. This is also complicated by the store selection process whereby the retailer, not the individual, determines which stores sales are completed by. This means that the attractiveness to stores and store choice has been altered by e-commerce (Davies, et al., 2019). In this sense, while the selection of products is likely to still influence a consumers choice to shop, this is more likely to be by brand rather than a physical store (Singleton, et al., 2016). Such changes therefore separate the size and attraction relationship in the models presented above.

It thus becomes important to identify what alternatives there may be to determine store location and potential revenue. In this sense, it must be recognised that spatial interaction models are but one tool in the arsenal of retail location planners to decide where to locate stores (Clarkson, et al., 1996). Indeed, in many cases spatial interaction models are not used at all in retail location decisions (Reynolds & Wood, 2010). Thus, techniques that are used instead of, or alongside, gravity models include: site visits by managers and staff (Clarke & Hayes, 2013), checklisting site location requirements (Robinson & Balulescu, 2018), spotting customers based on local foot traffic (Applebaum, 1966), creating travel time buffers and overlaying that onto demographic information (Benoit & Clarke, 1997), and regression methodologies that relate local area and store characteristics to store performance (Clarke & Hayes, 2013). Even major grocery retailers in the UK still use gut feeling and location visits to determine store location (Hood, et al., 2015). Thus, these techniques could potentially be used as alternatives to spatial interaction models to determine how desirable a store location may be. The issue, however, is that these models were originally not able to accurately and consistently estimate total store revenue and hence this is why spatial interaction models were adopted.

Nevertheless, some of the models and new datasets presented in the sections above, such as data science models or big social datasets could be used to augment the potential of these techniques. The aim of which is to ensure that they could be used reliably and on a scale which would supplant the usage of spatial interaction models. For example, big datasets on geographical features, such as those extracted from OSM or Google Maps (Pearson, 2007), could be used alongside disaggregated population estimates from micro-simulated data (Nakaya, et al., 2007), and travel time buffers in checklist analysis. This could then be used to determine the optimal location of convenience stores and how much local revenue would be able to be attracted to the store by adding new data layers that would generate a more accurate picture of local demand (Widaningurm, 2015). Such analysis could also be used in combination with machine learning methods such as support vector classification to identify whether a location is optimal for a new store based on the success characteristics of existing stores (Widaningrum, 2017).

Furthermore, regression techniques could be enhanced in a similar way with the introduction of new large datasets and taking advantage of methodologies such as geographically weighted regression (Ozuduru & Varol, 2011), random forest regression with spatial lags (Barzin, et al., 2022), or the usage of neural networks. While the original usage of regression was criticised for its inability to accurately account for the influence of geography in modelling revenue (Birkin, et al., 2017), these new methods and data could resolve this by adding in geographically weighted features. Additionally, the integration of new datasets could highlight which features are most important in determining store revenue than traditional census and competition datasets. The aim of this therefore, provided that the retailer had enough stores for comparison, would be to accurately and reliably estimate the probability of store success or even the amount of revenue it could generate. Thus the "fourth age" of retail location planning could involve adaptation away from the spatial interaction model. Although it has also to be acknowledged that the implementation of many of these models may be limited by access to data or computing power (Reynolds & Wood, 2010).

While both traditional and newer methods could be adopted to replace spatial interaction modelling to estimate store revenue, it is also worth asking the question as to whether grocery shopping behaviour has become too complex to be accurately modelled and predicted reliably. Batty (2018) discussed that we could model well highly routinous decisions with some regularity but he also identified that at some scale, especially when it comes to complex decision making processes, the predictability of this behaviour can begin to break down. This is because while some individual decisions can be modelled, when they are aggregated they can create emergent behaviour that is inherently difficult to predict. To an extent this can be seen in grocery shopping behaviour due to the complex decisions as to the type of shop to undertake, which store to choose and what channel to engage with. While individual decisions could be modelled and predicted, such as whether a consumer is likely to use e-commerce channels or not, what products they will buy, or whether they will use a convenience or large store for a particular shop, when aggregated up to predict overall store revenue the modelling focus begins to shift into the domain of aggregations of complex behaviours that become difficult to model as a whole. This is especially so at the scale this thesis has attempted to develop spatial interaction models at due to the variety of decisions that individuals can take. Thus, it could be argued that the spatial interaction model only applied to a small subset of behaviour, that which is was originally designed to do, as opposed to being able to be adapted to the whole set of modern shopping behaviours. This is such that the model is thus unlikely to be able to predict total revenue to within a degree of accuracy that would be required in practice, which can be further complicated when we then try to add in multiple models, with their own errors, to create an overall model of behaviour.

However, this author believes that while grocery shopping behaviour is becoming more complex and intwined with other decision making processes, there is still likely to be a degree of separation from these processes and an element of predictability. This means that, while the application of the spatial interaction model in this thesis has not been able to accurately model grocery store revenue, it is believed that some of the potential developments mentioned so far should be able to with some degree of accuracy. The extent to which this continues into the future however is yet to be seen, at which point it will become necessary to identify when decision making processes have become too complex to be reliably modelled (Batty, 2018).

8.3) Changes in behaviour

While the above discussion has presented ways in which the grocery retail spatial interaction modelling literature could potentially progress, it is also worth discussing how changes in consumer behaviour, alongside social, economic and technological changes, may depart from the underlying assumptions of the original gravity model formulation. The results from the previous chapters have highlighted the relatively poor fit of existing model formulations and data to grocery retailing scenarios and while some of the new models and methods presented above may resolve some of the highlighted issues, it could also be argued that changes in consumer behaviour have meant that gravity models are no longer relevant. These changes include the increasing diversity of retailers that consumers could be expected to shop at (Birkin, et al., 2017), the shift towards convenience shopping that began in the late 2000s (Buckley, et al., 2007; Hallsworth, et al., 2010), the rise of online grocery retailing (Kirby-Hawkins, et al., 2019), more recent developments in on-demand grocery delivery services (Butler, 2021), and the current affordability challenges in light of recent inflationary pressures (Butler, 2023). These changes could be argued to lead to departures from the underlying assumptions of the gravity model in its current form and thus mean that these modelling methods are no longer valid in estimating grocery store revenue.

8.3.1) Increasing brand diversity

The first of these changes is the increasing diversity and change in make-up of grocery retailing brands over the last twenty years, primarily driven by the increasing market share of deep-discounters from the continent and the potential influence of new entrants. This has been predominantly influenced by the entrance of the deep discounters such as Lidl, Aldi and Netto that began to eat into the market share of the “big four” (Tesco, Sainsburys, Asda and Morrisons) who in 2010 made up over three quarters of the grocery retailing market in the UK (Kollewe, 2022). They were able to do this because of their highly competitive prices relative to these big brands as the cost of unpacking and selection was borne by the consumer, rather than the store, with a focus on cheaper non-branded or own brand products being offered (Kor, 2019). This allowed these discounters to enter the market and capture the lower end of the value chain, locating themselves in areas of major urban deprivation which larger retailers had previously written off as being unprofitable (Birkin, et al., 2017). The success of this strategy has been borne out by their increasing market share in recent years and their continued investment plan with Aldi and Lidl planning to open over 100 stores each over the next four years (Nazir, 2021). This is therefore likely to have further impacts on existing retailers revenues and their market shares.

This changing market structure is likely to influence consumer behaviour and depart from gravity model as currently implemented. This is because, whilst in the early 2000s and 2010s with the market primarily dominated by four main brands, consumers could be influenced simply by which brand had a store located near them, now, with the increasing diversity of brand options across the value chain, the dynamics of competition and choice have become increasingly complex. Therefore, keeping track of the relative attractiveness of brands to different consumers becomes more difficult and thus subsequently more challenging to model. This is highlighted by Aldi overtaking Morrisons to become the UK’s fourth largest supermarket for the first time in 2022 (Kollewe, 2022), alongside the increasing inflationary pressure over the last year leading to more consumers shopping at discounters (Butler, 2023) and shopping around for better deals (Romei, 2023). This therefore suggests that the current dynamics of competition between brands has changed and is still changing which is likely to influence consumers choices and behaviour beyond what we can currently model. This will continue in the future as influenced by macro-economic conditions and consumer attitude, alongside the influence of new entrants such as Amazon Fresh grocery stores which offers a cashier-less experience (Lee, 2023).

8.3.2) The shift towards convenience shopping

Another trend that arguably departs from the assumption of the gravity model is the increasing influence and market share of convenience shopping. It was previously identified that by the mid-2000s consumer behaviour was shifting towards increased convenience shopping characterised by lower travel distances, active models of travel, increased frequencies of shops, smaller basket sizes, multi-purpose and multi-origin trips (Buckley, et al., 2007; Hallsworth, et al., 2010; Elms, et al., 2010). This was due to a change in lifestyle that necessitated a different type of shopping that conformed less to regularity and more towards convenience, driven by time-poor households and individuals (Buckley, et al., 2007; Hallsworth, et al., 2010). This shift in behaviour therefore was a departure from the single grocery shop, undertaken at the same time each week originating from the home, at a large grocery store that was the dominant mode of shopping in the early 2000s (East, et al., 1994; Popkowski Leszczyc, et al., 2004). Retailers responded to this shift by developing new convenience format stores; from 2003 to 2012 most of the increase in grocery stores nationwide was due to the opening of convenience store formats by retailers such as Co-op, Tesco and Sainsbury's (Hood, et al., 2015). This is to the extent that by 2015, the grocery convenience market was reported to be worth an estimated 22% of the total grocery market (Hood, et al., 2015), and this did not include convenience shopping behaviour that was undertaken at large format stores.

These changes are therefore seen as a departure from the assumptions of the gravity model. This is because the gravity model in its current form and data requirements assumes that consumers undertake single-purpose, regular shops with large basket sizes and trips that originate from the home (Birkin, et al., 2017). However, convenience shopping is characterised by shops that are undertaken frequently but on a less regular basis, with small basket sizes and using active travel modes, with many shops originating from places such as work, school, attractions and transport stations and terminating at small format convenience stores located near to consumers (Buckley, et al., 2007; Hallsworth, et al., 2010; Elms, et al., 2010). This therefore requires the integration of additional demand layers that assign revenue from locations other than consumers' homes to account for the different sources of demand, alongside alternative methods of calibration and data sources in order to potentially estimate the relationship assumed by the gravity model (Newing, et al., 2015; Waddington, et al., 2019). Thus, complicating an already difficult model to calibrate, further adding in the complexity of choice of destination format and origin of trip alongside the choices of brand and final destination. This is highlighted in the research presented in this thesis as the focus is primarily on large format stores due to the recognised difficulty of modelling convenience store shopping with spatial interaction models such as estimating non-household based demand sources and willingness to travel from these origins.

8.3.3) The increase in online grocery retailing

The shift towards convenience shopping habits goes hand in hand with the more recent rise of grocery e-commerce, which, while lagging behind the adoption in other retailing sectors (Van Droogenbroeck & Van Hove, 2017), was accelerated during the COVID-19 pandemic due to limited in store shopping (Song, 2021). This has been facilitated by the benefits of online shopping as it reduces search costs of finding the right product, grants convenient access to product and price information, enables quick and easy comparison between products, has no restriction on shopping hours and no associated travel costs for the consumer (Hamad & Schmitz, 2019). This therefore parallels the shift towards convenience shopping, as identified in the previous section, as greater convenience, choice, lower prices, and increased store accessibility are all facilitated by the development of e-commerce offerings (Kirby-Hawkins, et al., 2019). Grocery retailers have thus responded in the UK as currently most major brands have some form of e or m-commerce offering with options such a click and collect, home delivery or locker usage (Vyt, et al., 2017). This includes the rise of pure online retailer Ocado whose market share continues to grow in light of increasing online demand (Butler, 2021), and Amazon's own online grocery retail offering developed alongside their foray into brick and mortar stores (Lee, 2023). However, the direction of this increasing trend is likely to be affected by the current "cost-of-living crisis" as consumers search around for deals that are only found in store (Eley, 2022; Romei, 2023), potentially halting the gains in market share of grocery e-commerce at least for the current moment.

Nevertheless, In terms of the effect on the gravity model formulation, this is expected to influence the distance decay and attractiveness relationship assumed by the model as retailers bear the cost of transporting the goods (Kirby-Hawkins, et al., 2019) whilst also determining which store the purchase is attributed to (Davies, et al., 2019). In terms of the distance decay relationship, consumers do not have to travel to the store themselves (unless it is a click-and-collect order), which instead shifts the costly burden onto retailers (Hood, et al., 2020). This means that in theory distance to the store does not matter for the consumer, but instead the retailer sets the distance threshold that they are willing to deliver to. Whilst there are theories of the geographical relationship within e-commerce shopping, such as the diffusion or innovation theory (Hood, et al., 2020), and attempts have been made to alter the functional form of the distance decay relationship to account for this (Beckers, et al., 2021), it could be argued that e-commerce would fundamentally alter the distance decay relationship that we see with in-store shopping. This is also complicated by the store or brand selection process whereby the retailer, not the consumer, determines which store a sale is attributed or assigned to. This means that the attractiveness relationship to stores is also altered by e-commerce as it is no longer the size of the store that determines a consumers preference, but rather the availability of products and the convenience of service (Davies, et al., 2019). While the selection of products is still likely to influence a consumer's choice to shop online this is more likely to be influenced by brand rather than a physical store (Singleton, et al., 2016). Such changes therefore separate the size and attraction relationship that is present in the models utilised in the previous chapters and disrupts the expected geographical relationship assumed by the gravity model.

8.3.4) New technology and rapid grocery delivery

The latest change in behaviour could be suggested to be a natural continuation of the convenience and e-commerce trends already discussed, with the rise of m-commerce and rapid delivery services in the UK. These services rose to prominence during the pandemic where a number of new technology firms began offering cheap groceries with free delivery on your doorstep in 10 to 20 minutes (Butler, 2021). They operate primarily using "dark stores" which locate in residential areas but have no in-store shopping experience, and focus primarily on major urban areas such as London, Paris, Amsterdam and Berlin (Bradshaw, 2022), although some began to enter smaller cities in the UK such as Manchester, Leeds and Liverpool (Wallop, 2021). Like convenience shops and e-commerce offerings they aim to target cash-rich and time-poor households such as young professionals wanting a quick meal, parents stuck at home with their children, and dinner party hosts missing a last minute ingredient (Butler, 2021). While online supermarket orders and deliveries have become a regular part of the grocery retailing market, these new rapid delivery services are aiming to target the convenience store market share (Butler, 2021). This is achieved through cheaper to operate "dark stores" (Wallop, 2021), and offering a 1,000-4,000 items which is much fewer than large format stores but is greater than what most convenience stores can offer (Butler, 2021).

While the prominence of these rapid delivery apps has prompted a response from several large retailers to either develop their own rapid grocery delivery offering or to partner with existing rapid delivery services (Bradshaw, 2022), there is also evidence of several firms going out of business already (Levingston, 2023). This was driven by the strategy of aiming to get as many customers as possible at any cost, funded by venture capitalists looking to support the next Amazon of grocery delivery (Nargi, 2022). This has meant that recent consolidation within the market, such as Getir buying out Gorillas in 2022, and reduction in competition as a result of firm failures, has left investors in the remaining companies hoping that cash-rich and time-poor consumers are still willing to pay a premium for rapid delivery of grocery products, even during a "cost of living crisis" (Levingston, 2023). This is despite a recent fall in orders in light of inflationary pressures and tightening purse strings, alongside potential back-lash against "dark stores" and their impact of the local high-street (Bradshaw, 2022; Nargi, 2022).

While this trend of rapid grocery delivery services is still clearly a developing market, the implications for gravity models are clear. Namely that adding in another channel through which consumers can get their groceries, alongside new brand names and reputations, is likely to further complicate an already complex grocery retailing landscape. This combines the potential influences of convenience store shopping, in terms short distances, irregular shops and small baskets, with the influence of e-commerce solutions, in terms of the effect of distance and attraction, which are either opposite to the assumptions of the gravity model or would fundamentally alter the form of the model chosen to be applied. Thus this shows that recent trends and developments in the grocery retailing market could be suggested to be moving even further away from the assumption of the existing gravity model formulation.

Overall, these changes therefore paint a picture of an evolving grocery retailing landscape in the UK over the last twenty years, with consumer behaviour departing from the underlying assumptions of the application of the gravity model in the grocery retailing sector. Namely, these changes generate a number of potential options for consumers in the form of new brands, new store formats, and new channels through which to engage with their grocery retailing shop. This adds complexity to the decisions that consumers have to make with regards to where they spend their money, a decision that is further influenced by socioeconomic conditions of the day. The more the market continues to evolve and the more options that become available for consumers, the more difficult it will be to try to estimate grocery store revenue using gravity models in the form and data that they have been used in this thesis. This is highlighted by the relatively poor results that have been achieved in previous chapters, with many of the changes in behaviours identified above potentially influencing this result.

8.4) Beyond gravity modelling

With the results presented in previous chapters and the discussion above about changing behaviours it thus becomes important to identify what alternative methods there may be to determine store location and potential revenue. In this regard it must be recognised that spatial interaction models are but one tool in the arsenal of retail location planners to decide where to locate stores (Clarkson, et al., 1996). Indeed, in many cases spatial interaction models are not used at all in retail location decisions, even by some of the biggest firms (Reynolds & Wood, 2010). Techniques that are often used instead of, or alongside, gravity models include: site visits by managers and staff (Clarke & Hayes, 2013), using a checklist of site location requirements (Robinson & Balulescu, 2018), spotting customers based on local foot traffic (Applebaum, 1966), creating travel time buffers and overlaying that onto demographic information (Benoit & Clarke, 1997), and regression methodologies that relate local area and store characteristics to store performance (Clarke & Hayes, 2013). Even major grocery retailers in the UK were still found to use gut feeling and site location visits to determine store locations, suggesting that gravity models were not the sole determinant of new store locations (Hood, et al., 2015). Thus, it is worth exploring how these techniques could potentially be used as alternatives to spatial interaction models to determine how desirable a store location may be. The issue however is that these models were originally not able to accurately and consistently estimate total store revenue and hence this is why spatial interaction models were originally adopted.

Nevertheless, some of the models and new datasets presented in the sections above, such as data science models or big social datasets could be used to augment the potential of these techniques. The aim of which would be to ensure that these existing methods could be used reliably and on a scale which would supplant the usage of spatial interaction models in store revenue estimation and store location decisions. For example, big datasets on geographical features, such as those extracted from OSM or Google Maps (Pearson, 2007), could be used alongside disaggregated population estimates from micro-simulated data (Nakaya, et al., 2007), and travel time buffers in a method developed around checklisting analysis. This could potentially be used to determine the optimal location of convenience stores and how much local revenue could be attracted to those stores by adding in new data layers that would generate a more accurate picture of local demand (Widaningurm, 2015). Such analysis could also be used in combination with machine learning methods such as Support Vector Machine (SVM) classification that can be used to classify whether a location is optimal for a new store based on the success and characteristic of existing stores (Widaningrum, 2017). This is aided by convenience stores' relatively small catchment area but could be complicated by which consumer group they target, whether that is residential, workplace or transient populations (Hood, et al., 2015).

Alternatives could also include the enhancement of regression techniques with the introduction of new large datasets and taking advantage of methodologies such as geographically weighted regression (Ozuduru & Varol, 2011), random forest regression with spatial lags (Barzin, et al., 2022), or the usage of neural networks. While the original usage of regression methodologies were criticised for their inability to accurately account for the influence of geography in modelling revenue (Birkin, et al., 2017), these new methods and data could resolve this issue by adding in geographically weighted features (Barzin, et al., 2022). Additionally, the integration of new datasets could highlight which features were most important in determining store revenue rather than traditional census and competition datasets. The aim of this methodology, therefore, would be to accurately and reliably estimate the probability of store success and potentially the amount of revenue it could generate. Thus, the "fourth age" of retail location planning could involve adaptation away from the spatial interaction model and taking advantage of machine learning and big data. However, it has also to be acknowledged that the implementation of many of these models may be limited by access to data or computing power, especially for smaller retailers (Reynolds & Wood, 2010).

8.5) Beyond Modelling

While both traditional and more recent methods could be adopted to replace spatial interaction modelling to estimate store revenue it is also worth asking the question as to whether grocery shopping behaviour, as discussed above, has become too complex to be accurately modelled and predicted reliably. That is, with the changes in consumer behaviour and market conditions, including the increase in brand diversity, the shift towards convenience shopping, the rise of e-commerce and the development of rapid delivery platforms, are we still able to estimate a stores revenue even with some of the new methodologies suggested above?

Batty (2018) presented the idea that we could model well highly routinous decisions with some regularity but argued that at some scale, especially when it comes to complex decision making processes, the predictability of behaviour can begin to break down. This is because while some individual decisions can be modelled, when the decisions are aggregated then this can create emergent behaviour that is inherently difficult to predict. To an extent this can be seen in grocery shopping behaviour due to the complex decisions as to the type of shop to undertake, which type of store to visit, which brand to shop with and what channel to engage with. While individual decisions could be modelled and predicted, such as whether different consumer groups are likely to use e-commerce channels or not, what products they are likely to buy and whether they will use convenience or large format stores, when aggregating these decisions to predict overall store revenue then the modelling focus begins to shift into the domain of the aggregation of complex behaviours that become difficult to model as a whole. This is especially so at the scale at which this thesis has attempted to develop spatial interactions due to the variety of decisions that individuals could take which would result in them shopping at a given store.

It could thus be argued that the spatial interaction model should only be applied to a small subset of behaviour of consumers that still perform their regular weekly shop, using their car, with large baskets at a large format store, as opposed to being adapted to attempt to model the complete set of modern grocery shopping behaviours. This is such that using the spatial interaction model is unlikely to predict total revenue to within a degree of accuracy that would be required in practice when attempting to model the whole behaviour set of consumers. However, there is further difficulty in being able to not only identify the subset of population that conforms with this assumption but also with identifying the size of the market that they represent in geographically diverse areas. A difficulty highlighted in the analysis performed in section 7.4. This would be further complicated when attempting to then integrate multiple different models, with their own errors and issues, to create an overall model of behaviour to estimate total store revenue.

On this basis therefore, this author believes that with grocery shopping behaviour becoming more and more complex and entwined with other decision-making processes, there is likely to become a wider separation between modelling individual behaviour and modelling the whole. That is, modelling total grocery store revenue at the scale which this thesis has been focusing on is likely to become more difficult and potentially be subject to the difficulties of emergent behaviour that make modelling and predictability challenging. While some of the developments highlighted above may resolve some of these issues in modelling individual decisions, modelling the whole may be beyond our current capabilities. It is however left up to future researchers to identify whether grocery retail decision making processes have become too complex to be reliably modelled (Batty, 2018).

8.6) Conclusion

In light of the results presented in the previous chapters, this chapter has highlighted potential avenues for future research to explore which could potentially lead to improved modelling performance. This includes the development of new model forms and adapting existing ones, such as utilising methods from the data science domain, the adaptation of agent-based-model, the development of new model formulations such as the radiation model, or the extension of existing modelling formulations such as the geo-lagged models. The aim of which is to resolve some of the issues of the Wilsonian form of the model such as being able to capture a wider range of relationships, understand individual level behaviour or account for spatial spillovers. The development of such models could also be supported by the integration of new datasets that can be used to show flows from origins to destinations, the attractiveness of stores or the emissiveness of origins. These would be used to support or even replace existing datasets by further illuminating the expected behaviour within the spatial interaction model such as the influence of non-residential origin grocery shopping, trip training or the influence of store location and age on store attractiveness. However, it must also be acknowledged that there may be issues with these developments including the computing power available or consistent access to these datasets that may limit their usage.

While developing new model formulations and utilising new datasets the literature must also be aware of how it is evaluating new model improvements and ensuring that these can be replicated. This involves developing a consistent approach to the metrics that are used to evaluate a spatial interaction model relative to other model formulations and the underlying data. This is to ensure that the models actually lead to improvements in modelling performance whilst still also being able to replicate the underlying data. A lack of consistency so far has meant that it is often not clear which models in the existing literature are most appropriate under different modelling scenarios. This then goes hand-in-hand with an open source approach to the implementation of spatial interaction models, with the development of open source tools, code and data. The aim of this would to allow new and existing researchers to easily implement existing model formulations and replicate previous results, thereby facilitating the verification of modelling performance. However, it must also be acknowledged that this may not be practical in all circumstances and domains such as due to issues over privacy, commerciality and sensitivity.

Finally, it must also be recognised that grocery shopping behaviour over the last 20 years has changed and become more complex with a variety of different types, formats and channels that consumers can engage with. This may mean that spatial interaction models, at least in their current format, may no longer be relevant to modelling grocery store revenue due to changes in behaviour away from those underlying the model assumptions. Thus, the literature has to also acknowledge how different retail models, such as checklists, buffer and overlay analysis, and regression may be utilised in light of new techniques and data. This must include an understanding of the way in which these models may be applied and the types of decisions they are modelling, paying attention to the urban system that grocery shopping decisions are made within (Batty, 2018). While it is expected that individual decision making processes are still able to be modelled, it is suggested that modelling the complete grocery retailing landscape in a full region may be beyond our current capabilities.

Chapter 9

Thesis conclusions

9.1) Introduction

The work within this thesis has successfully addressed the aim set out in the introduction: to advance the current understanding of the behaviour of spatial interaction models in a grocery retailing scenario. In completing this research there have been a number of specific achievements: a thorough analysis of the spatial interaction modelling literature, the identification of market changes that are currently affecting the grocery retailing industry in the UK, a spatially and temporally large scale application of spatial interaction models across three consumer regions in the UK, an evaluation of the robustness and consistency of the results presented in the existing literature, the development and application of a competing destination model at a regional scale. These achievements have thus contributed to the completion of the main thesis aim.

This chapter concludes the thesis by summarising the main research findings and achievements. Firstly, Section 9.2 will address the seven research objectives that were laid out in Chapter 1 and how they were achieved. Section 9.3 will then summarise some of the limitations of the research through a critique of the methodology and data limitations. Finally section 9.3 will provide a concluding statement on the success of the project and where research can continue in the future.

9.2) Summary of Research Findings

In the introduction to this thesis the broad aim was established as being able to advance our current understanding of spatial interaction model behaviour in a grocery retail setting in the UK. This aim was to be met by addressing a series of research objectives, for which this section will now demonstrate how each of these were met through the research carries out within this thesis.

1. **To examine and review the current spatial interaction modelling literature in terms of the development and usage of models to understand which models would be most appropriate for the application to a grocery retailing scenario in the UK.**

In order to contextualise the analysis undertaken as part of this thesis, Chapter 2 began with a thorough examination of the current state of the spatial interaction modelling literature, with an emphasis on how the models have developed over time. In doing, this chapter started with a review of retail location theories that were originally developed at the beginning of the 20th Century. The aim of this was to contextualise the choice of the spatial interaction model as the theory that could be practically implemented to determine the ideal retail store location and revenue. This discussion therefore presented an overview, alongside the strengths and weaknesses, of the theories of the principle of minimum differentiation, central place theory, and bid rent theory relative to that of spatial interaction models. It was identified that while the first three theories have continued to be developed and analysed since their initial conception, the benefit of spatial interaction models is that they are able to estimate an individual stores revenue and suggest ideal store locations rather than offering general conclusions. To this end, they have been routinely operationalised in both academic and industrial settings with the aim of locating retail stores in the urban environment.

Once the relative merits of the spatial interaction model were identified, the chapter then turned to an evaluation of the evolution of the spatial interaction model since the original conception of the idea by Reilly in 1929 and how this relates to form of the model that is most practical for an application in grocery retailing. This evaluation encompassed the evolution of the model through initial conception, formulation adjustments, the integration of utility theory, the attempts to draw parallels with physical laws, the development of a family of spatial interaction models and subsequent model iterations and adaptations. While this highlighted the variable history of the model formulation, and the many subsequent model formulations that have developed, it was concluded from this analysis that the most appropriate model formulation would be that of a production constrained model from the Wilsonian family of models. This was because of the ability of this formulation to constrain the total outflow of estimated revenue from each origin, thereby adding in additional information to the model estimation, and its adaptability to different scenarios such as the ability to disaggregate the model implementation to account for different behaviours.

In addition to the work in Chapter 2, this objective was also further addressed by the work in Chapter 4, 6 and 8. Chapter 4 explored current issues in the literature surrounding methods of model calibration and evaluation in relation to the Wilsonian family of spatial interaction models whilst also developing an initial model implementation at a city level scale. This chapter concluded that the most appropriate calibration method was that of Poisson Regression and that multiple metrics should be used to evaluate models to ensure consistency in conclusions. Chapter 6 then built on the results from previous chapters by providing a more thorough insight into the contributions of recent applications of the spatial interaction model in a grocery setting, particularly in reference to the scale and calibration methods they developed. This work highlighted that recent models implemented using anonymised loyalty card data focused on temporally and spatially small subsets of stores that were calibrated using an iterative calibration method, but that the results suggested could not be consistent repeated. Thus finally Chapter 8 evaluated recent developments in and around the spatial interaction modelling literature and how they could potentially improve modelling performance in the future. This highlighted the potential for new models and data to continue to advance the literature forward but that implementation in a grocery retailing environment may be restricted by the type of data that is typically used in these scenarios.

2. **To review the literature on grocery retailing in the UK to be able to identify social, economic and political pressures that have, and are currently, influencing the market and to relate these influences and market developments to models that have been used to determine store location.**

It then became necessary to identify how the grocery retailing market in the UK has changed in response to social, economic and political pressures over the years so as to determine how this form of the spatial interaction may behave when applied to a grocery retailing scenario. Chapter 3 therefore tackled this objective by providing a comprehensive overview of both the history of grocery retailing in the UK since the 1960s, as to recognise factors that may relate to spatial interaction model usage, and how these developments have gone hand-in-hand with the adaptation and implementation of existing and new methods for evaluation store locations.

Firstly this chapter discussed changes that the grocery retailing market in the UK underwent over the last 60 years. From this discussion it was identified that the industry went through several key periods in response to different economic, social and political pressures. The market was initially dominated in the 1960s by small format stores owned by independent retailers, but land and pricing regulation changes enabled the development of the superstore format. This change facilitated the development of nationally focused brands who built large format stores all over the country, resulting in increased concentration of market share among a small group of brands. Further regulation change, along with social and economic pressures, in the late 1990s and early 2000s then forced a change in focus towards the expansion of existing stores, the development of convenience store offerings and initial exploration of e-commerce platforms. This therefore foresaw a fundamental change in the market which continues until today with the increase in competitive pressures from expanding international discounters, the continued rise of convenience retailing and the growing importance of e-commerce sales. Therefore highlighting that while the industry was focused on the development of large format stores until the early 2000s, more recent changes in consumer behaviour are leading to increased focus on convenience store formats and e-commerce platforms.

The Chapter then continued by exploring the influence that these changes had on the methods that were used by retailers to decide where to locate new stores or close down existing ones. It was shown that when competition was mostly between local small format retailers, managers or owners could rely on their local knowledge and gut feeling to pick new store locations. However, with increasing competition, costs of constructing new stores and expanding brand reach, retailers sought more advanced and objective techniques. Early non-computational methods involved processes such as check-listing and customer-spotting, which were later enabled by advances in data and technology in the mid-1980s to allow for improved visualisation and objectivity. This new technology also enabled the use of more advanced techniques such as regression analysis and spatial interaction models. Thus, while there was a considerable history of spatial interaction models in the literature, grocery retailers themselves did not necessarily adopt spatial interaction models until the 1990s when advances in technology and data enabled their consistent use and evaluation. Furthermore, retailers have begun to adapt large data sets and data science techniques to model store location in light of changes in consumer behaviour that may longer reflect the expected behaviour of a spatial interaction model.

This chapter therefore highlights that the adoption of spatial interaction models in grocery retailing location decisions began in the 1980s at the height of new large format store building, when shopping was regular, with a single purpose, undertaken by car and with large baskets. The adoption of this technique then continued as the costs increased of locating new stores, with a specific focus on large format stores. However, in the late 2000s and early 2010s new shopping behaviours of multi-purpose trips, convenience shopping, and the development of e-commerce platforms began to influence store location, format and channel usage. Thus, some retailers have been developing and using new techniques from the domain of data science in conjunction with new large datasets. Therefore showing that while the spatial interaction model assumptions were likely to be reflect reality in the 1990s and 2000s, it could be suggested that new shopping behaviours and formats are likely to influence current modelling performance.

3. **To identify issues surrounding the implementation of spatial interaction models in practice and to develop a working model based on anonymised loyalty card data.**

While Chapter 2 and 3 identify the theoretical considerations of spatial interaction modelling it is also important to identify the practical implications of applying a spatial interaction model to a grocery retailing scenario using anonymised loyalty card data. To this extent, Chapter 2 recognises that there are many different model formulations that could be used in practice, but the discussion makes it clear that the most appropriate model formulation, and one that is commonly used in practice, is the production constrained spatial interaction model from the Wilsonian family of spatial interaction model. This is because of its ability to incorporate constraints on the total outflow of revenue from each origin and the total of flow within the system overall. Chapter 3 then highlights potential factors that may influence modelling performance, notably the consumer behaviour associated with convenience shopping, multi-purpose trips, and e- and m-commerce usage.

Chapter 4 then builds on this foundation by discussing how to implement this form of the model in practice. The first issue in this regard was recognised as how to calibrate the model based on anonymised loyalty card data. Thus, section 4.1 discussed the calibration methods identified in the existing literature including linear regression, maximum likelihood calibration methods, iterative calibration, Poisson Regression and Binomial Regression. The conclusion drawn from this discussion was that the most appropriate method for the practical implementation of the production constrained model, based on the data that was available, was that of the Poisson Regression. Such a conclusion was reached on the basis that this method, relative to the other discussed, was able to naturally incorporate zero flows, allow for the integration of constraints within the model, and there are programmatic tools available for the simple implementation of this modelling formulation in practice.

The second issue then identified was that of modelling evaluation whereby it is important to be able to evaluate and validate the output of the models relative to the underlying data and previous modelling implementations, but that there was a variety of methods that were used in the existing literature. It was thus recognised that commonly used metrics included the average trip distance (ATD), R^2 and SRMSE measures but that often these fail to capture the true performance of spatial interaction models. Thus, further discussions highlighted other potential metrics that could be used to inform the debate of model performance, including the use of the Common Part of Commuters (CPC) metric, the Mean Absolute Error (MAE) and the Sørensen Similarity Index (SSI) that could be used alongside the previous metrics to ensure consistent model evaluation.

Finally section 4.3 utilises these findings by developing an initial application of the spatial interaction model at a city level scale. This section therefore discusses the methodology for implementation, including data gathering and integration. The results from this firstly show that these models can be implemented using the anonymised loyalty card data that we have access to and that the parameters derived from the anonymised loyalty card data can then be used to estimate total store revenue. Secondly, it showed that when scaling up from modelling anonymised loyalty card data using Poisson Regression to estimating total revenue using the trained parameters, the range of individual store errors increased. Finally, the model implemented performed considerably better at predicting large store format revenue, in line with the results from the existing literature that acknowledged that small store formats revenue were difficult to model with a spatial interaction model. Therefore, these findings laid the foundation for the implementation of the model at scale.

4. **To develop and apply a spatial interaction model at a regional and yearly scale so as to identify how modelling performance changes and responds at scale and whether there are any factors that influence the modelling performance.**

This objective speaks to the main contribution of this thesis, to extend our understanding of the behaviour of spatial interaction models when applied to a grocery retailing scenario, and is primarily addressed through the analysis in Chapter 5. This analysis builds on the exploration and evaluation presented in the previous chapters by applying a spatial interaction model to three consumer regions in the UK. To start with a system wide model is applied to all three regions using the exponential form of the model that is calibrated using Poisson regression. This application focused specifically on modelling large format stores revenue, given the results from the previous chapter and those presented in the literature in regards to modelling convenience format stores. This showed that when aggregating the loyalty card predictions at the store level, the majority of individual store errors were concentrated within a +/-20% boundary. When using the calibrated parameters to estimate total store revenue however the range of errors increased such that errors could be seen between 60% underprediction and 120% overprediction. The model is then further developed through the implementation of an origin-disaggregated form of the model whose results show minor improvements over the system wide implementation. Then data on travel time, store competition and households estimates are also integrated into the model for a single region, which while leading to increases in the amount of revenue attracted to stores, do not resolve the issue of a large variance in store errors. This exploration therefore showed that the current model implementations were unable to account for the variation in store conditions and consumer behaviour when scaled up to predict total store revenue.

It therefore became important to identify factors that could be related to, or affecting, these results at an individual store level across all regions and model implementations. Factors that were identified included store level characteristics, store revenue generation and surrounding area characteristics. This examination suggested that the only consistent factor related to modelling performance was the age of the store. This showed that an older store saw consistent overprediction from the model and that a younger store saw consistent underprediction. These results therefore suggested that a younger store was likely to be more attractive than just its store size would suggested, which is unaccounted for in the current model implementations. While data limitations restricted further exploration of this relationship, the result hinted at potential avenues for future research to explore, which could lead to improvements in the performance of spatial interaction models.

The robustness of these results and conclusions were then validated by the application of the models across every week in a year and through the implementation of a cross-validation study. In the first instance the yearly results showed that while there was some seasonal variation over the year in model performance, the results presented for the single week application were not an outlier. Notably, only region 1 showed considerable influence of tourist revenue where there was considerable variation in model performance, notably during the school holidays, but that the individual week the model was first applied on was one where there was little tourist influence and good overall model performance. In contrast, region 2 and 3 showed smaller amounts of variation throughout the year, suggesting that non-residential tourist demand was not a considerable influence of modelling performance. These results were then supported by the cross-validation study which showed that the modelling performance was consistent across parameters that were derived from other regions. Notably, the difference between regions was consistently larger than the variation in performance within regions when applying different parameter pairs, therefore suggesting that the current model specification is unable to accurately scale up to model total revenue at the regional scale.

The results presented in this chapter therefore contribute to the literature through the application and evaluation of a spatial interaction model at a geospatial and temporal scale that has not been seen before in the literature. In doing so it highlighted that with the current model specification and data availability these models were unable to replicate the performance achieved in the previous literature. Furthermore, the range of errors achieved at this scale suggests that these models could not be consistently used in practice. Thus, this chapter showed that there are issues in the current modelling specification or methodology that could influence these results.

5. **To replicate the modelling implementations from the most recent and up to date papers so as to examine whether the suggested results and performance from them can be replicated on data that we have available.**

In light of the results from Chapter 5, showing the performance of the models seen in the literature could not be replicated at the regional scale, it was then necessary to utilise the data available to replicate the scenarios that were previous explored. This was achieved through the analysis presented in Chapter 6 which begins by evaluating the differences between previous implementations of the spatial interaction model and our own. From this, it was identified that there were two main differences in model implementation: modelling scale and calibration method. In terms of the former the previous literature was only able to explore subsets of size four and sixteen of large format stores, thereby different to our own application to 29, 47 and 60 stores.

Furthermore, they calibrated through model through an iterative method focused on replicating the average trip distance metric from the underlying anonymised loyalty card data. Therefore, such differences could be explored in terms of how they may influence modelling performance and whether their results could be consistent replicated.

The Chapter then continues firstly with an analysis of the influence of scale on modelling performance in an attempt to replicate the results in the previously literature. This exploration was achieved through the implementation of the spatial interaction model on 47 different groups of stores of sizes 4, 5, 7, 10 and 16. The results from this analysis showed that while there could have been conditions that led to the performance suggested in the previous literature at scales of 4 and 16 stores, based on the distribution of group errors it is unlikely that these results could be consistently replicated. This was such that while a single group for both 4 and 16 store subsets showed performance similar to that seen in the literature, this was only for one group out of 47. Thus, while not being able to integrate non-residential demand, the models at this scale did not show consistent performance. This analysis was then extended over an entire year for both of the well-performing groups of stores, showing that there is likely to be seasonal variation in modelling performance at this scale such that the range of errors over a single year were larger than for the individual week. Therefore further supporting the conclusion of the inability to replicate the results of the previous literature.

The alternative was to then examine the influence of the method of calibration on modelling performance at the regional scale. This was achieved through the implementation of a grid search to explore a range of parameters that could produce a model where the average trip distance metric was equal to 1. The results from this analysis showed that, rather than a single parameter value, a range of parameter values produced an ATD value equal to 1. Further exploration of these range of parameters showed that while an ATD value of 1 could be achieved, in line with previous literature, there were no significant improvements in terms of the mean store error or range of errors relative to the model calibrated using Poisson Regression methodology. Indeed, across the range of parameters it was show that individual stores responses different such that no parameter pairings were likely to lead to a convergence in modelling performance to the results seen in the literature. Thus, this analysis showed that model calibration was not the issue in model implementation and that as before the models failed to account for the range of behaviour and store conditions at the regional levels.

The conclusions drawn from the analysis presented in this chapter, with reference to the objective above, was that while the results from the previous literature could not be replicated on a consistent basis. This was such that neither scale or calibration method was seen to resolve the issues inherent in the modelling application and that future research or application would be unlikely to replicate the results suggested. Therefore, this analysis supported the argument that at this scale, with the current data and modelling formulation, the spatial interaction model is unable to account for the variance in underlying store characteristics and consumer behaviour at the regional level.

6. **To implement and examine alternative forms of the spatial interaction model to identify the influence of additional store based factors on modelling performance.**

This objective was addressed through the work presented in Chapter 7 in which a competing destinations model was adapted, a model that integrated store age was developed, and a model that focused on large basket revenue was implemented. Firstly, as discussed in the chapter, the competing destinations models was originally developed to integrate the theory of two level decision making into the spatial interaction model. It was thus suggested that by not accounting for this influence, and thus either competitive or agglomerative forces, then the models implemented in the previous chapters could be misspecified. On this basis, a non-disaggregated competing destinations model implemented to estimate total store revenue for stores in Region 2. A grid search methodology was used to identify the influence of either agglomeration or competition on modelling performance across a range of potential distances. The results from this analysis showed that there was a trade-off in modelling performance throughout the parameter space in terms of the mean store error, standard deviation of errors and average trip distance. Notably, the parameters that produce a mean store error close to zero were associated with a high standard deviation of errors and average trip distance and vice versa. The results presented therefore showed that the competing destinations model was unlikely to resolve the performance issues of the models presented in Chapters 4, 5 and 6. Further examination of the individual store results also showed that the forces of competition or agglomeration also lead to a divergence of individual store errors away from zero, suggesting that there was no parameter pairing that would lead to improved modelling performance.

The second modelling formulation explored therefore was one that integrated store age into the model as a measure of store attractiveness. The development of this model was based on the results introduced in Chapter 5, Section 5.5. which showed a clear and consistent correlation between store age and individual store errors across three regions and model specifications. The aim was thus to identify whether the integration of store age would reduce the strength of the correlations with store errors and lead to improved model performance. The results presented however showed that for the sixteen stores that the model was trained on, the integration of store age did not alter the original model correlation or improve modelling performance. This was driven by a positive store age attractiveness parameter which suggested that the older a store was then the more attractive it was to consumers. It was argued therefore that this could be related to information available about a store to consumers. This performance was consistent when the model was implemented across all weeks in the year and across all other subsets of sixteen stores within the region. Thus, while these results suggested that the integration of additional factors into the model, as with the competing destinations model, affects modelling performance, in this case store age did not lead to improvements in the results.

The final model in the chapter, in contrast to the previous two, thus attempted to examine whether subsets of behaviour could be identified and modelled more accurately than the total regional behaviour. This aimed to reduce the influence of convenience and multi-purpose trip behaviour on model calibration by developing a model trained on large basket shopping. It was expected that large baskets would better represent the behaviour aligned with spatial interaction modelling implementations of travel from home for single purpose regular shopping trips by car. The results, while showing different behaviours to the whole region, did not show improvements in modelling performance at predicting either total loyalty card or total revenue sales in terms of the distribution of individual store errors. Thus, further highlighting the difficulty of capturing and modelling all subsets of behaviour at the regional scale with spatial interaction models.

Thus, three alternative modelling implementations and their influence on modelling performance were explored. Importantly, these formulations were chosen based on results presented in the previous chapters. However they were unable to resolve the issues of large mean and variance of store errors at the regional scale. This exploration therefore supports the previous conclusions of the inability of the model, in its current form and with current data, to be able to accurately and consistently estimate total store grocery revenue. Furthermore, the analysis contributes to the literature by examining the influence of competition, agglomeration, store age and behavioral subsets on modelling performance. Therefore, the results presented could be used to determine future research directions into the influence of store age on a larger scale and the development of alternative methods to accurately estimate the total revenue available to spend at local stores. The identification of such avenues of future research was thus further discussed in Chapter 8.

7. **To offer potential avenues for future research to explore so as to continue the development of the spatial interaction model in reference to its application in grocery retailing.**

In light of the results from the analysis presented in chapters 4, 5, 6 and 8 the research reported in this thesis has the potential to be taken forward in a number of directions. This aim was thus addressed in Chapter 8 which evaluated potential future avenues for research, identifying those that are likely to be most fruitful in reference to the application of spatial interaction models in a grocery retailing environment. To this end, the first research future research direction identified was the development of new model formulations or the adaptation of existing methods towards spatial interaction modelling purposes. New modelling methods were suggested to come from a variety of potential directions including the integration of data science methods such as neural networks or tree based algorithms, the adaptation of agent-based-models to account for individual behaviour in shopping decision making, the development of new modelling formulations such as the radiation model, or the adaptation of existing formulations to account for perceived issues of the Wilsonian form of the model. It was thus suggested that the most fruitful areas of potential future research was the development of data science methods or the construction of agent based models. This was because the Data Science methods would be able to identify and work with new forms of spatial interaction relationships beyond the restrictive form of the Wilsonian formulation. Furthermore, the development of agent-based-models could account for the variety of behaviours that individuals exhibit across a range of consumer groups, including the influence of convenience shopping, multi-purpose trips and usage of e- and m-commerce channels. This future research however has to be able to overcome issues of calibration, data integration and evaluation to ensure that they lead to actual improvements in model performance.

Alongside the development of new model formulations, it was also suggested that the exploration of new datasets could improve the performance of spatial interaction models in grocery retailers. The idea was that these new datasets could be used to support our understanding, or even replace, existing data on the flows from origins to destinations, the attractiveness of individual stores and the emissiveness of origins. Potential datasets identified included mobile phone generated data or large scale spending data (such as cross brand loyalty cards or credit cards) that could be used to examine individual behaviour in terms of convenience shopping, multi-purpose trips and non-residential shopping behaviour, thus how they may influence the assumptions of the spatial interaction model. However, this would likely be complicated by difficulty in cleaning the data and linking it to actual store purchases. Other datasets could include the use of open source image datasets or footfall data that could be used to inform estimates of the attractiveness of individual stores such as the fascia, likely age of store, entrance direction and consumer traffic which could improve the identification of store attractiveness, particularly for small format stores. The difficulty with this however would be the ability to collect the relevant data at scale and to generate accurate and reliable estimates of store attractiveness. Finally, methods such as micro-simulation or receipt level data could be used to inform estimates of origin emissiveness or available revenue by generating more reliable estimates of population, spending and brand attractiveness.

However, to benefit from the exploration of both new models and datasets it is important for future research to carefully examine the influence of evaluation metrics and to develop open source infrastructures. This is because the literature needs a reliable measure, or group of measures, to be able to identify how accurate new model or data are relative to the underlying data and previous model formulations. Without a clear and consistent way evaluating modelling improvements it was argued that it would become difficult to identify where the literature future research and for current researchers and practitioners to select the most relevant models and data. It was also further argued that this would be facilitated by the development of open source infrastructure for the implementation of spatial interaction models. This would likely include the development of open source packages in a variety of programming languages and the open sourcing of code and data for model implementations. Such infrastructure would thus support the continued development and implementation of models by allowing quick and easy replication of existing results and the adaptation of code or models to new areas of research and data. Arguably, the two avenues of research above cannot proceed without future research spent achieving these two goals of clear and consistent evaluation methods and the development of open source infrastructure.

Finally, it was put forward that the literature should also acknowledge the potential of other retail location methods through the utilisation of advancements in both new data sources and modelling methods. This would encompass an evaluation of how changes in consumer behaviour influence grocery shopping attitudes and trends relative to those assumed by the spatial interaction model and identify new methods of locating stores. These results could then be used to inform the future implementation of checklisting, buffer and overlay analysis, and regression techniques to account for these behaviours. Examples utilising such information was suggested to include the use of big geospatial datasets and taking advance of data science methods. The aim of which would be to identify whether traditional techniques, with updated data and methods, could be used to account for changes in behaviour more accurately than spatial interaction models. Thus, it would advance our understanding of behaviour in grocery retailing.

9.3) A Critique of the Methodology

This thesis has advanced the understanding of how spatial interaction models behave when applied to grocery retailing scenario in the UK and their potential limitations. However, despite this success there are inevitable issues in the methodological approach which, whilst noted in various places within the thesis, should also be explicitly acknowledged here.

In developing regional scale models following the initial city level implementation in Chapter 4, the focus was on exclusively modelling large format stores. The purpose of this was to limit the complexity of the modelling implementations whilst acknowledging the results from the previous literature which highlighted the difficulty of modelling small format stores with spatial interaction models (Waddington, et al., 2019). However, it could be suggested that this decision was taken too early and in light of limited results. Notably, this could have affected modelling performance in terms of predicting total store revenue by adding in the complexity of having to scale down the estimated revenue to that only spent at large format stores. This was achieved by calculating the percentage amount of revenue derived from large format stores from our partner retailer and then scaling down the estimates expenditure from each output area. While it was expected that since our partner is a national grocery retailer with a large market share, thus being representative of the market, an alternative could have been to use a value of convenience market share or to adjust the percentage relative to the surrounding convenience store penetration. The former solution would likely have a similar critique to the current solution used, while the latter could more accurately account for the influence of format availability on shopping behaviour. Nevertheless, this calculation would likely add additional complexity into the modelling formulation and would likely require an accurate estimate of small format store revenue. Thus, with limited data and ability to model small format store revenue, this is an avenue that future research could explore in terms of its influence on modelling performance.

Nevertheless, the underlying estimation of total available revenue could also be suggested to be improved. In this implementation estimates of household revenue available are based on the Living Costs and Food Survey output area supergroup classification for 2017. While this allows the attribution of expected revenue from each output area according to their sociodemographic group, as opposed to the application of a single value, it could be suggested that this disaggregation does not go far enough. Indeed, it could be suggested that the LCFS estimates for supergroups do not account for regional or socioeconomic variance in expenditure across the whole country in the same way that regional values or the OAC group estimates would (ONS, 2022). Thus affecting the estimates of total available revenue per output area. Furthermore the number of households in each output area is estimated using data from the 2011 census while the model application is developed using anonymised loyalty card data and total store revenue from 2017. This could thus lead to differences in the number of estimates households and the actual number of households in each output area, thereby affecting the estimates of total output area revenue (Newing, et al., 2015).

While this was addressed extent in the new data model implementation in Chapter 5, a more accurate estimate of households may have led to more considerable effects in model performance.

Furthermore in this thesis store attractiveness is proxied by the total store size in square feet. To this extent it has been acknowledged in previous research that factors other than store size are likely to affect how attractive a store may be to consumers (Newing, et al., 2020; Fornari, et al., 2020). This may include factors such as the distance to the street, store accessibility, store frontage, and other services locating in the store (Birkin, et al., 2017). The potential influence of this can be seen in Section 5.5 where the age of stores was seen to be consistently related to individual store performance within the model. While the identification of attractiveness was in line with the usage in previous research (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021), it could thus be suggested that a more comprehensive index of store attractiveness may lead to more accurate and consistent models and hence to alternative conclusions. Thus, while we were limited in the data that could used consistently to estimate store attractiveness in this thesis, future research may examine whether other store attractiveness factors may influencing modelling results. Notably, inclusion of other variables may lead to a reduction in the variance of individual store errors by more accurately accounting for how attractive an individual store is.

The estimation of distance and travel time within the model could also be suggested to influence the performance of the models presented so far. Firstly, the majority of models that were implemented in Chapter 5 used as-the-crow-flies distance between origins and destinations. This is likely to affect modelling performance by failing to account of the true distance or travel time consumers travel to the store from home. The effect of this could be seen in the model presented in section 5.4.3 whereby the use of travel time resulted in an increase in revenue attributed to our partners store in the model due to the relative accessibility of the stores compared to the competition. Nevertheless, the introduction of drivetime moved the mean store error further away from zero and also did not lead to improvement in the range of store errors. This could potentially be related to the use of the OSRM API for the calculation of travel time which uses estimates of potential travel speed across various road types (Huber & Rust, 2016). This could potentially influence modelling performance by not accounting for the true travel speed on different roads in response to traffic, data would could have been extracted from sources such as the Google Maps API (Salonen & Toivonen, 2013; Google, 2022). However, in the absence of other available open source datasets, this was the best estimate that could be used. Therefore, future research could examine the influence of different travel time estimates.

Finally, it is worth mentioning the potential influence that behaviour change could have in the model implementation. This includes the influence of multi-purpose shopping, the increase in convenience shopping across all formats, and the development and integration of e- and m-commerce in the grocery retailing sector. These behaviours are not accounted for in this modelling implementation due to limited access to data on non-residential demand in terms of both estimates of population or their behaviour, and the lack of penetration of e-commerce at an output area scale to allow for consistent evaluation due to GDPR constraints. Arguably, these limitations have been what the recent literature (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021) has been attempting to address. In doing so they have identified the influence that these factors can have on modelling performance in terms of the ability to estimate grocery store revenue. To mitigate their potential influence however, this thesis has focused on analysis where seasonal demand fluctuations are limited and on large format stores where daytime demand is limited and e-commerce demand has been removed. However, it could still be argued that not accounting for these factors, such as by integrating further demand layers, may be influencing the results, and thus future research should continue the exploration of the strength and direction of this influence at the regional scale.

9.4) Concluding Remark

This thesis has explored the performance of spatial interaction models in the UK grocery retailing sector using anonymised loyalty card data from a national grocery retailer. This thesis has thus developed our understanding of the performance of these models at scale and how robust these modelling formulations are across different scenarios. While a stated aim of the spatial interaction modelling development in grocery retailing is to create a national scale model, the results from this thesis has suggested that a spatial interaction model not be appropriate. While there is still research to be undertaken, methods to be tweaked, techniques to be improved, and data collected, this work has succeeded in exploring the complexities of the application of spatial interaction models in grocery retailing. It is hoped that the novel application of the models developed in this thesis will help to guide future research to estimate grocery store revenue at scale in the UK.

Bibliography

Aiello, L., Quercia, D., Schifanella, R. & Prete, L., 2020. Tesco Grocery 1.0, a large-scale dataset of grocery purhcases in London. *Scientific Data,* 7(57).

Alexander, N. & Morlock, W., 1992. Saturation and Internationalization: The future of grocery retailing in the UK. *International Journal of Retail & Distribution Management,* 20(3), pp. 33-39.

Applebaum, W., 1966. Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales. *Journal of MArketing Research,* 3(2), pp. 127-141.

Arentze, T., Oppewal, H. & Timmermans, H., 2005. A Multipurpose Shopping Trip Model to Assess Retail Agglomeration Effects. *Journal of Marketing Research,* 42(1), pp. 109-115.

Aversa, J., Doherty, S. & Hernandex, T., 2018. Big Data Analytics: The New Boundaries of Retail Location Decision Making. *PApers in Applied Geography,* 4(4), pp. 390-408.

Balali, V., Rad, A. & Golparvar-Fard, M., 2015. Detection, classifictaion, and mapping of U.S. traffic signs using google street view images for roadway inventory management. *Visualization in Engineering,* 3(15).

Baron, S., Harris, K., Leaver, D. & Oldfiel, B., 2001. Beyond convenience: the future for independent food and grocery retailers in the UK. *The International Review of Retail, Distribution and Consumer Research,* 11(4), pp. 395-414.

Barzin, S., Avner, P., Rentschler, J. & O'Clery, N., 2022. *Where Are All the Jobs ? : A Machine Learning Approach for High Resolution Urban Employment Prediction in Developing Countries,* Washington: World Bank.

Batty, M., 2018. Predictability, Complexity, and Inventing the Future. In: *Inventing Future Cities.* s.l.:MIT Press, pp. 1-17.

Batty, M. & Mackie, S., 1972. The calibration of gravity, entropy, and related models of spatial interaction. *Environment and Planning,* Volume 4, pp. 205-233.

Beckers, J. et al., 2021. Incorporating E-commerce into Retail Location Models. *Geographical Analysis,* pp. 1-20.

Beckers, J., Cárdenas, I. & Verhetsel, A., 2018. Identifying the geography of online shopping adoption in Belgium. *Journal of Retailing and consumer services,* Volume 45, pp. 33-41.

Benoit, D. & Clarke, G., 1997. Assessing GIS for retail location planning. *Journal of Retailing and Consumer Services,* 4(4), pp. 239-258.

Birkin, M., Clarke, G. & Clarke, M., 2010. Refining and Operationalizing Entropy-Maximizing Models for Business Applications. *Geographical Analysis,* 42(4), pp. 422-445.

Birkin, M., Clarke, G. & Clarke, M., 2017. Chapter 2. In: M. Birkin, G. Clarke & M. Clarke, eds. *Retail Location Planning in an Era of Multi-Channel Growth.* 1st ed. New York: Routledge.

Birkin, M., Clarke, G. & Clarke, M., 2017. *Retail Location Planning in an Era of Multi-Channel Growth.* 1st ed. New York: Routledge.

Black, J. & Salter, R., 1975. A statistical evaluation of the accuracy of a family of gravity models. *Proceedings of the Institution of Civil Engineers,* 59(1), pp. 1-20.

Black, W., 1991. A note on the use of correlation coefficients for assessing goodness-of-fit in spatial interaction models. *Transportation,* Volume 18, pp. 199-206.

Black, W., 1995. Spatial interaction modeling using artifical neural networks. *Journal of Transport Geography,* 3(3), pp. 159-166.

Blum, B. & Goldfarb, A., 2006. Does the Internet defy the law of gravity?. *Journal of International Economics,* 70(2), pp. 384-405.

Bonnetain, L. et al., 2021. TRANSIT: Fine-grained human mobiluty trajectory inference at scale with mobile network signaling daa. *Transportation Research Parrt C: Emerging Technologies,* Volume 130.

Bradshaw, T., 2022. *Financial Times.* [Online]
Available at: https://www.ft.com/content/aba36d58-f1d5-45f5-bb6d-2ee7d9aa07d5
[Accessed 06 May 2023].

Bradshaw, T., 2022. *Financial Times.* [Online]
Available at: https://www.ft.com/content/be07c7ae-765d-4d29-b8cd-26ad8b4ccaed
[Accessed 06 May 2023].

Brown, S., 1992. The wheel of retail gravitation. *Environment and Planning A,* 24(0), pp. 1409-1429.

Brown, S., 1993. Retail location theory: evolution and evaluation. *International Review of Retail, Distribution and Consumer Research,* 3(2), pp. 185-229.

Buckley, M., Cowan, C. & McCarthy, M., 2007. The convenience food market in Great Britain: Convenience food lifestyle (CFL) segments. *Appetite,* 49(3), pp. 600-617.

Burger, M., Van Oort, F. & Linders, G., 2009. On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflation Estimation. *Spatial Economic Analysis,* 4(2), pp. 167-190.

Butler, S., 2012. Tesco and rivals turn against huge stores as internet shopping takes over. *The Guardian*, 4 March, pp. https://www.theguardian.com/business/2012/mar/04/online-shopping-changes-hypermarket-strategy.

Butler, S., 2021. *The Guardian.* [Online]
Available at: https://www.theguardian.com/business/2021/may/22/fast-food-the-new-wave-of-delivery-services-bringing-groceries-in-minutes
[Accessed 06 May 2023].

Butler, S., 2023. *The Guardian.* [Online]
Available at: https://www.theguardian.com/business/2023/feb/28/british-shoppers-grocery-prices-kantar
[Accessed 06 May 2023].

Carrothers, G., 1956. An historical Review of the Gravity and Potential Concepts of Human Interactions. *Journal of the American Institute of Planners,* 22(2), pp. 94-102.

CDRC, 2022. *Consumer Data Research Centre.* [Online]
Available at: https://www.cdrc.ac.uk/
[Accessed 6 November 2022].

Cesario, F., 1974. More on the generalised trip distribution model. *Journal of Regional Science,* 14(3), pp. 389-397.

Cesario, F., 1975. Least-squares estimation of trip distribution parameters. *Transportation Research,* 9(1), pp. 13-18.

Cesario, F., 1975. Linear and Nonlinear Regression Models of Spatial Interaction. *Economic Geography,* 51(1), pp. 69-77.

Chen, Y., 2015. The distance-decay function of geographical gravity model: Power law or exponential law?. *Chaos, Solitons & Fractals,* Volume 77, pp. 174-189.

Clarke, G., 1998. Changing methods of location planning for retail companies. *GeoJournal,* Volume 45, pp. 289-298.

Clarke, G. & Hayes, S., 2013. GIS and Retail Location Models. In: G. Cliquet, ed. *Geomarketing: Methods and Strategies in Spatial Marketing.* s.l.:s.n.

Clarke, G., Thompson, C. & Birkin, M., 2015. The emerging geography of e-commerce in British retailing. *Regional Studies, Regional Science,* 2(1), pp. 371-391.

Clarke, I., Kirkup, M. & Oppewal, H., 2012. Consumer satisfaction with local retail diversity in the UK: effects of supermarket access, brand variety, and social deprivation. *Environment and Planning A,* 44(8), pp. 1896-1911.

Clarke, K., 2014. Cellular automata and agent-based models. *Handbook of regional science,* pp. 1217-1233.

Clarke, M. & Birkin, M., 2018. Spatial Interaction Models: from Numerical Experiments to Commercial Applications. *Applied Spatial Analysis and Policy,* Volume 11, pp. 713-729.

Clarkson, R., Clarke-Hill, C. & Robinson, T., 1996. UK supermarket location assessment. *International Journal of Retail & Distribution Management,* 24(6), pp. 22-33.

Converse, P., 1949. New Laws of Retail Gravitation. *Journal of Marketing,* 14(3), pp. 379-384.

Cullinan, J. & Duggan, J., 2016. A School-Level Gravity Model of student Migration Flows to Higher Education Institutions. *Spatial Economic Analysis,* 11(3), pp. 294-314.

Davies, A., Dolega, L. & Arribas-Bel, D., 2019. Buy online collect in-store: exploring grocery click&collect using a national case study. *International Journal of Retail & Distribution Management,* 47(3), pp. 278-291.

Davies, R. & Guy, C., 1987. The Statitical Modeling of Flow Data When the Poisson Assumption is Violated. *Geographical Analysis,* 19(4), pp. 300-314.

De Beule, M., Poel, D. & Weghe, N., 2014. An extended Huff-model for robustly benchmarking and predicting retail network performance. *Applied geography,* Volume 46, pp. 80-89.

De Mello-Sampayo, F., 2016. A Spatial Analysis of Mental Healthcare in Texas. *Spatial Economic Analysis,* 11(2), pp. 152-175.

de Vries, J., Nijkamp, R. & Reitveld, P., 2009. Exponential or Power Distance-Decay for Commuting? An Alternative Specification. *Environment and Planning A: Economy and Space,* 41(2), pp. 461-480.

Dearden, J. & Wilson, A., 2011. The Relationship of Dynamic Entropy Maximising and Agent-Based Approaches in Urban Modelling. In: A. Heppenstall, A. Crooks, L. See & M. Batty, eds. *Agent-BAsed Models of Geographical Systems.* Dordrecht: Springer, pp. 705-720.

dunnhumby, 2022. *HuYu.* [Online]
Available at: https://www.huyuapp.co.uk/
[Accessed 5 December 2022].

East, R., Lomax, W., Willson, G. & Harris, P., 1994. Decision Making and Habit in Shopping Times. *European Journal of Marketing,* 28(4), pp. 56-71.

Eley, J., 2022. *Financial Times.* [Online]
Available at: https://www.ft.com/content/bf79741d-7530-4865-b4e1-51a6fa098a38
[Accessed 06 May 2023].

Ellickson, P. & Grieco, P., 2013. Wal-Mart and the geography of grcoery retailing. *Journal of Urban Economics,* Volume 75, pp. 1-14.

Elms, J. et al., 2010. 30 years of retail change: where (and how) do you shop?. *Internatoinal Journal of Retail & Distribution Management,* 36(11/12), pp. 817-827.

Ewing, G., 1974. Gravity and Linear Regression Models of Spatial Interaction: A Cuationary Note. *Economic Geography,* 50(1), pp. 83-88.

Fischer, M. & Gopal, S., 1994. Artifical neural networks: a new approach to modeling interregional telecommunication flows. *Journal of Regional Science,* 34(4), pp. 503-527.

Fischer, M., M, R. & Hlavackova-Schindler, K., 2003. Neural network modelling of constrained spatial interaction flows: design, estimation, and performance issues. *Journal of Regional Science,* 43(1), pp. 35-61.

Fischer, M. & Reggiani, A., 2004. Spatial Interaction Models: From the Gravity to the Neural Network Approach. *Contributions to Economic Analysis,* Volume 266, pp. 317-346.

Fischer, M. & Reismann, M., 2002. A methodology for Neural Spatial Interaction Modeling. *Geographical Analysis,* 34(3), pp. 207-228.

Flowerdew, R. & Aitkin, M., 1982. A METHOD OF FITTING THE GRAVITY MODEL BASED ON THE POISSON DISTRIBUTION. *Journal of Regional Science,* 22(2), pp. 191-202.

Fornari, E., Grandi, S., Menegatti, M. & Fornari, D., 2020. Discounters versus Supermarkets and Hypermarkets: what drives store-switching. *The International Review of Retail, Distribution and Consumer Research,* 30(5), pp. 555-574.

Fotheringham, A., 1983. A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning A,* Volume 15, pp. 15-36.

Fotheringham, A., 1988. Note-Consumer Store Choice and Choice Set Definition. *Marketing Science,* 7(3), pp. 211-312.

Fotheringham, A. & O'Kelly, M., 1989. The Calibration of Spatial Interaction Models. In: A. Fotheringham & M. O'Kelly, eds. *Spatial Interaction Models: Formulations and Applications.* London: Kluwer Academic Publishers, pp. 43-65.

Gale, C., Singleton, A., Bates, A. & Longley, P., 2016. Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Informatino Science,* Volume 12, pp. 1-27.

Geolytix, 2022. *Geolytix.* [Online]
Available at: https://drive.google.com/file/d/1VQFcL1ysZDrHY7gLPjjoJ3b9nLE3nVub/view
[Accessed 8 November 2022].

Georganos, S. et al., 2019. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International,* 36(2), pp. 121-136.

Gibson, M. & Pullen, M., 1972. Retail turnover in the East Midlands: A regional application of a gravity model. *Regional Studies,* 6(2), pp. 183-196.

Giraud, T., Cura, R., Viry, M. & Lovelance, R., 2021. *project osrm.* [Online]
Available at: http://project-osrm.org/
[Accessed October 2022].

Google, 2022. *Google.* [Online]
Available at: https://developers.google.com/maps/documentation/distance-matrix/overview
[Accessed 8 November 2022].

Griffith, D., Fischer, M. & LeSage, J., 2017. The spatial autocorrelation problem in spatial interaction modelling: a comparison of two common solutions. *Letters in Spatial Resource Sciences,* 10(1), pp. 75-86.

Guy, C., 1987. Recent advances in spatial interaction modelling: an application to the forecasting of shopping travel. *Environment and Planning A,* 19(2), pp. 173-186.

Guy, C., 1988. Retail planning policy and large grocery store development: a case study in South Wales. *Land Development Studies,* Volume 5, pp. 31-45.

Guy, C., 1991. Spatial Interaction modelling in retail planning practice: the need for robust statistical measures. *Environment and Planning B: Planning and design,* 18(2), pp. 191-203.

Guy, C., 1992. Estimating Shopping Centre Turnover: A Review of Survey Methods. *International Journal of Retail & Distribution Management,* 20(4), pp. 18-23.

Guy, C., 1995. Retail store development at the margin. *Journal of Retailing and Consumer Services,* 2(1), pp. 25-32.

Guy, C., 1996. Corporate strategies in food retailing and their local impacts: a case study of Cardiff. *Environment and Planning A: Economy and Space,* 28(9), pp. 1575-1602.

Hallsworth, A. & Jayne, M., 2000. Problems surrounding the retail property market in a complex connurbation: the case of Stoke-on-Trent, UK. *Property Management,* 18(4), pp. 206-217.

Hallsworth, A., Kervenoael, R., Elms, J. & 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.

Hamad, H. & Schmitz, M., 2019. The Impact of Demographic Variables and Consumer Shopping Orientations on the Purchasing Preference for Different Product Categories in the Context of Online Grocery Shopping. *Journal of Marketing and Consumer Research,* Volume 52.

Harris, B., 1964. A note on the probability of interaction at a distance. *Journal of Regional Science,* 5(2), pp. 31-35.

Harris, B. & Wilson, A., 1978. Equilibrium Values and Dynamics of Attractiveness Terms in Production Constrained Spatial Interaction Models. *Environment and Planning A: Economy and Space,* 10(4), pp. 371-388.

Hasova, L., 2021. *Lenka's blog.* [Online]
Available at: https://lenkahas.com/post/gsoc.html
[Accessed 21 October 2022].

Haynes, K. & Fotheringham, A., 1985. *GRavity and Spatial Interaction Models.* s.l.:s.n.

Heppenstall, A. et al., 2020. Future Developments in Geographical Agent-Based Models: Challenges and Opportunities. *Geographical Analysis,* pp. 1-16.

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

Hilton, B., Sood, A. & Evans, T., 2020. Predictive limitations of spatial interaction models: a non-Gaussian analysis. *Scientific Reports.*

Hood, N., Clarke, G. & 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.

Hood, N., Urquhart, R., Newing, A. & Heppenstall, A., 2020. Sociodemographic and spatial disaggregation of e-commerce channel use in the grocery market in Great Britain. *Journal of Retailing and Consumer Services,* Volume 55.

Huber, S. & Rust, C., 2016. Calculate travel time and distance with OpenStreetMap using the Open Source Routing Machine (OSRM). *The Stata Journal,* 16(2), pp. 416-423.

Hu, C. et al., 2020. Classification and mapping of urban canyon geometry using Google Street View Images and deep multitask learning. *Building and Environment,* Volume 167.

Huff, D., 1963. A Probabilisic Analysis of Shopping Center Trade Areas. *Land Economics,* 39(1), pp. 81-90.

Huff, D., 1964. Defining and Estimating a Trading Area. *Journal of Marketing,* 28(3), pp. 34-38.

Huff, D., 1966. A Programmed Solution for Appoximating an Optimum Retail Location. *Land Economics ,* 42(3), pp. 293-303.

Hu, P. & Pooler, P., 2002. An empirical test of the competing destinations model. *Geographical Systems,* Volume 4, pp. 301-323.

Hu, T. et al., 2021. Human mobility data in the COVID-19 pandemic: characteristics, applications and challenges. *International Journal of Digital Earth,* 14(9), pp. 1126-1147.

Joseph, L. & Kuby, M., 2011. Gravity Modeling and its Impacts on Location Analysis. In: H. Eiselt & V. Marianov, eds. *Foundations of Location Analysis.* New York: Springer, pp. 423-443.

JRF, 2022. *Joseph Rowntree Foundation.* [Online]
Available at: https://www.jrf.org.uk/data/household-spending
[Accessed 7 December 2022].

Kang, J. et al., 2018. Building instance classification using street view images. *ISPRS Journal of Photogrammetery and Remote Sensing,* 145(A), pp. 44-59.

Karlaftis, M. & Vlahogianni, E., 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies,* 19(3), pp. 387-399.

Kerkman, K., Martens, K. & Meurs, H., 2017. A multilevel spatial interaction model of transit flow incorporating spatial and network autocorrelation. *Journal of Transport Geography,* Volume 60, pp. 155-166.

Khawaldah, H., Birkin, M. & Clarke, G., 2012. A review of two alternative retail impact assessment techniques: the case of Silverburn in Scotland. *Town Planning Review,* 82(2), pp. 233-260.

Kirby-Hawkins, E., Birkin, M. & Clarke, G., 2019. An investigation into the geography of corporate e-commerce sales in the UK grocery market. *Urban Analytics and City Science,* 46(6), pp. 1148-1164.

Kishore, N. et al., 2020. Measuring mobility to monitor travel and physical distancing interventions: a common framewok for mobile phone data analysis. *The Lancet Digital Health,* 2(11), pp. 622-628.

Knudsen, D. & Fotheringham, S., 1986. Matrix Comparison, Goodness-of-Fit, and spatial Interaction Modelling. *International Regional Science Review,* 10(2), pp. 127-147.

Kollewe, J., 2022. *The Guardian.* [Online]
Available at: https://www.theguardian.com/business/2022/sep/13/average-uk-food-bill-rises-grocery-inflation-aldi-morrisons
[Accessed 06 May 2023].

Kor, Y., 2019. *The Conversation.* [Online]
Available at: https://theconversation.com/shopping-trends-mean-blocking-the-big-sainsburys-asda-merger-may-not-protect-customers-116083
[Accessed 22 October 2020].

Krisztin, T. & Fischer, M., 2015. The Gravity Model for International Trade: Specification and Estimation Issues. *Spatial Economic Analysis,* 10(4), pp. 451-470.

Kritsztin, T. & Fischer, M., 2015. The Gravity Model for International Trade: Specification and Estimation Issues. *Spatial Economic Analysis,* 10(4), pp. 451-470.

Langford, M., 2013. An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data. *Geographical Analysis,* 45(3), pp. 324-344.

Lata, R., Proff, S. & Brenner, T., 2018. The influence of distance types on co-patenting and co-publishing in the USA and Europe over time. *Annals of Regional Science,* Volume 61, pp. 49-71.

Lee, D., 2023. *Financial Times.* [Online]
Available at: https://www.ft.com/content/f132299d-3b31-4248-8ddf-81182c5930fc
[Accessed 06 May 2023].

Lee, M. & Pace, R., 2005. Spatial Distribution of Retail Sales. *The Journal of Real Estate Finance and Economics,* 31(1), pp. 53-69.

Lenormand, M., Bassolas, A. & Ramasco, J., 2016. Systematic comparison of trip distribution laws and models. *Journal of Transport Geography,* Volume 51, pp. 158-169.

Lenormand, M., Huet, S., Gargiulo, F. & Deffuant, G., 2012. A Universal Model of Commuting Networks. *PLoS ONE,* 7(10).

Levingston, I., 2023. *Financial Times.* [Online]
Available at: https://www.ft.com/content/23858b1a-b105-416d-859b-be3f95cd4290
[Accessed 06 May 2023].

Li, X. et al., 2021. Emerging geo-data sources to reveal human mobility dynamics during COVID-19 pandemic: opportunities and challenges. *Computational Ubran Science,* 22(1).

Li, Y. L. L., 2012. Assessing the impact of retail location on store performance: A comparison of Wal-Mart and Kmart stores in Cincinnati. *Applied Geography,* 32(2), pp. 591-600.

Lovelace, R. & Nowosad, J., 2022. *CRAN.* [Online]
Available at: https://cran.r-project.org/web/packages/simodels/simodels.pdf
[Accessed 7 November 2022].

Macal, C. & North, M., 2014. *Introductory tutorial: agent-based modeling and simulation.* s.l., IEEE.

Marchetti, S., Giusti, C. & Pratesi, M., 2016. The use of Twitter data to improve small area estimates of households' share of food consumption expenditure in Italy. *AStA Wirtschafts- und Sozialstatistisches Archiv ,* Volume 10, pp. 79-93.

Masucci, A., Serras, J., Johansson, A. & Batty, M., 2013. Gravity v Radiation Models: On the Importance of Scale and Heterogeneity. *arXiv.*

McCann, P., 2013. *Modern urban and regional economics.* 2nd ed. Oxford: Oxford University Press.

McCulloch, K. et al., 2021. Ensemble model for estimating continential-scale patterns of human movement: a case study of Australia. *Scientific Reports,* Volume 11.

Mendes, A. & Themido, I., 2004. Multi-outlet retail site location assessment. *International Transactions in Operational Research,* Volume 11, pp. 1-18.

Merino, M. & Ramirez-Nafarrate, A., 2016. Estimation of retail sales under competitive location in Mexico. *Journal of Business Research,* 69(2), pp. 445-451.

Motta, V., 2019. Estimating Poisson pseudo-maximum-lieklihood rather than log-linear model of a log-transformed dependent variable. *RAUSP Management Journal,* 54(4), pp. 508-518.

Mozolin, M., Thill, J. & Usery, E., 2000. Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. *Transportation Research Part B: Methodological,* 34(1), pp. 53-73.

Nakaya, T. et al., 2007. Combining microsimulation and spatial interaction models for retail location analysis. *Journal of Geographical Systems,* Volume 9, pp. 345-369.

Nargi, L., 2022. *The Guardian.* [Online]
Available at: https://www.theguardian.com/environment/2022/jul/29/dark-stores-ultra-fast-delivery-app-bad-for-workers-and-communities
[Accessed 06 May 2023].

Nazir, S., 2021. Lidl and Aldi to be two years late in reaching store opening targets. *Retail Gazette*, 1 July`.

Newing, A., Clarke, G. & Clarke, M., 2013. Identifying seasonal variations in store-level visitor grocery demand. *International Journal of Retail & Distribution Management,* 41(6), pp. 477-492.

Newing, A., Clarke, G. & Clarke, M., 2013. Visitor expenditure estimation for grocery store location planning: a case study of Cornwall. *The International Review of Retail, Distribution and Consumer Research,* 23(3), pp. 221-244.

Newing, A., Clarke, G. & Clarke, M., 2014. Exploring small area demand for grocery retailers in tourist areas. *Tourism Economics,* 20(2), pp. 407-427.

Newing, A., Clarke, G. & Clarke, M., 2015. Developing and Applying a Disaggregated Retail Location Model with Extended Retail Demand Estimations. *Geographical Analysis,* 47(3), pp. 219-239.

Newing, A., Clarke, G. & Clarke, M., 2018. Applied spatial modelling for retail planning in tourist resorts. *Aplied Spatial Modelling,* 46(11/12), pp. 1117-1132.

Newing, A., Hood, N. & Sterland, I., 2020. Planning support systems for retail location planning. In: S. Geertman & J. Stillwell, eds. *Handbook of Planning Support Science.* Cheltenham, UK: Edward Elgar Publishing, pp. 459-470.

O'Kelly, 2011. Applied Retail Location Models Using Spatial Interaction Tools. In: A. Fotheringham & P. Rogerson, eds. *The SAGE Handbook of Spatial Analysis.* London: SAGE Publications, Ltd, pp. 420-442.

ONS, 2013. *Office for National Statistics.* [Online]
Available at:
https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography#:~:text=There%20are%20now%20181%2C408%20OAs,MSOA)%20in%20England%20and%20Wales.
[Accessed 19 July 2021].

ONS, 2020. *Open Geography Portal.* [Online]
Available at: https://geoportal.statistics.gov.uk/
[Accessed October 2021].

ONS, 2021. *Living Costs and Food Survey,* s.l.: ONS.

ONS, 2022. *ONS.* [Online]
Available at:
https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure/datasets/familyspendingworkbook3expenditurebyregion
[Accessed 8 November 2022].

Openshaw, S., 1998. Neural network, genertic, and fuzzy logic models of spatial interaction. *Environment and Planning A,* Volume 30, pp. 1857-1872.

Oshan, T., 2016. A Primer for working with the Spatial Interaction modeling (SpInt) module in the python spatial analysis library. *Region,* 3(2), pp. 11-23.

Oshan, T., 2017. *Arizona State University.* [Online]
Available at: https://keep.lib.asu.edu/_flysystem/fedora/c7/187558/Oshan_asu_0010E_17476.pdf
[Accessed 16 December 2022].

Ozuduru, B. & Varol, C., 2011. Spatial Statistics Methods in Retail Location Research: A Case Study of Ankara, Turkey. *Procedia Environmental Sciences,* Volume 7, pp. 287-292.

Pappalardo, L., Simini, F., Barlacchi, G. & Pellungrini, R., 2019. scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data. *arXiv.*

Pearson, J., 2007. A Comparative Business Site-Locayion Feasability Analysis using Geographic Information System and the Gravity Model. *Papers in Resource Analysis,* Volume 9, pp. 1-10.

Perktold, J., Seabold, S. & Taylor, J., 2022. *statsodels.* [Online]
Available at: https://www.statsmodels.org/stable/index.html
[Accessed 9 November 2022].

Picci, L., 2010. The internationalization of inventive activity: A gravity model using patent data. *Research Policy,* Volume 39, pp. 1070-1081.

Piovani, D. et al., 2018. Measuring accessibility using gravity and radiation models. *Royal Society Open Science,* Volume 5.

Popkowski Leszczyc, P., Sinha, A. & Sahgal, A., 2004. The effect of multi-purpose shopping on pricing and location strategy for grocery stores. *Journal of Retailing,* 80(2), pp. 85-99.

Pourebrahim, N., Sultana, S., Niakanlahiji, A. & Thill, J., 2019. Trip distribution modeling with Twitter data. *Computers, Environment and Urban Systems,* Volume 77.

Puga, D., 2010. The Magnitude and causes of agglomeration economies. *Journal of Regional Science,* 50(1), pp. 203-219.

Rains, T. & Longley, P., 2021. The provenance of loyalty card data for urban and retail analytics. *Journal of Retailing and Consumer Services.*

Recker, W. & Schuler, H., 1981. Destination choice and processing spatial information: some empirical tests withy alternative constructs. *Economic Geography,* 57(4), pp. 373-383.

Reigahinha, T., Godhino, P. & Dias, J., 2017. Portugese food retailers - Exploring three classic theories of retail location. *Journal of Retailing and Consumer Services,* Volume 34, pp. 102-116.

Reilly, W., 1929. *Methods for the study of retail relationships,* s.l.: University of Texas.

Reynolds, J. & Wood, S., 2010. Location decision making in retail firms: Evolution and challenge. *International Journal of Retail and Distribution Management,* 38(11), pp. 828-845.

Robinson, D. & Balulescu, A., 2018. Comparison of methods for quantifying consumer spending on retail using publicly available data. *International Journal of Geographic Information,* 32(6), pp. 1061-1086.

Roig-Tierno, N., Baviera-Puig, A., Buitrago-Vera, J. & Mas-Verdu, F., 2013. The retail site location decision process using GIS and the analytical hierarchy process. *Applied Geography,* Volume 40, pp. 191-198.

Romei, V., 2023. *Financial Times.* [Online]
Available at: https://www.ft.com/content/9efcc6b4-688b-4004-bb87-440e0438da79
[Accessed 06 May 2023].

Rowe, F., Lovelace, R. & Dennet, A., 2022. Spatial Interaction Modelling: A Manifesto. In: s.l.:Edward Elgar Books.

Safegraph, 2022. *Safegraph.* [Online]
Available at: https://www.safegraph.com/
[Accessed 11 November 2022].

Salonen, M. & Toivonen, T., 2013. Modelling travel time in urban networks: comparable measures for private car and public transport. *Journal of Transport Geography,* Volume 31, pp. 142-153.

Santos Silva, J. & Tenreyro, S., 2006. The Log of Gravity. *The Review of Economics and Statistics,* 88(4), pp. 641-658.

Santos Silva, J. & Tenreyro, S., 2010. On the existence of the maximum likelihood estimates in Poisson regression. *Economics Letters,* Volume 107, pp. 310-312.

Santos Silva, J. & Tenreyro, S., 2011. Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics letters,* 112(2), pp. 220-222.

Seabold, S. & Perktold, J., 2010. *statsmodels: Econometric and statistical modeling with Python.* s.l., 9th Python in Science Conference.

Simini, F., Barlacchi, G., Luca, M. & Pappalardo, L., 2021. A Deep Gravity model for mobility flows generation. *Nature communications,* 12(6576).

Simini, F., Gonza´lez, M., Maritan, A. & Baraba´si, A., 2012. A universal model for mobility and migration. *Nature,* Volume 484, pp. 96-100.

Simkin, L., 1990. Evaluating a Store Location. *International Journal of Retail and Distribution Management,* Volume 18, pp. 33-38.

Singleton, A., Dolega, L. & Riddlesden, L. P., 2016. Measuing the spatial vulnerability of retail centres to online consumptino through a framework of e-resilience. *Geoforum,* Volume 69, pp. 5-18.

Smith, L., Ma, M., Widener, M. & Farber, S., 2022. Geographies of grocery shopping in major Canadian Cities. *EPD: Urban Analytics and City Science.*

Song, Z., 2021. The geography of online shopping in China and its key drivers. *Environment and Planning B: Urban Analytics and City Science,* 0(0), pp. 1-16.

Spadon, G., Carvalho, A., Rorigues-Jr, J. & Alves, L., 2019. Reconstructing commuters using machine learning and urban indicators. *Science Reports,* 9(11801).

Sprinboard, 2022. *Sprinboard.* [Online]
Available at: https://www.spring-board.info/
[Accessed 6 November 2022].

Statista, 2022. *Statista.* [Online]
Available at: https://www.statista.com/statistics/330171/grocery-retail-market-value-in-the-united-kingdom-uk/
[Accessed 7 December 2022].

Stefanouli, M. & Polyzos, S., 2017. Gravity vs radiation model: two approaches on commuting in Greece. *Transportation Research Procedia,* Volume 24, pp. 65-72.

Stewart, J., 1941. An inverse distance variation for certain social influences. *Science,* 93(2404), pp. 89-90.

Stewart, J., 1947. Empirical mathematicla rules concerning the distribution and equilibirum of population. *Geographical Review,* 37(3), pp. 461-485.

Stewart, J., 1947. Suggested Principles of "Social Physics". *Science,* 106(2748), pp. 179-180.

Stewart, J., 1948. Concerning "social physics". *Scientific American,* 178(5), pp. 20-23.

Stewart, J., 1950. The Development of Social Physics. *American Jrournal of Physics,* 18(239).

Sturley, C., Newing, A. & Heppenstall, A., 2018. Evaluating the potential of agent-based modelling to capture consumer grocery retail store choice behaviours. *The International Review of Retaill, Distribution and Consumer Research,* 28(1), pp. 27-56.

Suhara, Y., M, B., Bozkaya, B. & Pentland, A., 2019. *Validating Gravity-Based Market Share Models Using Large-Scale Transactional Data.* [Online]
Available at: https://arxiv.org/abs/1902.03488
[Accessed 21 April 2021].

Themido, I., Quintino, A. & Leitão, J., 1998. Modelling the retail sales of gasoline in a portugese metropolitan area. *International Transactions in Operational Research,* 5(2), pp. 89-102.

Thompson, C., Clarke, G., Clarke, M. & Stillwell, J., 2012. Modeling the future opportunities for deep dicsount food retailing in the UK. *The International Review of Retail, Distribution and Consumer Research,* 22(2), pp. 143-170.

Thorsen, I. & Gitlesen, J., 2002. Empirical evaluation of alternatiive model specifications to predict commuting flows. *Journal of Regional Science,* 38(2), pp. 273-292.

Tiefelsdorf, M. & Boots, B., 1995. The specification of constrained interaction models using the SPSS loglinear procedure. *Geographical Systems,* Volume 2, pp. 21-38.

Trasberg, T., Chesire, J. & Longley, P., 2018. *Integration New Measures of Retail Unit Attractiveness into Spatial Interaction Models.* Leicester, GIS Research UK (GISRUK).

University of Leeds, 2021. *University of Leeds.* [Online]
Available at: https://environment.leeds.ac.uk/faculty/news/article/5383/lida-announces-partnership-with-sainsbury-s
[Accessed 29 November 2022].

Van Droogenbroeck, E. & Van Hove, L., 2017. Adoption of Online Grocery Shopping: Personal or Household Characteristics. *Journal of Internet Commerce,* 16(3), pp. 255-286.

Vickers, D. & Rees, P., 2007. Creating the UK National Statistics output area classification. *Statistics in Society,* 170(2), pp. 379-403.

Vyt, D., Jara, M. & Cliquet, G., 2017. Grocery pickup creation of value: Customers' benefit vs. spatial dimension. *Journal of Retailing and Consumer Services,* Volume 39, pp. 145-153.

Waddington, T. , Clarke, G., Clarke, M. & Newing, A., 2019. Accounting for Temporal Demand Vairations in Retail Location Models. *Geographical Analysis,* 51(4), pp. 426-447.

Waddington, T., Clarke, G., Clarke, M. & Newing, A., 2018. Open all hours: spatiotemporal fluctuations in U.K. grocery store sales and catchment area demand. *The International Review of Retail, Dsirtibution and Consumer Research,* 28(1), pp. 1-26.

Wallop, H., 2021. *The Guardian.* [Online]
Available at: https://www.theguardian.com/lifeandstyle/2021/jun/12/the-rise-of-on-demand-grocery-deliveries
[Accessed 06 May 2023].

Wang, F., Chen, C., Xiu, C. & Zhang, P., 2014. Location analysis of retail stores in Changchun, China: A street centrality perspective. *Cities,* 41(A), pp. 54-63.

Wang, J., Tsia, C. & Lin, P., 2016. Applying spatial-temporal analysis and retail location theory to public bikes site selection in Taipei. *Transportation Research Part A,* Volume 94, pp. 45-61.

Widaningrum, D., 2017. *GIS and SVM Approach for Convenience Store Location Analysis.* New York, NY, United states, Association for Computing Machinery.

Widaningurm, D., 2015. A GIS - Based Approach for Catchment Area Analysis of Convenience Stores. *Procedia Computer Science,* Volume 72, pp. 511-518.

Willberg, E., Järv, O., Väisänen, T. & Toivonen, T., 2021. Escaping from Cities during the COVID-19 Crisis: Using Mobile Phone Data to Trace Mobility in Finlnd. *International Journal of Geo-Information,* 10(2).

Williams, H. & Senior, M., 1977. A Retail Location Model with Overlapping Market Areas: Hotelling's Problem Revisited. *Urban Studies,* Volume 14, pp. 203-205.

Wilson, A., 1967. A statistical theory of spatial distribution models. *Transportation research,* 1(3), pp. 253-269.

Wilson, A., 1971. A family of Spatial Interaction Models, and Associated Developments. *Environment and Planning A: Economy and Space,* 3(1), pp. 1-32.

Wilson, A., 2007. Boltzmann, Lotka and Volterra and spatial structure evolution: an integrated methodology for some dynamical systems. *Journal of the Royal Society Interface,* Volume 5, pp. 865-871.

Wilson, A., 2010. Entropy in Urban and Regional Modelling: Retropsect and Prospect. *Geographical Analysis,* 42(4), pp. 364-394.

Wilson, S., 1976. Statistical notes on the evaluation of calibrated gravity models. *Transportation Research,* 10(5), pp. 343-345.

Woelwer, A. et al., 2022. *CRAN.* [Online]
Available at: https://cran.r-project.org/web/packages/gravity/index.html
[Accessed 7 November 2022].

Wood, S. & Browne, S., 2007. Convenience store location planning and forecasting - a practical research agenda. *International Journal of Retail & Distribution Management,* 35(4), pp. 233-255.

Wood, S., Lowe, M. & Wrigley, N., 2006. Life after PPG6 - recent UK food retailer responses to planning regulation tightening. *The International Review of Retail, Distribution and Consumer Research,* 16(1), pp. 23-41.

Wood, S. & McCarthy, D., 2014. 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. & Reynolds, J., 2011. The intrafirm context of retail expansion planning. *Environment and Planning A,* 43(10), pp. 2468-2491.

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.

Wu, C., Smith, D. & Wang, M., 2021. Simulating the urban spatial structure with spatial interaction: A case study of urban polycentricty under different scenarios. *Computers, Environment and Urban Systems,* Volume 89.

Yeghikyan, G. et al., 2020. *Learning Mobility Flows from Urabn Features with Spatial Interaction Models and Neural Networks.* s.l., s.n.

Ying, F. et al., 2019. Customer mobility and congestion in supermarkets. *arXiv.*

Zipf, G., 1942. The Unity of Nature, Least-Action, and Natural Social Science. *Sociometry,* 5(1), pp. 48-62.

Zipf, G., 1946. Some determinants of the Ciruclation of Information. *The American Journal of Psychology,* 59(3), pp. 401-421.

Zipf, G., 1946. The P1 P2/D Hypothesis: on the Intercity Movement of Persons. *American Sociological Review,* 11(6), pp. 677-686.

Appendix

Appendix A

Exponential Decay Regional Results

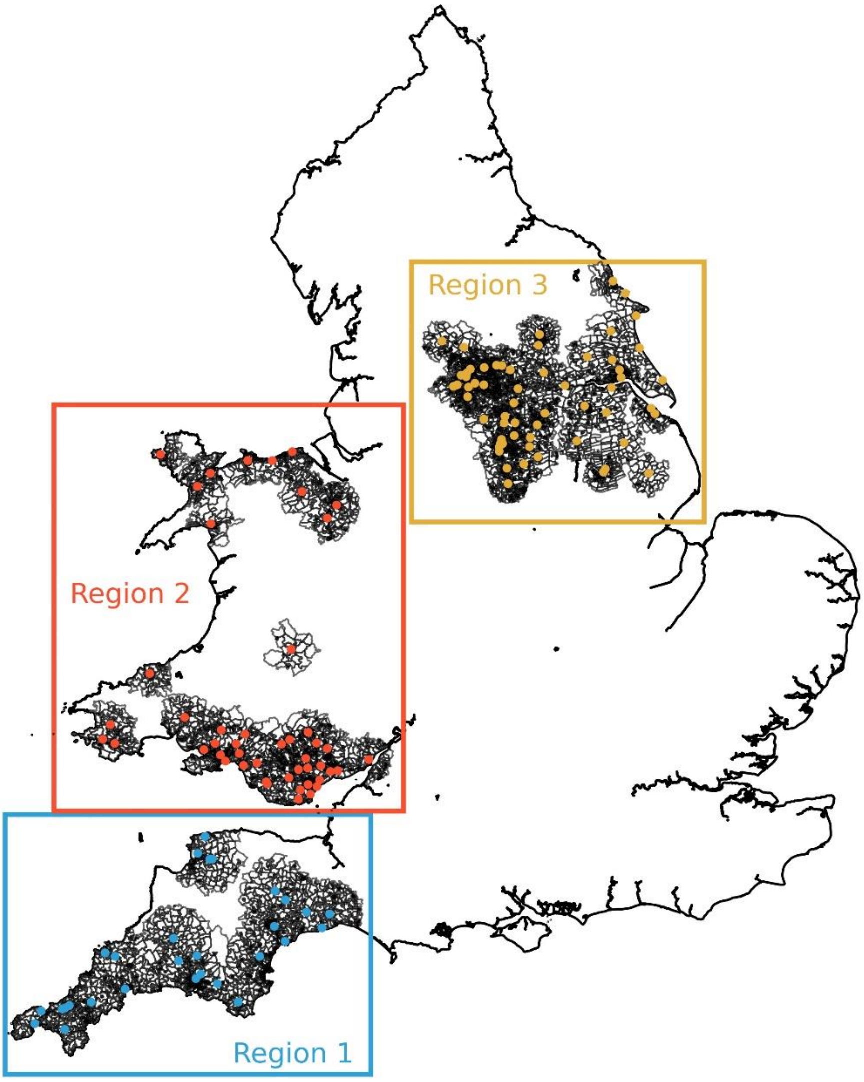

Figure 45 - Regional representations of partner stores modelled and output areas revenue is assumed to be derived from

Table 12 - Exponential decay metrics for each region by each supergroup

Metrics		Region 1		Region 2		Region 3	
	Supergroup	Loyalty card	Total Revenue	Loyalty card	Total Revenue	Loyalty card	Total Revenue
Pseudo R^2	1	0.913	N/A	0.893	N/A	0.927	N/A
	2	0	N/A	0.981	N/A	0.978	N/A
	3	0	N/A	0	N/A	0.93	N/A
	4	0	N/A	0.965	N/A	0.982	N/A
	5	0.947	N/A	0.935	N/A	0.951	N/A
	6	0.933	N/A	0.918	N/A	0.954	N/A
	7	0.960	N/A	0.942	N/A	0.974	N/A

	8	0.923	N/A	0.906	N/A	0.964	N/A
AIC	1	564885	N/A	473559	N/A	285854	N/A
	2	0	N/A	9724	N/A	9428	N/A
	3	0	N/A	0	N/A	2674	N/A
	4	0	N/A	34708	N/A	55083	N/A
	5	201556	N/A	318312	N/A	324027	N/A
	6	228234	N/A	809927	N/A	547383	N/A
	7	42254	N/A	123978	N/A	71852	N/A
	8	301877	N/A	1098647	N/A	366450	N/A

Table 13 - Metrics for travel time data in Region 2

Metrics	**Base model**		**Disaggregated model**	
	Loyalty card	Total Revenue	Loyalty card	Total Revenue
R^2	0.949	0.708	0.950	0.701
Pseudo R^2	0.898	N/A	N/A	N/A
RMSE	149.09	1209.82	147.45	1228.62
SRMSE	0.589	4.780	0.583	4.854
ATD	1.000	1.098	1.000	1.091
MAE	53.84	482.09	51.58	479.24
AIC	2504948	N/A	N/A	N/A
SSI	0.176	0.118	0.177	0.118
CPC	0.894	0.499	0.898	0.501

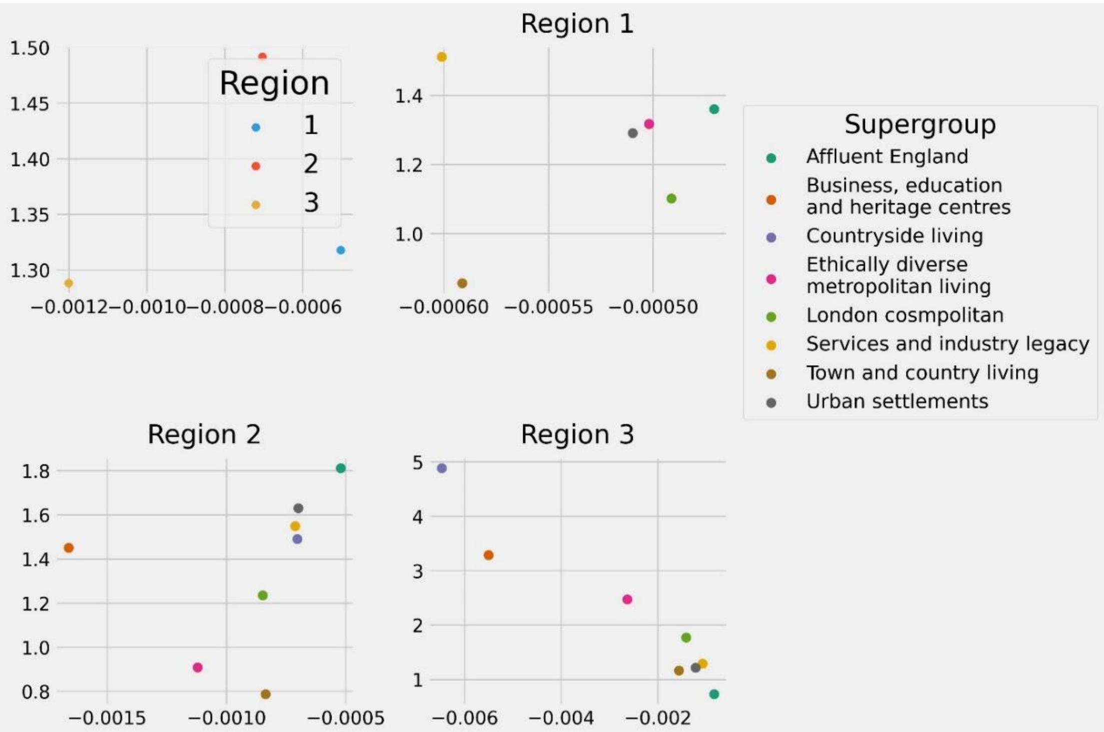

Figure 46 - Regional parameter values for the base and disaggregated spatial interaction model implementation for a single week

Table 14 - Regional parameter values for both the base model (system wide) and each supergroup

Supergroup	Region 1		Region 2		Region 3	
	Beta	Gamma	Beta	Gamma	Beta	Gamma
System wide	-0.000502	1.318	-0.000703	1.491	-0.001202	1.288
1	-0.000471	1.361	-0.000521	1.812	-0.000841	0.729
2	-0.000502	1.318	-0.001661	1.491	-0.005506	3.291
3	-0.000502	1.318	-0.000703	1.451	-0.006473	4.888
4	-0.000502	1.318	-0.001121	0.910	-0.002636	2.472
5	-0.000491	1.102	-0.000848	1.235	-0.01417	1.773
6	-0.000601	1.511	-0.000713	1.549	-0.001078	1.292
7	-0.000591	1.291	-0.000837	0.787	-0.001223	1.168
8	-0.000510	1.318	-0.000703	1.631	-0.001202	1.214

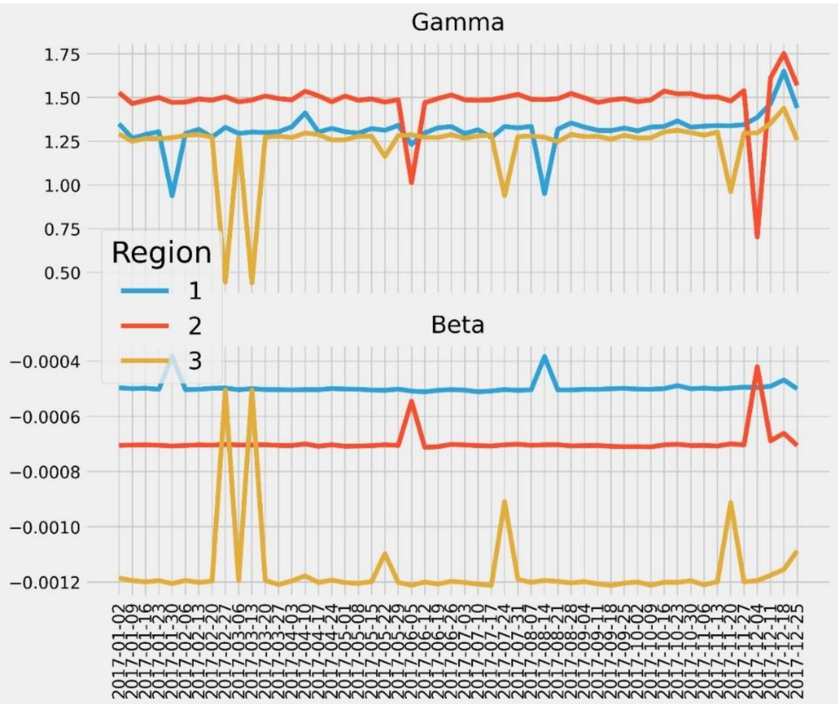

Figure 47 - Variation in system wide parameter values for all three regions across the year

Appendix B

Inverse Power Decay Regional Results

Table 15 - Performance metrics for the non-disaggregated base model for each region in comparison to the loyalty card flows

Metrics	Region 1		Region 2		Region 3	
	Loyalty card	Total Revenue	Loyalty card	Total Revenue	Loyalty card	Total Revenue
R^2	0.941	0.746	0.952	0.725	0.959	0.694
Pseudo R^2	0.907	N/A	0.918	N/A	0.950	N/A
RMSE	178.66	1009.12	146.04	1288.79	86.82	1129.11
SRMSE	0.590	3.332	0.577	5.092	0.694	9.026
ATD	1.066	1.168	1.019	1.105	1.020	1.087
MAE	73.59	412.21	52.89	488.58	19.62	359.91
AIC	1837994	N/A	3095182	N/A	2105707	N/A
SSI	0.199	0.149	0.176	0.114	0.113	0.062
CPC	0.878	0.569	0.895	0.486	0.922	0.389

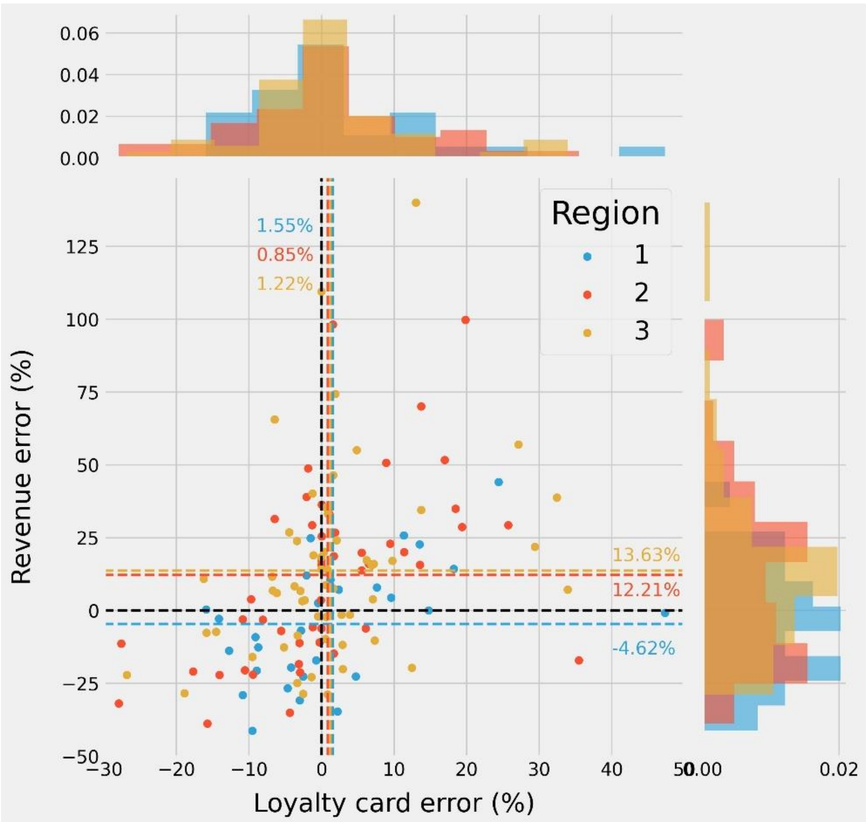

Figure 48 - Results from the non-disaggregated inverse power model application across all three regions in terms of the store level errors for both the loyalty card and total revenue in terms of the individual store percentage error

Table 16 - Performance metrics for the origin disaggregated model for each region in comparison to the loyalty card data (and the metrics from the base model) * Disaggregated metrics reproduced in

Metrics	**Region 1**		**Region 2**		**Region 3**	
	Loyalty card	Total Revenue	Loyalty card	Total Revenue	Loyalty card	Total Revenue
R^2	0.947 (0.941)	0.764 (0.746)	0.952 (0.952)	0.723 (0.725)	0.962 (0.959)	0.676 (0.694)
Pseudo R^2	*	N/A	*	N/A	*	N/A
RMSE	170.68 (178.66)	1026.06 (1009.12)	145.30 (146.04)	1291.84 (1288.79)	83.55 (84.82)	1160.15 (1129.11)
SRMSE	0.563 (0.590)	3.389 (3.332)	0.574 (0.577)	5.104 (5.092)	0.667 (0.694)	9.274 (9.026)
ATD	1.053 (1.066)	1.152 (1.168)	1.020 (1.019)	1.107 (1.105)	1.018 (1.020)	1.059 (1.087)
AIC	*	N/A	*	N/A	*	N/A
MAE	69.67 (73.59)	422.36 (412.21)	52.62 (52.89)	487.21 (488.58)	19.06 (19.62)	365.37 (359.91)
SSI	NaN (0.199)	0.151 (0.149)	0.176 (0.176)	0.114 (0.114)	0.113 (0.113)	0.061 (0.062)
CPC	0.884 (0.878)	0.567 (0.569)	0.896 (0.895)	0.486 (0.486)	0.924 (0.922)	0.386 (0.389)

Table 17 - Inverse power decay metrics for each region by supergroup

Metrics		**Region 1**		**Region 2**		**Region 3**	
	Supergroup	Loyalty card	Total Revenue	Loyalty card	Total Revenue	Loyalty card	Total Revenue
Pseudo R^2	1	0.872	N/A	0.891	N/A	0.921	N/A
	2	0	N/A	0.986	N/A	0.975	N/A
	3	0	N/A	0	N/A	0	N/A
	4	0	N/A	0.955	N/A	0.978	N/A
	5	0.935	N/A	0.925	N/A	0.937	N/A
	6	0.930	N/A	0.908	N/A	0.947	N/A
	7	0.953	N/A	0.931	N/A	0.961	N/A
	8	0.913	N/A	0.905	N/A	0.954	N/A
AIC	1	830927	N/A	482770	N/A	307066	N/A
	2	0	N/A	7157	N/A	10674	N/A
	3	0	N/A	0	N/A	0	N/A
	4	0	N/A	44796	N/A	66890	N/A
	5	366	N/A	370027	N/A	414393	N/A
	6	236478	N/A	910801	N/A	636950	N/A
	7	50043	N/A	147685	N/A	105407	N/A
	8	342704	N/A	1108995	N/A	466582	N/A

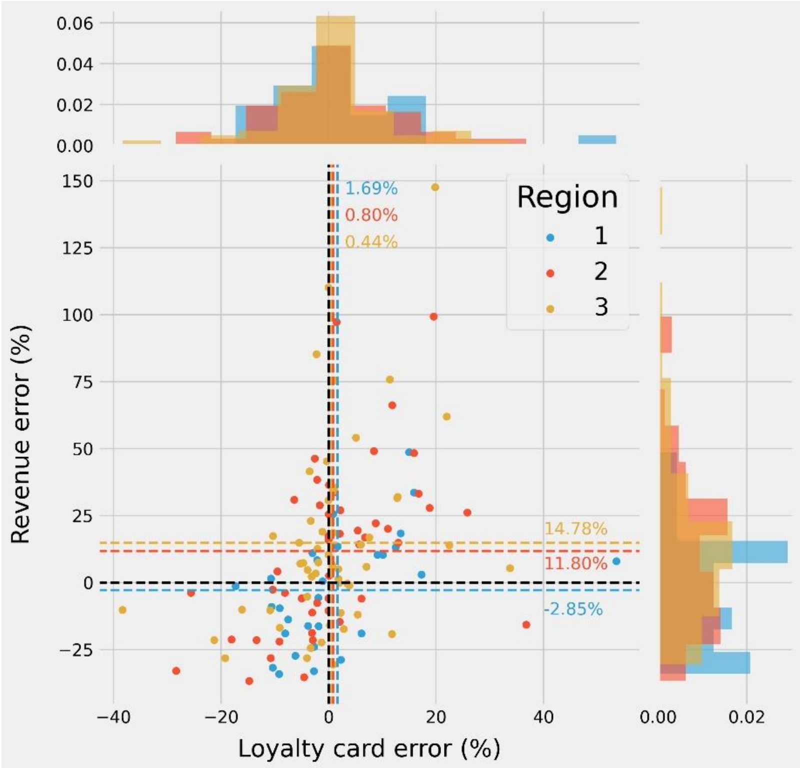

Figure 49 - Results from the origin disaggregated inverse power decay spatial interaction model across all three regions in terms of the individual and average store error for both the loyalty card and total revenue predictions

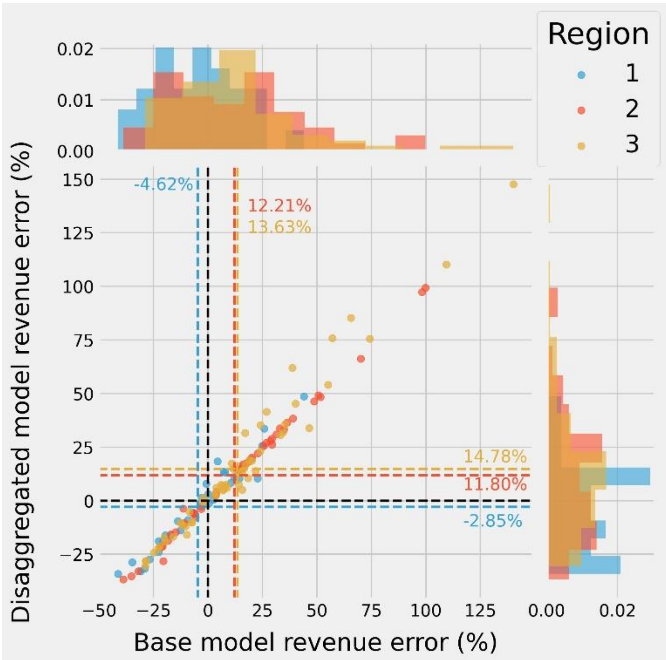

Figure 50 - System wide and origin disaggregated inverse power decay spatial interaction model errors compared across all three regions

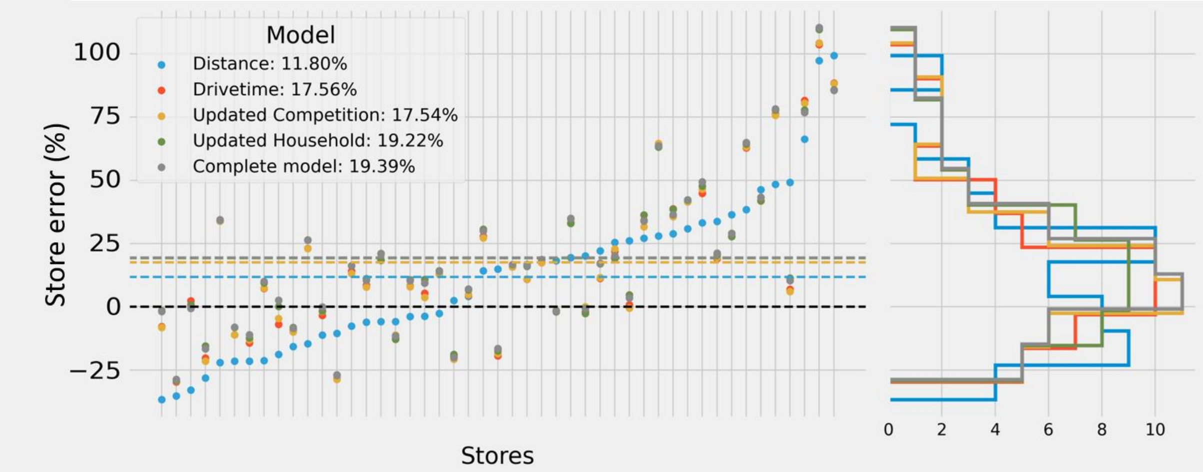

Figure 51 - Individual store error and the overall distribution in region 2 in response to additions of new data. 1) The origin disaggregated model as already presented, 2) The origin disaggregated model with drivetime data between origins and destinations, 3) The origin disaggregated model with drivetime and updated attractiveness values to account for competition, 4) The origin disaggregated model with drivetime and updated household count, 5) the origin disaggregated model with all new datasets integrated.

Table 18 - Pearson correlation statistic between store characteristics and inverse power decay model errors

Correlate		Region 1		Region 2			Region 3	
		Base model	Disaggregated model	Base model	Disaggregated model	Complete model	Base model	Disaggregated model
Store characteristics	Store size (sqft)	-0.1777	-0.0706	-0.0653	-0.0707	-0.1884	-0.1656	-0.0669
	Gross store size (sqft)	-0.1068	-0.0046	-0.1018	-0.1059	-0.2091	-0.1945	-0.0987
	Age of store (months)	-0.1056	-0.0804	-0.2841	-0.2795	-0.2928	-0.2035	-0.2336
Store revenue	Total store sales	-0.2449	-0.1343	-0.1673	-0.1593	-0.1799	-0.3508	-0.2922
	Total Grocery sales	-0.2854	-0.1859	-0.1857	-0.1748	-0.1802	-0.3843	-0.3283
	Total loyalty card grocery sales	-0.2347	-0.1475	-0.1165	-0.1039	-0.0408	-0.3733	-0.3568
	Percentage of sales of	-0.1415	-0.2328	-0.1111	-0.0877	0.0634	-0.2743	-0.3019

	grocery revenue							
	Total number of baskets	-0.1747	-0.0837	-0.1243	-0.1084	-0.1315	-0.2121	-0.1497
Surrounding area	Number of output areas	-0.0248	0.0767	-0.1638	-0.1733	-0.1252	0.0863	0.1292
	Average distance to output areas	-0.1287	-0.1568	0.2678	0.2661	0.3600	0.0547	0.0658

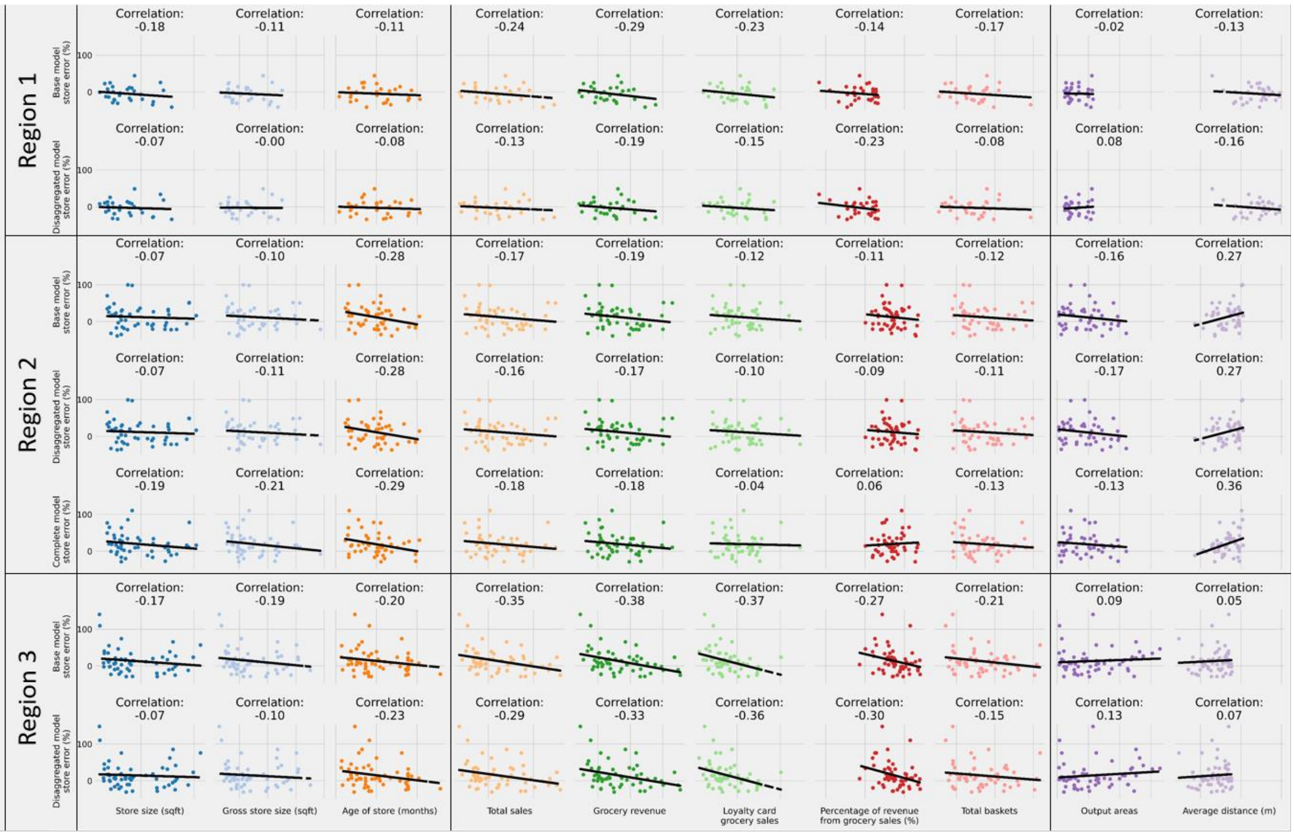

Figure 52 - Scatter plot of model errors and store level characteristics including the line of best fit and the Pearson correlation statistic

Figure 53 - Mean store total revenue errors for the base and origin disaggregated inverse power decay model for each region across the year

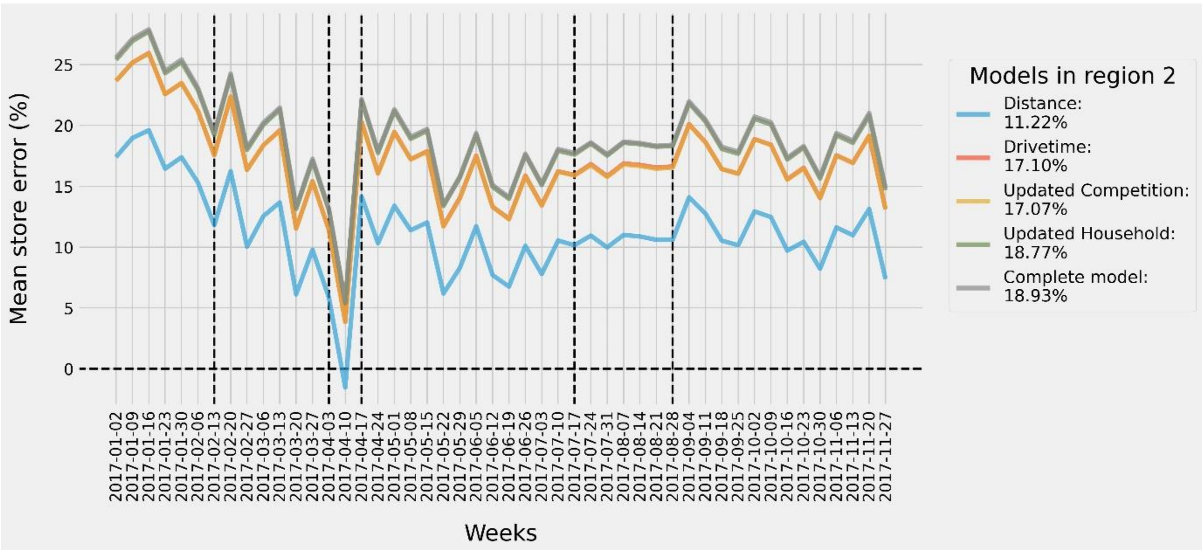

Figure 54 - Weekly average store errors in Region 2 across the different model implementations for the origin-disaggregated inverse power decay model

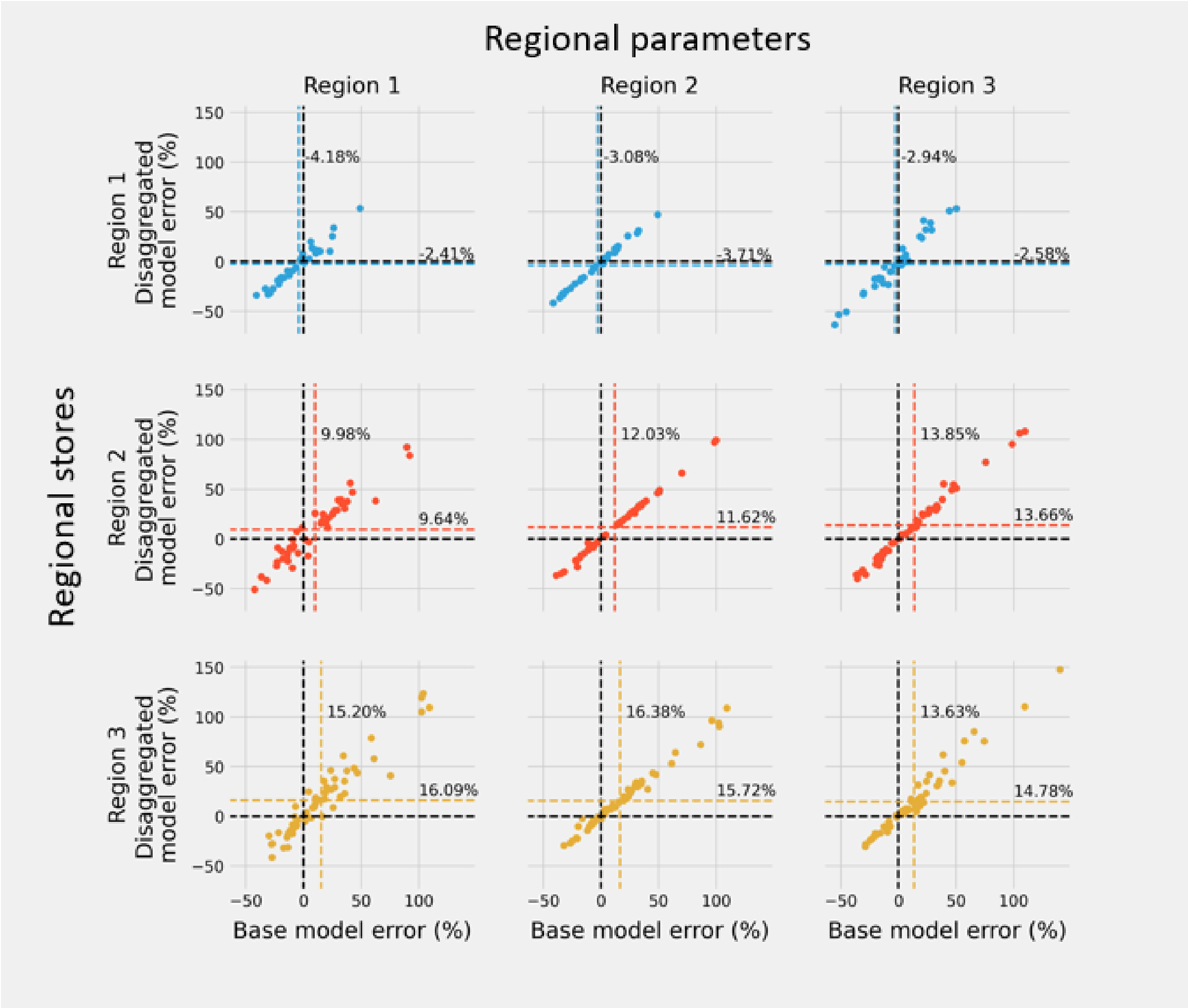

Figure 55 - Regional cross validation results with the regions trained parameters (column) against the regions stores and origins (rows) for the base model (x-axis) and the disaggregated model (y-axis) for the inverse power decay model

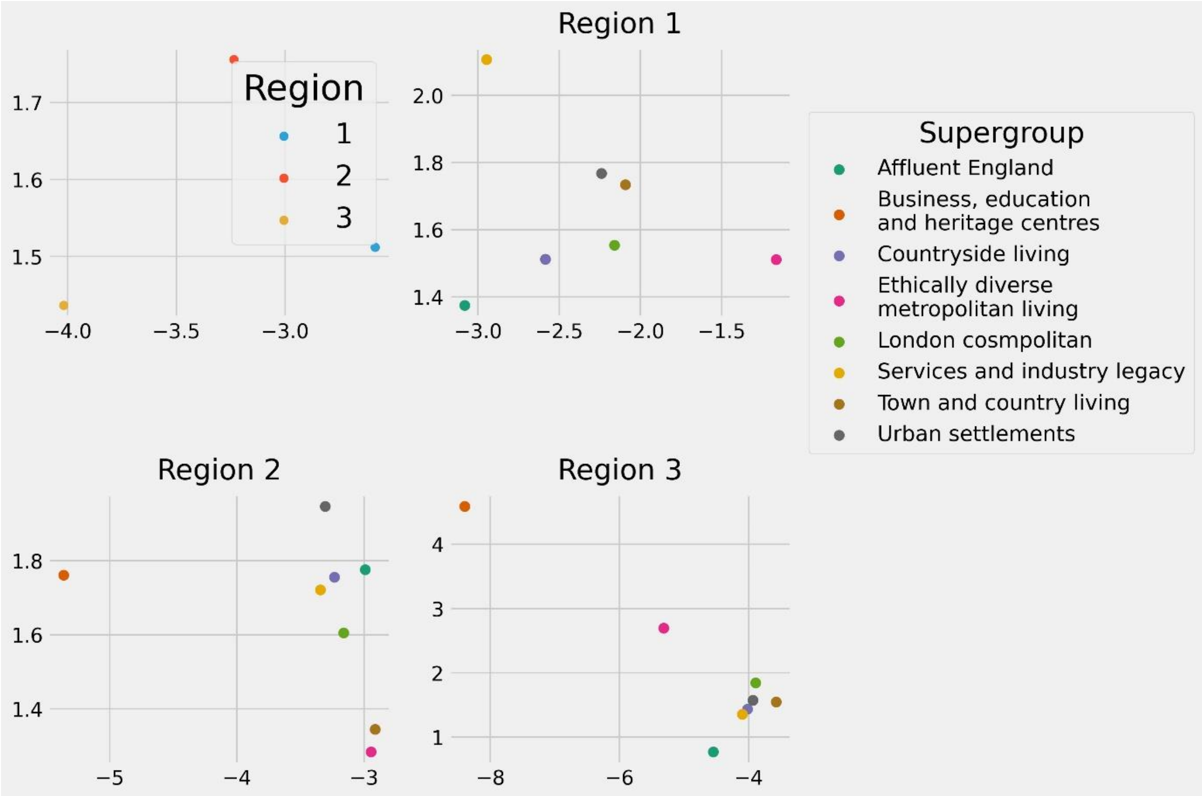

Figure 56 - Regional parameter values for the base and disaggregated spatial interaction model implementation for a single week for the inverse power decay model

Table 19 - Regional parameter values for both the base model (system wide) and each supergroup for the inverse power decay model

Supergroup	Region 1		Region 2		Region 3	
	Beta	Gamma	Beta	Gamma	Beta	Gamma
System wide	-2.586	1.512	-3.236	1.756	-4.019	1.436
1	-3.084	1.374	-2.992	1.777	-4.549	0.770
2	-2.586	1.512	-5.364	1.762	-8.401	4.594
3	-2.586	1.512	-3.236	1.756	-4.019	1.436
4	-1.165	1.511	-2.947	1.283	-5.318	2.696
5	-2.160	1.554	-3.161	1.606	-3.897	1.849
6	-2.948	2.106	-3.344	1.722	-4.102	1.356
7	-2.093	1.734	-2.914	1.345	-3.579	1.550
8	-2.241	1.767	-3.307	1.948	-3.935	1.576

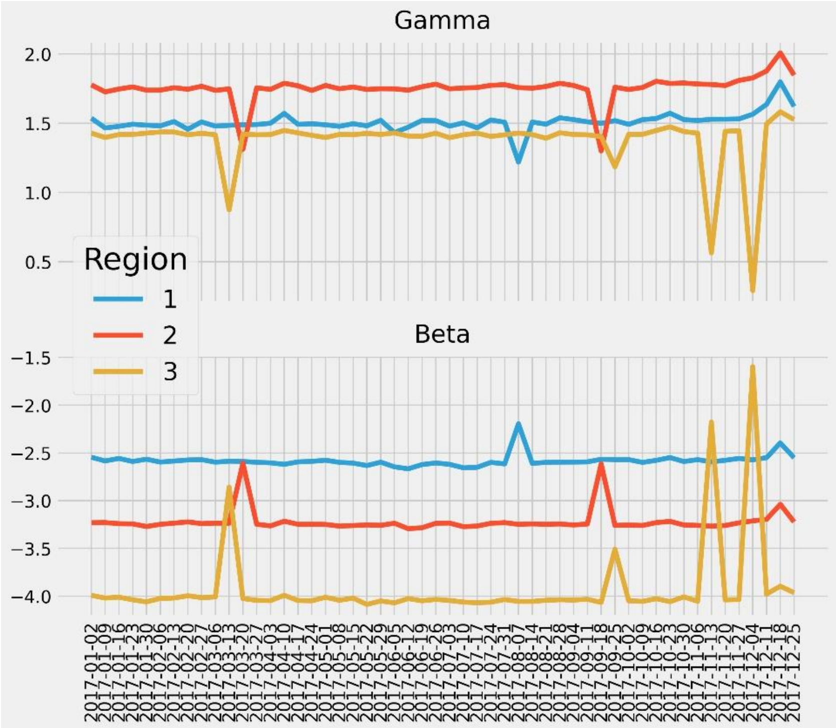

Figure 57 - Variation in system wide parameter values for all three regions across the year for the inverse decay model

Appendix C

Chapter 6 Inverse Power Decay Replication Results

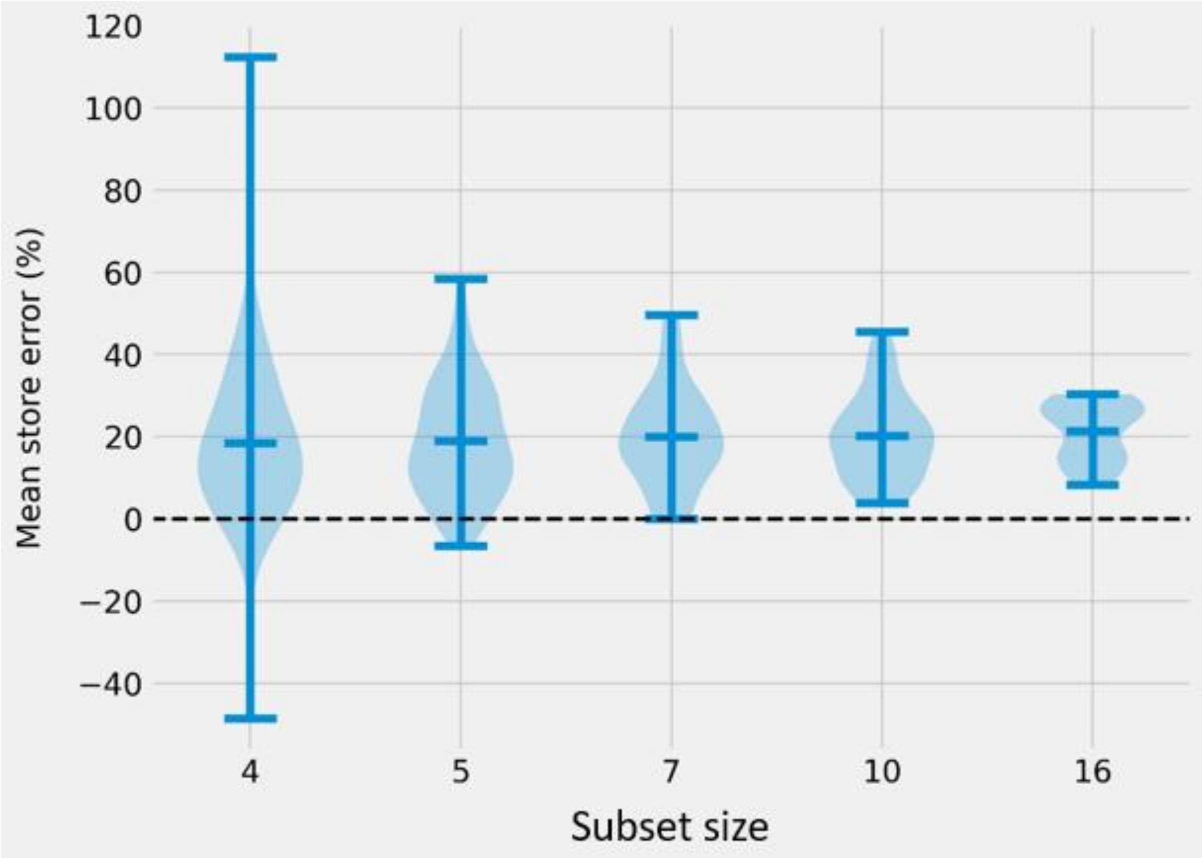

Figure 58 - A Violin plot showing the ranges of mean percentage error for each group for each subset size in region 2 for the inverse power distance decay model

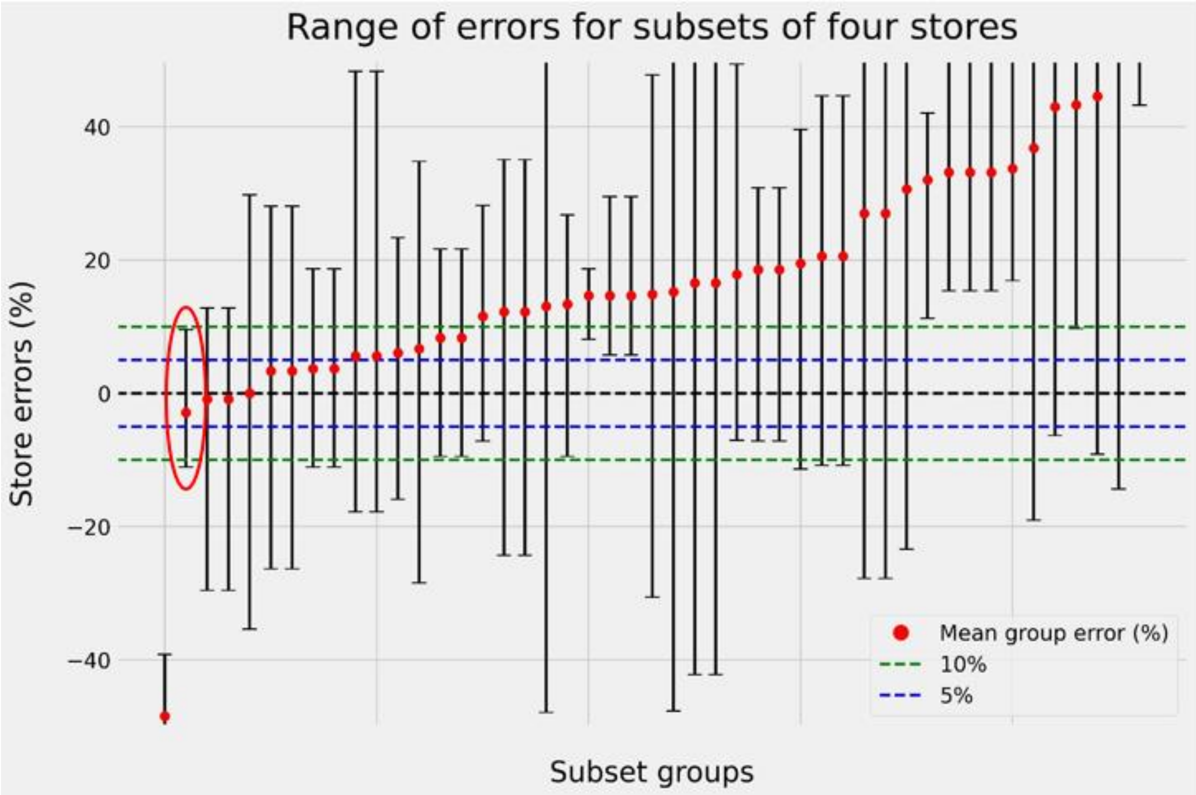

Figure 59 - Error range for groups of four stores including the mean error in comparison to the ranges suggested by Newing et al. (2015) for the inverse power decay model

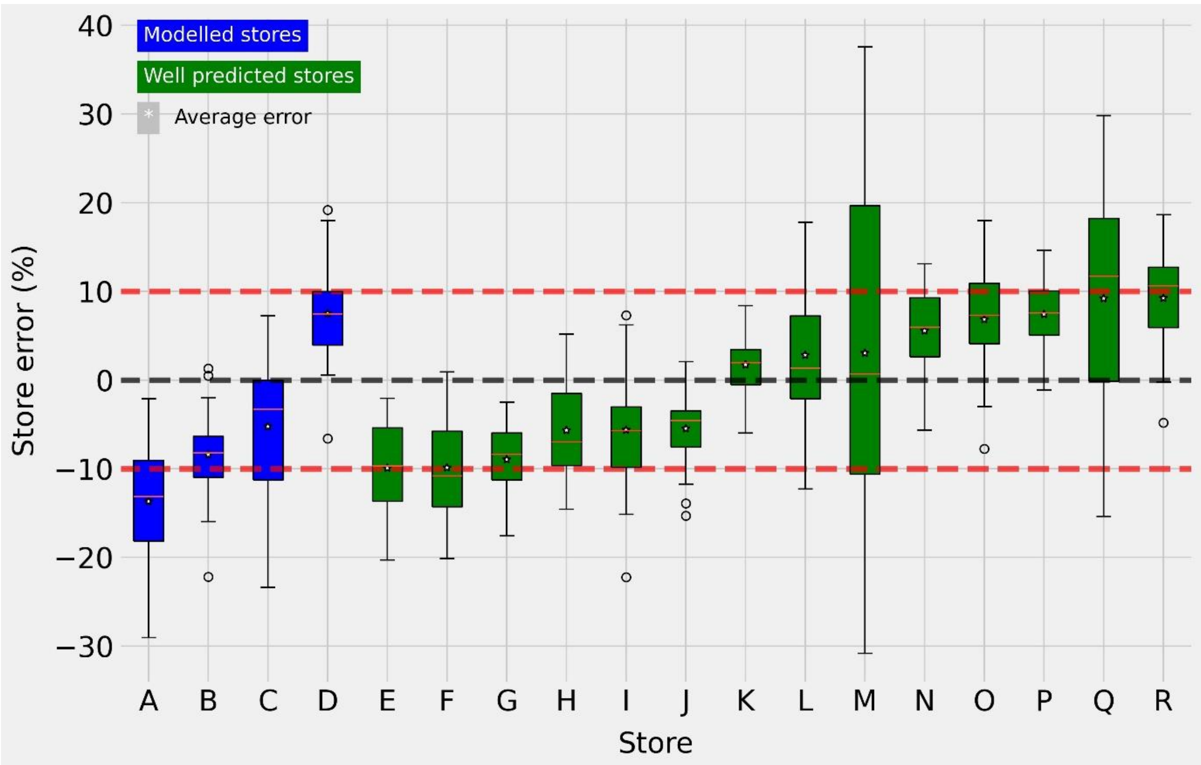

Figure 60 - The range of weekly store errors for each store in the well modelled store group of four stores and other "well-predicted" stores in the region for the inverse power decay model

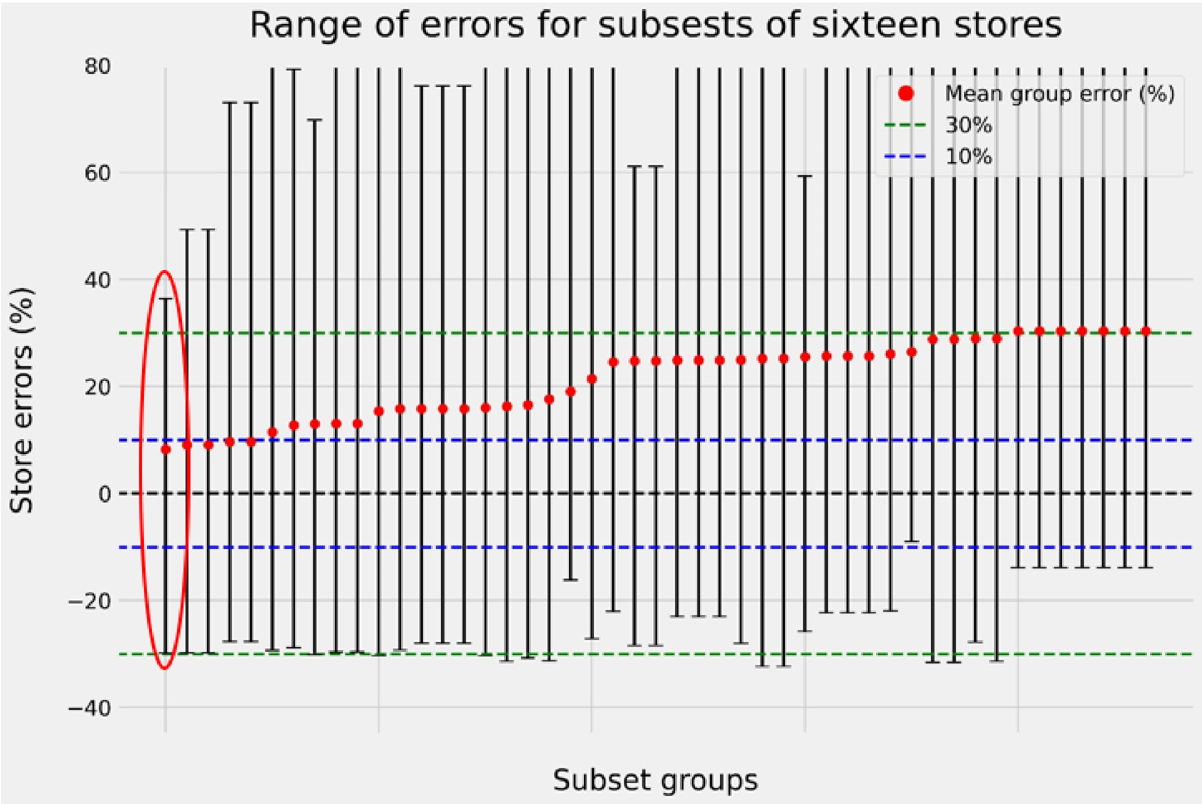

Figure 61 - Range of errors for groups of 16 stores along with the groups mean error. The horizontal bars indicate the error performance achieved by Waddington et al. (2019)

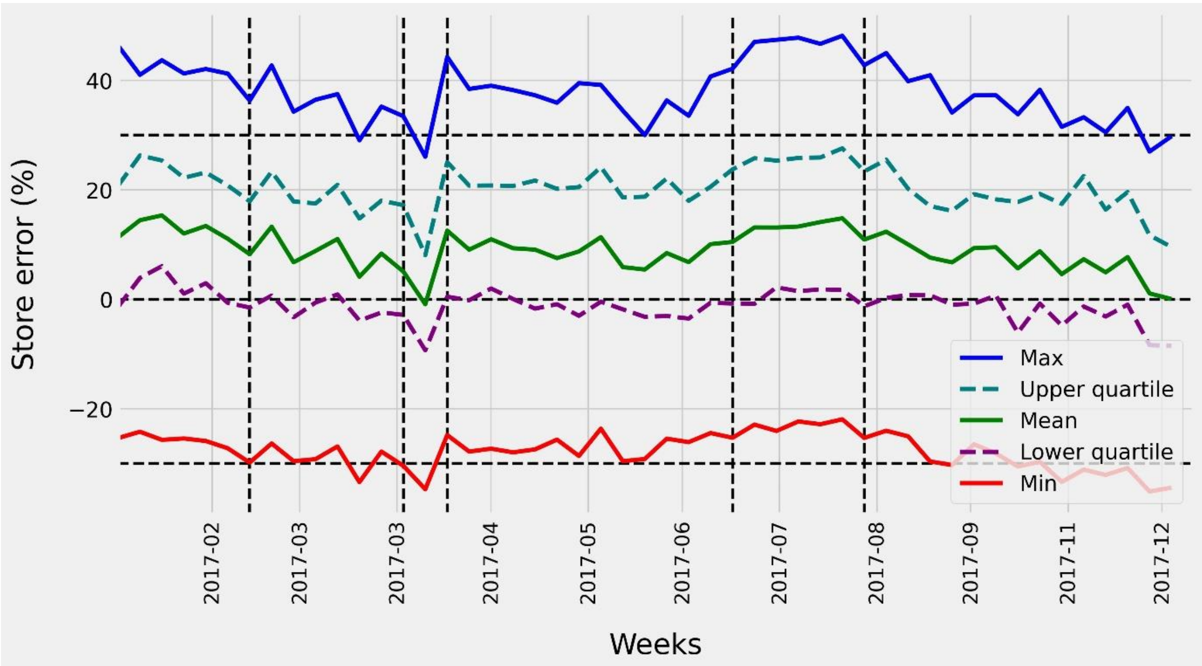

Figure 62 - Sixteen store subset well modelled group performance as applied over the entire year, including the maximum, upper quartile, mean, lower quartile and minimum store error from within the group for the inverse power decay model

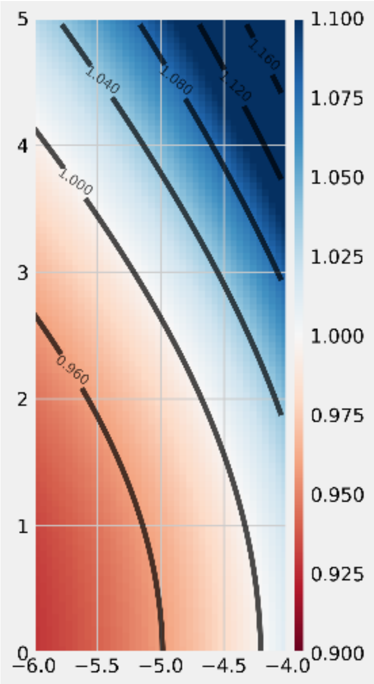

Figure 63 - Average Trip Distance (ATD) for a grid search of parameter values trained on the whole region for the inverse power decay base model

Table 20 – Inverse Power Decay parameter pairs derived from the grid search producing total revenue estimates close to ATD = 1

Pair	Attractiveness (γ)	Distance decay (β)
1	4.10	-5.97
2	3.49	-5.53
3	2.73	-5.05
4	1.64	-4.53
5	0.23	-4.22
Poisson Regression Calibrated	1.67	-3.95

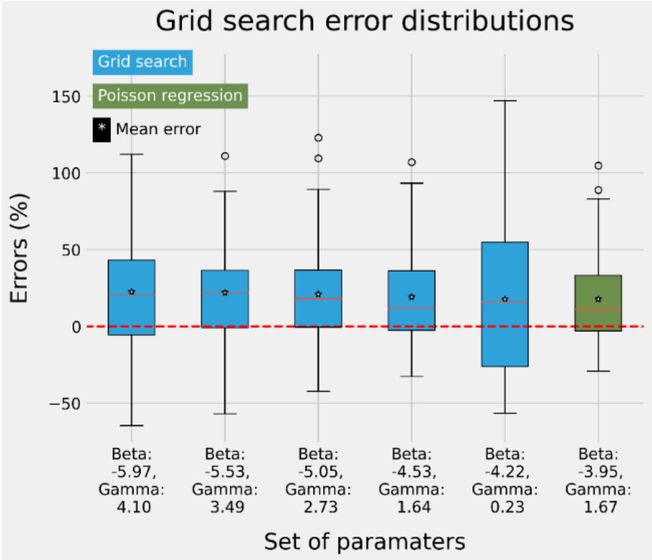

Figure 64 - Error ranges for each set of parameter pairs derived from the grid search for the inverse power decay base model with parameters that most closely generate revenue predictions with an ATD value equal to 1

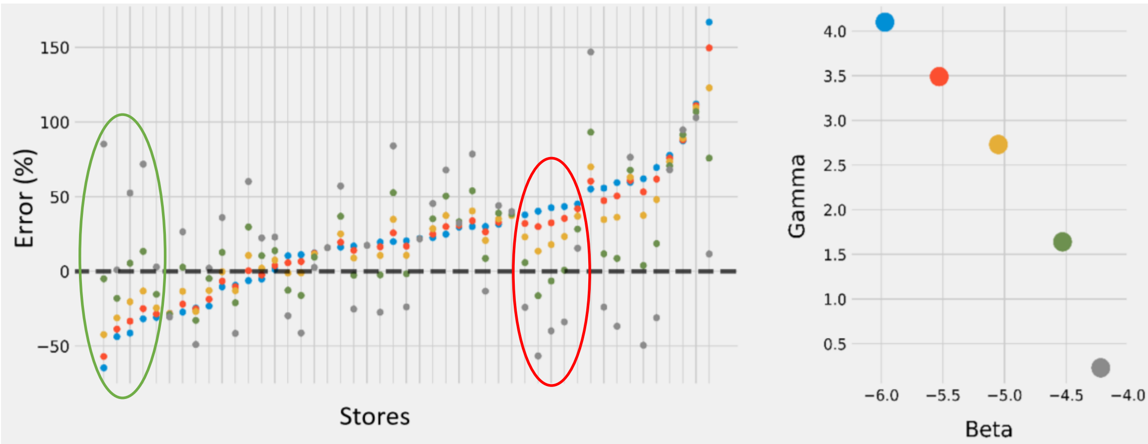

Figure 65 - Individual stores changes in error in response to parameter pairings

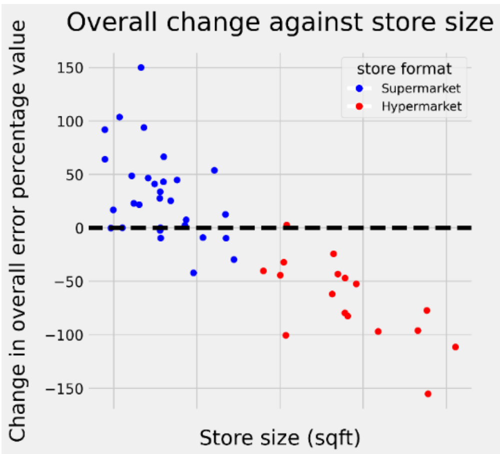

Figure 66 - Absolute change in percentage error in response to moving from parameter pair 1 to parameter pair 5 from the grid search model

Appendix D

Limitations Discussion

The results from Chapters 4, 5, 6, and 7 show that a production constrained spatial interaction model with its current form and data is not able to achieve the suggested performance in the previous literature when applied to a regional scale (Mendes & Themido, 2004; Newing, et al., 2015; Waddington, et al., 2019). This is despite the application of the model to several regional datasets across the UK for several weeks, the analysis of smaller groups of stores within a region, the application of an iterative calibration method and the implementation of a form of the competing destinations model. As can be seen from the results presented so far in this thesis, not even a majority of the individual store errors fall within a 15% error band and the variation in underlying errors from beyond 100% overprediction to below 50% underprediction means that these models are unlikely to be used as a reliable source of store revenue estimation in practice. In reaching this conclusion however it is important to acknowledge the limitations of the current research and how they may impact the performance of the model implementations. This includes the accurate estimation of total revenue available to spent at large grocery retail stores, the estimation of store attractiveness, distance and travel time calculations between origins and destinations, the calibration of the model parameters and the influence of behavioral change. While each of these are likely to have had an impact on the final outcome, as will be discussed below, it is nevertheless expected that the adjustment of the model to account for these factors is unlikely to significantly alter the conclusions that are drawn from the results so far.

Revenue Estimation

One limitation with the current model implementation, and one that is difficult to overcome within data regulation constraints, is the estimation of total revenue available to spend exclusively at large grocery stores. The presence of this limitation is primarily driven by the focus of this thesis of exclusively modelling large format grocery stores in the UK. This decision was taken in light of the results of the city level implementation, presented in chapter 4, and the issues of modelling convenience store revenue that were highlighted in previous research (Waddington, et al., 2019). Thus, removing these smaller store formats for the modelling implementation meant that total revenue available to spend had to represent only the revenue that would be spent at large format stores, not the total amount of revenue on all grocery spend. This created several potential issues in the modelling implementation including estimating the number of households, accurate disaggregation of potential revenue sources to reflect true grocery spend, estimation of non-residential expenditure and estimation of large format grocery revenue spend.

In this thesis, as in most application of the spatial interaction model to grocery retailing (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021), origins are determined as output areas in the UK. This is the smallest scale of geography that is available for which there are reliable estimates of the number of households within each area, alongside related socioeconomic data (Newing, et al., 2015). The latter of which allows for the classification of output areas into different socioeconomic groups (Gale, et al., 2016), as defined by their underlying demographic and economic characteristics. The estimate of revenue available within each output area then follows the formula:

$$O_i^{kt} = e^{kt} n_i^{k2011} \quad \textit{Eq. 75}$$

Where O_i^{kt} is a measure of total expenditure available in origin i by household type k during the year t, e^{kt} is a measure of the average weekly grocery spend for a household of type k in year t, while n_i^{k2011} is a measure of the number of households within origin i of type k from 2011. In this, the number of households in each output area are obtained from the 2011 census, explaining the 2011 superscript, while the average expenditure on groceries per household by socioeconomic group is obtained from the Living Costs and Food Survey for each year. This implementation is therefore consistent with the way in which previous papers have calculated the estimated revenue available (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021), and leads to reliable and consistent estimates.

However, there are four potential issues with this calculation that could affect the modelling results. The first issue is that the number of households in each output area is estimated using data from the 2011 census, while the model application is developed using loyalty card and total store revenue data from 2017. This creates the potential limitation of a potential difference between the actual household numbers in an output area in 2017 as compared to that of the 2011 census, thus affecting the estimate of the total amount of revenue available. This limitation was addressed in the final model implementation presented in Chapter 5 where the number of households in 2017 was estimated using population change figures for the relevant years. While this calculation was not expected to represent the true change in the number of households across the years, due to differences in demographic trends, population changes and housing construction, the effect of this for each individual store was small compared to the underlying model performance. In this case, while the change influenced the revenue estimated for some stores, it had little effect on others, and the overall effect was smaller than adding in drivetime metrics. Furthermore more accurate data in terms of the actual number of households was unavailable. Thus, due to the relatively small impact on the modelling performance and the expected accuracy of the adjusted calculation, this factor was deemed to be unlikely significantly affecting the modelling performance.

Related to this then was the estimate for the revenue available to be spent on groceries per household. This is because the estimates that are used are only disaggregated across eight different output area classification supergroups. While this allows for more accurate attribution of expected revenue from each output area according to their sociodemographic group, as opposed to the application of a single flat value, it could be argued that this disaggregation does not go far enough to reflect the true variance in household expenditure across the country. To this extent, data from the ONS goes further by developing tables that represent both variation in household expenditure across regions and by output area classifications, where there are 27 different groups as opposed to eight (ONS, 2022). In 2019, while the variation between regions of average household expenditure (£58.80 - £71.40) was less than the variation in expenditure across output area supergroups (£45.80 - £68.90), the variation across output area classification groups was greater (£35.70 - £75.30). This suggests that the use of output area classification (OAC) group data could produce more reliable estimates of expenditure per output area depending on how they are distributed across the country. Further research could therefore examine the influence of these estimates relative to the OAC supergroup data in terms of the modelling performance. However, the use of the OAC supergroup data was in line with previous model implementations across the country (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021), and exploratory attempts at implementation suggested a worsening of the estimation errors rather than an improvement. It is therefore expected that the inclusion of the more disaggregated data is unlikely to significantly affect the underlying issue of large variance in store errors and hence lead to the alteration of model performance.

Beyond this however is the estimation and integration of revenue from non-residential sources and how that may have influenced individual stores revenue estimations. For this thesis, the only data that was able to be accessed and evaluated was the residential population in output areas which were used to estimate the total revenue available. However previous research has identified the influence of non-residential demand on grocery retailing revenue (Newing, et al., 2013; Newing, et al., 2013; Newing, et al., 2014; Birkin, et al., 2017; Waddington, et al., 2018) and has attempted to resolve this issue by integrating new demand layers into the model to account for this (Newing, et al., 2015; Waddington, et al., 2019). This research showed that integrating non-residential demand layers, in the form of tourist and daytime population, improved overall model performance. In this regard, the influence of tourist revenue can clearly be identified in the yearly analysis of region 1 as presented in chapter 5 where there is considerable variation in model performance throughout the year. This reason however is why the analysis was performed on multiple regions so as to limit the potential effects of seasonal demand on model performance, including the focus in chapter 6 and 7 on region 2 which showed relatively little influence of seasonal demand on modelling outcomes.

Furthermore, the focus on the thesis was on large format stores in order to reduce the potential influence of daytime populations such as workplace and student demands (Waddington, et al., 2019). Therefore limiting the potential influence of these factors. Nevertheless, while there is expected to be some effect of including these populations in modelling outcomes, based on previous research for large format stores the scale of the correct is unlikely to resolve the issue of store error variance (Waddington, et al., 2019). This is highlighted by the evidence of consistent overprediction in region 2 and 3 in all three chapters. Thus, while not including these demand layers is likely to influence model performance, previous research suggests that the scale of the correction is unlikely to resolve the scale of the errors seen in this model implementation in this thesis.

Finally, in implementing the regional scale models and all subsequent implementations, the focus is on modelling revenue for large format stores. To this end the estimates from the Living Costs and Food Survey cover the expenditure per household per week on all grocery shopping, which includes shopping across all store formats. Thus, if the total value from the Living Costs and Food Survey is used per household then this would be likely to lead to consistent overprediction in terms of the total revenue to be spent only at large format grocery stores. To adjust for this therefore, data from our partner organization on total grocery revenue across all stores in the UK is used to calculate the percentage of revenue per week that is spent only at large stores. This percentage value is then used to scale down the estimates available revenue from each origin so as to only estimate the total expenditure at large format stores. A limitation of this however is that it is assumed that the percentage spent at large store for our partner organization is representative of the total amount spent across all large stores across the UK an in each region. Thus, this may potentially introduce bias into the amount of revenue available to spend. However, since the partner organization is a national grocery retailer with considerable investments in both large and small format stores, it can safely be assumed the percentage of revenue spend across format is representative of the total grocery market. Nevertheless, this could be a potential limitation. Secondly, this percentage adjustment is applied uniformly across all output areas in each region. This is likely to be unrealistic due to different levels of access and provision of different store formats across the country, namely with a division between urban and rural store distributions. To resolve this an index may be created to account for this variation, but without data relating to the decline of influence over distance, particularly for residential populations, then this is likely to be inaccurately estimated and applied with the data that is currently available. Thus, the implementation of a single scaling factor was taken as the best alternative due to both the ease of implementation and the argument that this is likely to be as accurate as a more complex index. Nevertheless, future research may consider examining the influence of these two factors and how the relate to model performance and hence whether they are a limitation in the current research application.

Store Attractiveness

Another potential limitation of the current model implementation is the estimation of store attractiveness. In this thesis store attractiveness has been measured by store size in square feet, in line with previous application of spatial interaction models in grocery retailing (Newing, et al., 2015; Waddington, et al., 2019; Beckers, et al., 2021). The data for this comes from our funding partner and includes data on store location, net area (internal floorspace) and gross area (internal and external floorspace), along with the opening and closing dates for grocery stores in the UK. An issue however is that throughout this dataset that are variation in the sizes attributed to stores. In this data there are no stores smaller than 1,000sqft while some also appear to have their size rounded to the nearest 1,00sqft. These differences may be expected to influence the model performance if the values form the data do not reflect the true store size and hence the underlying store attractiveness. However, small errors, including potential rounding of store size estimates, are unlikely to exert considerable influence on model performance, especially when rounding errors are small relative to the overall size of the stores. Furthermore, the dataset covers the entirety of the UK and includes stores that have been closed, are currently trading or stores that have changed brands. To the knowledge of this author there are no open datasets that could substitute for this dataset. The nearest similar dataset that is openly available would be the Geolytix data (Geolytix, 2022), for which the openly available version only provides store size information in four discrete bands. This would therefore be inappropriate for this application as actual store size is needed. Thus, the current data is expected to be sufficient for the current modelling application.

Secondly, this data only provides information on overall store size, as opposed to the square footage in a store that is dedicated to grocery retailing. While this may not influence the performance of the model for supermarket stores, that mostly focus on grocery retailing, this could be an issue for estimating the true attractiveness of hypermarket stores. This is because the larger store formats are likely to have floorspace dedicated to non-grocery revenue items such as household items, clothes or services such as banking or cafes (Fornari, et al., 2020). This may therefore affect the modelling performance by skewing the revenue assigned to large stores because of their disproportionately large footplate. On discussion with the authors of Hood et al. (2016), it was suggested that using the floorspace dedicated to grocery revenue improve the overall spatial interaction model performance, thus highlighting a potential limitation of our model implementation. However, when examining the relationship between store size and modelling errors, there was no clear relationship that could be identified in any region or model specification as seen in Chapter 5. Thus, this suggests that any influence as a result of the discrepancy between total store floorspace and that dedicated to grocery sales may not be large enough, or consistent enough, to show up in any relationship with model performance. Nevertheless, it is acknowledged that this could be a limitation but that there are not openly available datasets that could be used to compensate for this factor. Thus, this limited any exploration of the potential influence but may be something that future research could explore.

The final issue with the use of store size is that this was used as the sole measure of store attractiveness. It has been previously acknowledged that other influences are likely to affect how attractive a store may be to consumers (Newing, et al., 2020). This could include factors such as the distance to the street, store accessibility, store frontage or other services locating within the store such as cafes or baking services (Birkin, et al., 2017). The potential influence of this could be seen in chapter 5 where there was a correlation between errors and the age of the store, and conversations with our funding partner indicated that the inclusion of other store attributes led to improvements into their implementation of a gravity model formulation. However, lack of consistent store attribute data has meant that the inclusion and evaluation of these factors was limited. Furthermore, the performance gains seen from their implementation were only small relative to the overall model performance meaning that they were not large enough to justify the resources required to collect the extra data on a large enough scale to implement the model consistently across the UK. Thus, while the inclusion of such factors may improve modelling performance, the cost to collect and analyse them is expected to outweigh any potential benefit. IT may be suggested however that with resources such as google maps, google street view and OSM, a model may be trained to collect to attractiveness information that may be fed into future implementations of the model (Wood & Browne, 2007). Thus, while this is acknowledged as a limitation of the current research, future research may explore this opportunity in the future to examine what effect it has on modelling performance.

Distance and Travel Time Calculation

Further modelling considerations that are likely to affect modelling performance is the implementation and calculation of both distance and travel time. Firstly, in the majority of models that were implemented in chapter 5, before the updated data model implementation, distance between origins and destinations was given as-the-crow-flies. This distance measure however is unlikely to be representative of either the distance or time which consumers will actually travel to their destinations. For example, a store may be located where it is mostly accessible through a road running North/South where using as-the-crow-flies distance may attribute revenue from East/West. Thus, revenue may be underestimated from the North/South direction will be overestimated from the East/West direction as the distance used is not reflective of the true travel distance. This utilisation is therefore likely to lead to issues in model performance, especially in relation to individual stores. This issue is rectified in the final model implementation in chapter 5, alongside the subsequent model implementations in chapter 6 and 6, where it can be seen that using travel time increases the revenue attributed to stores from our partner organisation, indicating that they are more accessible by car than their competitors. However, while for some stores this reduced the scale of the absolute error, for others given the current model implementation it increased the absolute error. Thus, while using travel time represented the actual time that consumers would travel from their homes, it did not resolve the main issue of variance in store error. To evaluate this further however it would have been useful to examine the effect that travel time had on the two other regions to see the effect it had on model performance. However, time and resource limitations meant that this application could not be extended.

The second potential issue then was how travel time was estimated for consumers. For this, due to cost implications, the open source tool of the OSRM API was used to estimate the travel time between origins and destinations (Huber & Rust, 2016). The implementation of this API shows that the calculation of travel time from this source I based on average travel speeds across different types of roads as derived form open street map tags. While this may be a useful approximation of the actual travel time, the reality of travelling on streets is likely to vary according to traffic conditions and usage (Salonen & Toivonen, 2013). Therefore, tools such as Google Maps API may generate more accurate estimates of vehicle travel times from origins to destinations (Google, 2022). However, such resources are also likely to have the limitation of estimates of travel times changing according to the current traffic conditions, whereas OSRM estimates will remain consistent (Huber & Rust, 2016). Thus, there are potential trade-offs between the different sources of travel time estimation that may affect the actual value of travel time between origins and destinations. While resource limits meant that this could not be evaluated in this thesis, future research could build on this by exploring the influence of each different source on store revenue estimation and how it affects the variation for each store.

Model Calibration

Model calibration is also something that has been acknowledged in previous chapters that is likely to influence modelling outcomes. This issue was discussed in detail in section 4.1 where linear regression, Poisson Regression and maximum likelihood calibration methods were discussed. The main conclusion from this section was that the use of Poisson Regression was the most appropriate and practical method of calibration due to the representation of positive values, being able to deal consistent with zero flows, ease of implementation, the ability to handle modelling constraints and prior research papers suggesting that this method produced consistently accurate models

(Flowerdew & Aitkin, 1982; Tiefelsdorf & Boots, 1995). This method was therefore used consistently throughout the thesis apart from the use of an iterative calibration method used in section 6.5.

To this end, the initial test application of the model on a single city in section 4.3 was implemented using the Poisson Regression method from the statsmodel.api library in Python (Perktold, et al., 2022). When attempting to replicate this on a regional scale however with the same module, while for the first region the inverse power distance decay model could be calibrated, due to computational resources limits an exponential distance decay model could not. This is because the addition of the exponential distance decay introduced additional complexity in the model calculations which meant that the process would be killed by the internal system before calibration could be completed. Further resources to develop this implementation were not available due to the closed nature of the computing environment in which the data could be used. Thus, subsequent modelling implementations had to take advantage of the sparse matrices calculations that were part of the SpInt package developed by Taylor Oshan (Oshan, 2016). This allowed for reduced computational resources required for implementation, especially on the larger region 2 and 3, which also allowed for the calibration of the exponential decay form of the model. Comparing this implementation against the statsmodel.api results for the region showed parameter estimates for the base model that were consistent to three decimal places for both parameters and relatively small differences in the overall modelling errors. Therefore, the use of the SpInt module was not expected to considerably influence the modelling outcomes and it was used in all calibration implementations from Chapter 5 onwards due to it allowing for the calibration of the two further regions and the exponential distance decay form of the model.

Behavioural Changes

Beyond the potential limitations presented in this section so far, in terms of model formulation or data availability, impacting the modelling performance, changing behaviour could also be expected to influence the outcomes. To this end, whilst the model implementations are developed using data for 2017, by the early 2010s evidence had already began to appear that suggested lifestyle changes were already affecting shopping habits, particularly weekly grocery shopping. This included changes such as working more hours per week per household and travelling further and longer for work, necessitating changes away from the weekly regular shops at a single retailer that characterised behaviour up until the early 2000s (East, et al., 1994; Popkowski Leszczyc, et al., 2004), towards increased convenience shopping (Buckley, et al., 2007; Hallsworth, et al., 2010). This included more people shopping within a much small travel time, such as shift away from long distance car journeys to walking and cycling, an increase in the total proportion of shoppers who were shopping greater than three times a week, including regular top-up shops, and the combination of shopping trips with other purposes such as leisure or work (Elms, et al., 2010). This change in behaviour is thus seen as a departure from the assumptions of the gravity model that shopping behaviour is regular, for a defined single-purpose, that was characterised by large baskets and shopping by car (Waddington, et al., 2018). Grocery retailers in the UK responded to this change by focusing on building smaller format stores, with convenience stores growing at a greater rate than any other format from 2003 to 2012 (Hood, et al., 2015), for which it has been previously acknowledged that gravity models find it difficult to model these types of stores (Waddington, et al., 2019). This why the model implementations have focused exclusively on large format grocery retail stores, but it is still likely to affect the modelling performance due to the convenience shopping behaviour is not accounted for in the model implementation.

This influence is likely to affect the modelling performance due to the modelling assumption of single-purpose residential based grocery shopping. In this application, all revenue assigned to large grocery retail stores is derived from revenue estimates of expenditure, and based on distance and drivetimes, from residential locations. However, there is also evidence to suggest that shopping, particularly grocery shopping, often takes the form of multi-purpose and multi-destination shopping trips (Brown, 1992; Arentze, et al., 2005). This means that demand for grocery retailing is likely to come not only from residential locations by also from workplace, daytime and student populations (Birkin, et al., 2017; Waddington, et al., 2019). While this influence has attempted to be minimised in the thesis by focusing only on large format stores, this can still influence the assumption in the relationship of the spatial interaction model by reducing the potential expenditure from residential origins and distributing total demand across non-residential populations (Waddington, et al., 2019). Previous literature has accounted for this potential influence by adding in new demand layers (Waddington, et al., 2019), but it has been acknowledged that accounting for this influence is difficult due to estimation of both the potential revenue and parameters (Newing, et al., 2015). For this thesis, no data was available that could be used to illuminate behaviour in this regard, along with the inability to bring in new data sources into the system, such that non-residential demand layers could not be estimated. However, the relative changes in previous research were small enough that if integrated into this model they are unlikely to completely resolve the performance issues identified in this thesis (Waddington, et al., 2019). Thus, while this is a potential limitation and is likely to affect individual store performance, the expected improvement in modelling accuracy is unlikely to resolve the issues presented in the model so far.

Finally, there is also the potential influence that e-commerce in grocery retailing is likely to have on grocery shopping behaviour. To this extent the adoption of grocery-ecommerce has followed the general trend of broader e-commerce adoption over the last 20 years with increasing utilisation and value relative to the overall market share. This growth has been facilitated by the benefits of online shopping in that it reduces the search costs, provides convenient access to product and price information, enables quick and easy comparison, no restrictions on shopping hours and no associated travel costs for the consumer (Hamad & Schmitz, 2019). This therefore goes hand in hand with the general increase in demand for convenience by customers (Kirby-Hawkins, et al., 2019). While grocery e-commerce uptake has been slower than broader retail levels of adoption (Van Droogenbroeck & Van Hove, 2017), it is still likely to have a significant effect on the geography of grocery retailing. This is because the traditional distance decay relationship suggested in physical shopping is unlikely to hold, with suggestions of either innovation or diffusion theories of grocery e-commerce adoption suggested (Hood, et al., 2020). The effect of this on the models developed in this thesis has attempted to be minimised by removing all sales using e-commerce removed from both anonymised loyalty card data and total revenue. However, as the share of this revenue continues to grow for retailers, understanding the way in which it interacts with revenue predicting models will become more important (Beckers, et al., 2021).

Limitations Summary

Therefore there are limitations of the current implementation of the model namely in the form of revenue estimation, calculation of store attractiveness, distance and travel time calculations, model calibration and the influence of behaviour changes on the assumptions of the model. However, for the majority of cases these limitations have been dealt with in model adaptations, are likely to have limited effects on the overall modelling performance, or which data has not been made available to explore. This therefore presents opportunities for future research to incrementally improve the model implementation and potentially lead to more accurate results. This includes an exploration of how the model responds to the usage of square footage dedicated to grocery revenue, adjustments in estimation of total revenue estimated for large format stores, an evaluation of the different time and travel measures, and understanding how different transport options may affect travel to grocery stores. Future research could also build on the existing work presented in this thesis and those of Newing et al. (2015) and Waddington et al. (2019) to attempt to understand the effects of changing behaviour on model performance and how this may be reliably and consistently integrated into the gravity model specification. Such work could potentially include an identification of subsets of behaviour that still fit within the assumption of the gravity model, such as the identification of consumer groups that still do their regular weekly shop by car, and thus whether the gravity model still is able to model this behaviour. Further model adaptations could then be used to supplement this for other expected behaviour. Thus, these limitations leave remaining open questions to be explored in the future.

Appendix E

Origin-Destination Matrix

An important part of working with spatial interaction models is the creation of the origin-destination matrix that represents the potential flows from origin to destinations. In the case of grocery retailing, and in this thesis, it is assumed that these flows originate at consumers' homes and terminate at grocery stores in the UK. Whilst it is acknowledged that not all grocery retailing trips begin from the household, with many trips being combined with other activities such as work and leisure (Waddington, et al., 2019), this thesis only has access to data that pertains to consumers potential trips from their homes. This appendix will therefore discuss how the origin-destination matrices were created for this thesis for the first city area and the further three consumer regions in the UK. In particular, this appendix section discusses the methodology, data sources used, potential limitations and how this may affect modelling results.

For this thesis, as already discussed in previous chapters, we have access to anonymised loyalty card data from a single UK national grocery retailer which currently holds a significant share of the UK grocery retailing market and is responsible for thousands of stores nationwide. The dataset that we have access to from this retailer includes their loyalty card scheme data which has been running for over 20 years. Our access is limited to data from 2015-2020, terminating during the pandemic, but the focus is primarily on the loyalty card data for 2017. Whilst the underlying loyalty card data includes information about individual customers in terms of where they live, where they shop, what they purchase and when, access in this thesis is limited to an aggregation of this data. This is because data regulations restrict our access such that no individual could be identified during the process of performing this research.

Therefore, in this case, our partner organisation aggregated the underlying loyalty card data to the output area (OA) scale for England and Wales across every week in the time period, focusing solely on sales of grocery products. This dataset was then further limited by our partner organisation, due to the same data regulations, to remove any output areas where the number of households in a single week fell below a threshold of five so as to further ensure no individual could be identified. Therefore, data that was made available for the purpose of this research was in the form of an origin-destination matrix which represented the total amount of loyalty card sales per week per output area per store over the time period. This meant that each individual row in the dataset contained an output area code, representing the origin, a store name, representing the destination, and sales information in terms of the number of households and loyalty card revenue for an individual week.

In terms of using this data to then generate an origin-destination matrix that could be used to model grocery store revenue flows, access to this data within the workflow is represented in Figure 67 below. This figure is used to show the process that was used to create the full origin-destination matrix that allowed for the estimation of total grocery revenue for stores within each of the studied regions. This includes showing data sources that were provided by our partner organisation, presented in blue, data sets that come from external sources, highlighted in orange, and the processing stages that were undertaken to generate the matrix, as highlighted in green. This figure is also supported by Figure 68[17] below which visualises some of the key processing stages in generating the origin-destination matrix, including how these representations aided in the processing stages.

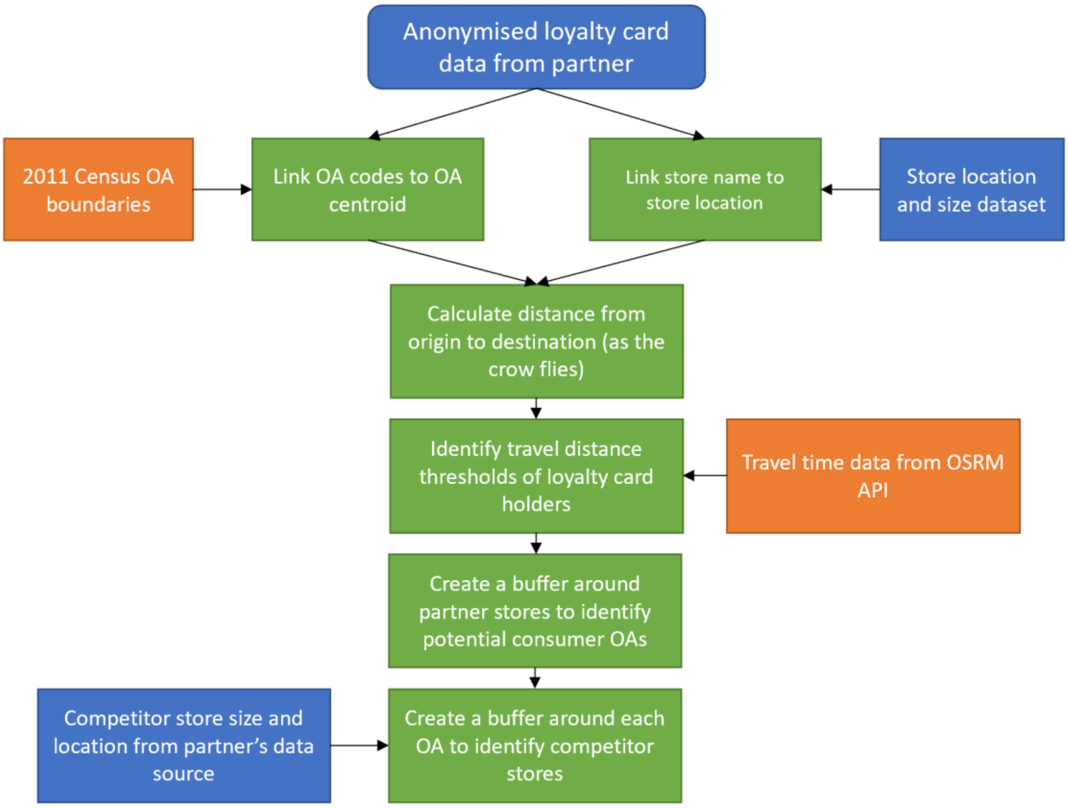

Figure 67 - Origin-Destination matrix generation workflow

As shown in Figure 67 the process for generating the origin-destination matrix begins with receiving the anonymised loyalty card data from our partner organisation. This dataset represents the flow of data that we know exists between origins and destinations, but which is limited by the form of aggregation that was undertaken due to data regulation, representing a biased sample of behaviour from only a single grocery retailer, and not showing the full flow of revenue from origins to

[17] Loyalty card data is not represented in this figure, rather generated data due to confidentiality reasons.

destinations (Birkin, et al., 2017; Rains & Longley, 2021). Firstly, the aggregation was performed by our partner organisation which could influence modelling outcomes in terms of the assignment of sales to grocery sales. This recognises that modern grocery retailers sell a variety of products, but that this is unlikely to significantly affect the results of this thesis because the aggregation of sales categories is performed by the retailer who follow standard industrial classification of products. In contrast, the use of a single retailer and the lack of data on the full flows between origins and destinations is likely to bias model estimations as behaviour of alternative consumer groups may differ from those that use loyalty cards and that shop at our partner organisation (Birkin, et al., 2017; Rains & Longley, 2021). However, this limitation is unlikely to be overcome by any researcher in the near future, due to confidentiality and competition across retailers.

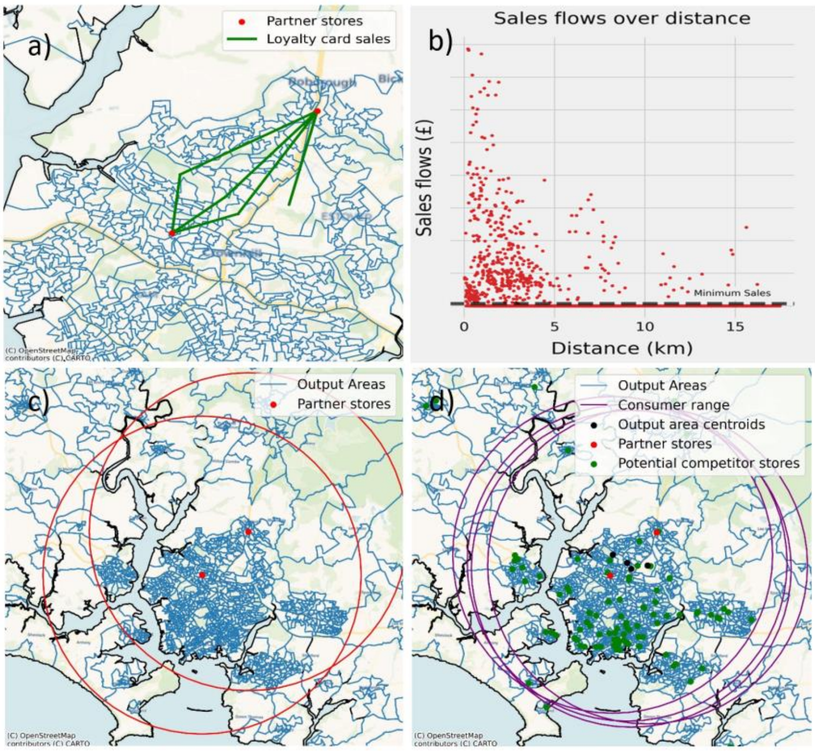

Figure 68 - Visual representation of the data processing steps that were undertaken to generate the origin-destination matrices. a) Linking loyalty card flows to output areas and stores to extract the distance loyalty card consumers travel, b) A representation of the flows of loyalty card sales over distance to identify a threshold distance consumers travel, c) the identification of potential consumer regions from each of our partner stores, d) the identification of all stores each output area could reasonably be expected to visit. *Representations not based on loyalty card data.

Once access to the underlying loyalty card origin-destination matrix was provided, the next stage was to then determine the distance that consumers were willing and able to travel to grocery retail outlets within the dataset. This is because the production constrained model used in analysis, as presented in Eq. 45, requires the distance between origin and destination, and we need to reasonably determine where consumers could be expected to travel from to complete their grocery shop. To identify these distances, the output area codes used in the loyalty card aggregation were linked to the 2011 output area census boundaries (ONS, 2020) from which centroids could be derived to generate a single point from which revenue could be reasonably expected to flow in the absence of individual house locations. This was achieved through a merging of the two datasets based on the output area code and then identifying the centroid from each output area geometry. Secondly, using a dataset provided by dunnhumby on grocery store locations in the UK, stores from the loyalty card data could be identified to a single point location. From these two datasets of output area centroids and stores, distance could be calculated between output areas in the loyalty card dataset to the individual stores. This allowed for the identification of flows from origins to destinations and their value, as represented in Figure 68a.

Once the distance between origins and destinations was calculated, this data could then be linked back to the underlying loyalty card origin to destination matrix. The next was to therefore identify the distances that loyalty card consumers were willing and able to travel for the their grocery shop, for the original city application and for all three regions in the UK. This process is represented in Figure 68b above which shows the distance decay relationship of sales flows over distance, and which is also highlighted in Figure 1 for the individual city application and Figure 8 for all three regions studied. These figures therefore represent the flow of grocery sales in the loyalty card data over distance, from which a clear and consistent decaying relationship can clearly be seen. These representations, alongside the distribution of sales over distance, could then be used to identify threshold distance over which it is reasonably to be expected that consumers would be willing to travel to perform their grocery shop. The need to calculate this threshold distance comes from two requirements of the spatial interaction model. The first is that it is unreasonable to expect consumers to consider visiting all grocery stores in a region, especially when that region is larger than most consumers would be willing to travel and that the data shows no consumers travelling beyond a given distance (Ellickson & Grieco, 2013). This therefore limits the potential destinations that a consumer would be expected to travel. Furthermore, if all origin to destination flows were considered in a region then this would considerably increase the computation required for both estimation of the parameter and assigning of revenue over flows, which could limit the implementation of the model. Therefore a distance threshold over which consumers are seen to travel is worth identifying.

In this however it is worth acknowledging limitations with this calculations and the estimation used in this thesis. Firstly, due to the data regulations requiring aggregation and the removal of output areas from the loyalty card origin-destination matrix with fewer than five households for each week, it is likely that some consumers would be willing to travel further than is seen in the aggregated dataset. This would be in line with the decaying relationships seen in the sales flow over distance figures and which could thus influence the estimation of distance would be willing to travel. Secondly, these distance calculations are based on loyalty card data from a single retailer, an acknowledgedly biased sample of consumers, beyond which it could be expected that consumers could be more or less willing to travel further. This then extends to non-loyalty card consumers who could be suggested to be less regular shoppers and thus potentially travelling further. Finally, these thresholds are calculated for all types and sizes of stores for our partner retailer, whereas consumers' willingness to travel will be highly dependent on the size and type of store. Indeed, this is an inherent assumption in the spatial interaction model that a larger store is more attractive than a small store and thus consumers are willing to travel further, therefore having a larger catchment area (Reilly, 1929; Newing, et al., 2015). Nevertheless, this issue of distance thresholds is resolved within the model itself due to the distance decay parameter influencing the distance that consumer are estimated to travel, but that the estimation of distance thresholds themselves could still influence model performance by identifying the range of consumers.

Whilst acknowledging these limitations and their potential influence on model performance, the next stage of the process of creating an origin-destination matrix was to use the distance threshold calculated to determine origins that could reasonably be expected to patronise our partner's stores. These are origins that fall within the distance threshold that surround each store and thus on the basis of loyalty card holders willingness to travel, could also be reasonably expected to travel the same distance. This process is represented in Figure 68c above which shows buffers being drawn around each partner store whereby each origin centroid that falls within each stores could be expected to potentially patronise that store. This then allows for an extension of the loyalty card origin-destination matrix by including other origins in the region whom for which there is no loyalty card data but which could be expected to be able to shop at our partner's stores. This means that the new origin-destination matrix represents potential flows from output areas in the region to stores from of our partner organisation, some of which we know have loyalty card data but others for which we have no data. This also contains the distance between each origin and destination in terms of as the crow-flies distance. In the analysis for Region 2 however this data is supplemented by travel-time distance that is generated from the OSRM API based on the centroid of each output area to each store within the origin-destination matrix.

This modified origin-destination matrix therefore now contains rows representing pairs of origins and destinations which grocery sales could be expected between for our partner organisation. In reality however it is expected that the output areas identified are also likely to visit stores from other competitors to undertake their grocery shop. The next, and final stage, therefore is to extend the origin-destination matrix to include potential pairs of origins to destinations from origins to stores from other brands that they would be reasonably expected to visit. This is achieved following a similar method to identifying the full set of origins but instead of creating a buffer around each store a buffer is created around each origin. The distance of this buffer is the same as that used around each store because it is the maximum distance that consumers are expected to travel to shop for groceries. This process is represented in Figure 68d above which shows the buffers being drawn around four output areas and stores being identified that lie within those buffered distances. Data for these stores and their locations come from a dataset provided by our partner organisation that includes information on store location, size, brand, fascia, opening data and closing data. This process therefore allows for the creation of a complete origin to destination matrix that includes pairs of output areas and stores that consumers are reasonably expected to potentially shop at based on the distance that consumers travel from loyalty card data.

From this complete origin to destination matrix for each study area, analysis can be performed based on a subset of the matrix and the complete matrix. As detailed in each respective analysis section, the first stage of analysis to estimate model parameters based on loyalty card data is based on the subset of origins and destinations that show actual and potential loyalty card flows from output areas to our retailers stores only. The second stage of the analysis, used to estimate total grocery store revenue, is then based on the full origin to destination matrix as described in this appendix section. For this analysis, the full origin to destination matrix is combined with other data sources such as the 2011 census data to get the number of households, the 2011 Output Area Classification (Gale, et al., 2016), and the Living Costs and Food Survey (ONS, 2021), to be able to estimate the total revenue flows between origins and destinations within the dataset. Full details on this estimation process is detailed in each respective chapter where a model is developed and applied, including Chapter 4, Chapter 5, Chapter 6 and Chapter 7, using the origin-destination matrix created.

Finally however, in addition to some of the limitations discussed above, it is also worth highlighting that some of the issues of model estimation within this thesis could likely be due to the fact that this origin to destination matrix is solely based on flows from consumers homes to stores. To this end it has been highlighted in previous research that expenditure and flows from non-residential sources, such as from tourism, workers and student populations, could contribute to a store's revenue generation especially in relation to grocery retailing (Newing, et al., 2015; Birkin, et al., 2017; Waddington, et al., 2019). Therefore future research could follow the above methodology to extend to the origin-destination matrix by including these non-residential sources of revenue, and potentially account for the influence of multi-purpose shopping. This would thus extend the origin-destination matrix to encompass flows to stores from a variety of different origins from which the potential catchment area is likely to differ to that from residential origins, especially across different formats of stores.