



Development of joint models for channel, store, and travel mode choice: Grocery shopping in London



Esra Suel*, John W. Polak

Centre for Transport Studies, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom

ARTICLE INFO

Article history:

Received 18 August 2016

Received in revised form 25 January 2017

Accepted 21 March 2017

Available online 5 April 2017

Keywords:

Online shopping

Multi-channel shopping

E-commerce

Store choice

Destination and mode choice

Discrete choice modelling

ABSTRACT

The nature of shopping activity is changing in response to innovation in retailing and the growth in online channels. There is a growing interest from transport researchers, policy makers, marketing and retail businesses in understanding the implications of this change. However, existing tools and techniques developed for analysing behaviour in traditional retail environments do not adequately represent emerging complexities resulting from digital innovation. In this paper, we advance existing destination and mode choice models by incorporating online channels in a unified framework. This is a critical extension to existing transport literature on destination choice which largely ignores online activity. Specifically, we develop discrete choice models using elemental store (including both online and in-store) alternatives for joint choice of channel, store, and travel mode. We demonstrate the use of a widely-accepted consumer panel dataset with minor modifications, for the first time in transport research, together with API based data mining tools that offer great potential for enrichment.

The analysis focuses on grocery shopping and uses consumer data collected from two selected boroughs in London; results from multinomial logit and nested logit estimations are reported. The extension presented here provides the tools to quantify the effects of increased online shopping on traditional store formats and travel patterns. Our results showed virtual alternatives currently offer an attractive substitute among early adopters for large basket shopping mostly for high income groups. This might suggest a significant reduction in shopping trips to hypermarkets often associated with large basket shopping potentially leading to store closures. Online deliveries mostly draw from driving trips and less so from walking and public transport trips. The present study also confirmed previous findings related to smaller stores and longer travel distances being associated with declining utilities, agglomeration and competition significantly influencing store choice.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The online retailing market has been growing steadily over the past decade and continued growth is expected. The share of online retailing in the UK, as a percentage of all sales, reached 12.5% in 2015 compared to only 0.3% in 1998 and 2.1% in 2002 (Office for National Statistics, 2015). In the USA e-commerce retail sales accounted for 7% of all sales in 2015, up from 2.8% in 2006 (Bucchioni et al., 2015). There is a growing interest from transport and town planning professionals, policy makers, as well as marketing and retail businesses in understanding how shopping behaviour is changing and potential

* Corresponding author.

E-mail addresses: esra.suel@imperial.ac.uk (E. Suel), j.polak@imperial.ac.uk (J.W. Polak).

implications for the future. These changes in the nature of retailing activity have implications for travel in general and more specifically for personal mobility. There is an extensive literature in transport research that looks at potential implications of increasing use of information and communication technologies (ICTs), as detailed in Section 2 of this paper. Further research, however, is needed for developing the modelling tools and identifying data requirements for studying choice behaviour in this increasingly complex retail environment. One area of particular interest is the analysis of choice behaviour involved in shopping when shoppers choose from a variety of in-store and online alternatives as well as delivery and travel options available to them. The modelling of this behaviour is valuable both from a research perspective due to the complexity of the choice process, and also from a business and policy perspective for providing insights into the potential implications of shifts in the retail market.

Existing models used by transport planners have very little representation of the complexity involved in multi-channel shopping. Shopping destination and travel mode choice models in transport research focus on spatial choice behaviour and exclude online alternatives. The literature on channel choice, on the other hand, has largely focused on modelling the choice between online and in-store alternatives ignoring the heterogeneity within each channel and linked choice dimensions resulting in trip generation. In this paper, we bridge this gap by developing discrete choice models for joint choice of channel, shopping destination (store), and travel mode. The retail sector in the UK offers an opportunity for studying channel choice due to its relatively developed online market ([Centre for Retail Research, 2016](#)). Still, finding empirical data sources for this line of work presents a challenge as travel surveys do not adequately capture online activity. Here, we also demonstrate the use of an existing and widely-accepted consumer panel data with a modest augmentation in a novel context to address this data gap. Collating and merging data from multiple additional sources for enrichment of panel data were also required.

The literature emphasizes the importance of acknowledging shopping itself is a heterogeneous class of activities when modelling behaviour ([Mokhtarian, 2004](#); [Visser and Lanzendorf, 2004](#); [Rotem-Mindali and Salomon, 2007](#); [Girard et al., 2003](#)). Shopping for daily groceries may involve different behavioural mechanisms with different mobility implications when compared to occasional shopping for white goods items. Changes in retailing are also likely to impinge on the problem of channel and mode choice very differently in different retail sectors (e.g., electronics, clothing, furniture). While it might be of interest in principle to investigate these differences, a comprehensive treatment is beyond the scope of this study. We limit our focus to grocery shopping occasions where a transaction occurs (i.e., excluding pre- and post-purchase trips). The reasons for this decision was fourfold. First, grocery shopping is the most common and frequent type of shopping, which makes it more relevant for travel implications and data collection less challenging. In Britain, food shopping was the most popular purpose accounting for half of all shopping trips and car was the main mode of transport for 78% of these journeys in 2014 ([DfT, 2016](#)). Average distances for food shopping trips were lower at 3.6 miles when compared to all shopping trips at 4.9 miles ([DfT, 2016](#)). Second, online grocery market is a very dynamic and fast growing segment at present and is expected to grow with an annual growth rate of more than 50% over the next five years in the UK ([Gladding, 2016](#)). Only within the past year Amazon Fresh expanded to the UK ([Butler, 2016](#)) and big players such as Google Express and Uber entered online grocery market ([Thielman, 2016](#); [Whipp, 2016](#)). Hence, potential implications on travel will be increasingly important. Third, shopping is increasingly seen as leisure activity introducing additional complexity when modelling the choice between online and in-store. Grocery shopping is among the most utilitarian with minimal (if any) recreational value. In our view, this allows to really get at the physical vs. virtual issue net of any of the complicating factors relating to recreational value of shopping. Fourth, pre-purchase (searching) and post-purchase (returns) stages are less relevant for additional trip generation reducing the complexity involved in modelling.

The remainder of this paper is organised as follows. In the next section, we present a brief review of previous work on online shopping and store choice behaviour. The third section describes the data used in this analysis, while the fourth section looks at methodology and model specification. The fifth section presents the results of model estimation. Finally section six summarises the findings of this research.

2. Previous work

Shopping has been subject to transport research as it has been one of the most popular reasons for travelling since the 1970s. In 2014, for instance, shopping was the most common trip purpose accounting for 19% of all trips in England ([DfT, 2015](#)). Shopping related travel is driven by the dynamics of the retail sector and hence highly influenced by its transformation. For example, average distances travelled for shopping trips have increased significantly in the UK over the past 20 years with out of town retail developments ([DfT, 2015](#)). Changes in retailing driven by ICT innovation is another important example of this, which have not really been extensively explored from a choice behaviour perspective. Transport researchers have been interested in understanding potential travel implications of increasing use of ICTs since the 1980s, with a focus on assessing its potential for reducing demand. At a conceptual level, potential travel impacts of increasing use of ICT have been classified into four types: substitution, complementarity, modification, and neutrality ([Salomon, 1985, 1986](#); [Mokhtarian, 2003](#); [Hjorthol, 2002](#); [Mokhtarian et al., 2006](#); [Mokhtarian, 2009](#)). More specifically, many studies in transport literature focus on quantifying the net effects of increased use of ICT on shopping journey frequency and miles travelled with a range of, sometimes contradictory findings as evidenced in comprehensive reviews by [Weltevreden \(2007\)](#), [Bhat et al. \(2003\)](#), [Rotem-Mindali and Weltevreden \(2013\)](#), and [Rotem-Mindali \(2014\)](#). The relationship between frequencies of in-store and

online shopping activity have also been studied extensively (Cao et al., 2012; Circella and Mokhtarian, 2010; Zhou and Wang, 2014; Farag, 2006; Farag et al., 2007; Dijst et al., 2008).

One other stream of the transport and marketing literature has focused on gaining insights to and developing a conceptual framework for the decision-making processes involved in channel (online vs. in-store) choice behaviour (Balasubramanian et al., 2005; Schoenbachler and Gordon, 2002; Salomon and Koppelman, 1988; Nagurney et al., 2001; Bosnjak et al., 2007; Dholakia et al., 2010; Couclelis, 2009; Crocco et al., 2013). Distinctive benefits (e.g. value of physical assessment, recreational aspects, time savings, lower prices, additional information) and costs (e.g. travel to store, delivery costs) associated with in-store and online shopping have been identified. The way in which channel attributes (e.g. prices, quality of products, in-store experience, delivery service quality, etc.) and personal preferences influence choice behaviour is also explored. Building on conceptual studies, empirical work in this area have focused on (i) adoption of online shopping and (ii) choice of channel for a specific shopping occasion. The focus of studies in the first group is to develop models of whether an individual adopts the channel (i.e., does any shopping online); numerous studies have looked at this behaviour as summarized by Chang et al. (2005) including Soopramanien and Robertson (2007), Teltzrow et al. (2007), Pauwels and Neslin (2011), Ha and Stoel (2009), Liu and Forsythe (2011), Bellman et al. (1999), Rhee et al. (2009), and Schröder and Zaharia (2008). Previous work that falls under the latter, on the other hand, is rather limited and has been largely separate from shopping destination and travel mode choice literature. These models of channel choice are mostly developed at the level of aggregated channel alternatives where the decision is between online versus in-store, hence the heterogeneity within each channel is not represented. Researchers face serious gaps in empirical data coverage as regional and national travel surveys typically collect quite limited information with which to establish how in-store and online shopping behaviour interact with one another. Most studies use one-off ad-hoc surveys (Hsiao, 2009; Cao et al., 2012; Chocarro et al., 2013; Mokhtarian and Tang, 2013; Crocco et al., 2013). Data are either sourced from stated-preference experiments where respondents are asked to choose between hypothetical online and in-store channels with simulated alternative attributes (e.g. website design, in-store atmosphere, prices, travel and delivery costs, delivery times) and contextual variables (e.g. time pressure, presence of other decision makers), or from revealed-preference surveys in which respondents are asked about their most recent purchase of a given product category and which shopping channel they used. Additional data on behavioural co-variables (e.g. online purchase frequencies, choice of channel in pre-purchase stages, experience with Internet) and socio-demographics are collected, and have been found to significantly correlate with channel choice. A novel approach was introduced by Chintagunta et al. (2012) who looked at channel choice for grocery shopping using scanner data from a single retail chain. He found that choice of channel is significantly influenced by what is being bought (e.g., total basket cost, presence of heavy/bulky or perishable items in the basket) in addition to situational variables (e.g., weekday/weekend, time of day, weather conditions, delivery costs, and in-store promotions). A significant limitation, however, was that the sample observations were restricted to shopping visits to the single retailer chain where the scanner data was sourced from.

This paper aims to contribute to the stream of literature that aims at modelling the channel choice behaviour for each shopping occasion. The objective is to represent channel choice as an additional dimension within the traditional shopping destination and travel mode choice framework. Modelling choice at the level of elemental store alternatives allows for representing the heterogeneity within each channel. Incorporating the travel mode dimension in a joint choice model framework helps to better understand how channel choice behaviour will influence personal mobility. There are, however, substantial challenges relating to the treatment of large choice sets. In the context of store location choice, it is possible to model the choice of individual stores (i.e. elemental units of attraction) at the most disaggregate level (Recker and Schuler, 1981; Rust and Donthu, 1995; Bell and Lattin, 1998). There are two main challenges for estimations at the level of elemental alternatives: (i) collecting data on all attributes for each of the elemental alternatives requires significant effort, and (ii) significant computational challenges arise when choice sets are very large. One approach to deal with the issue of very large choice sets is to use aggregated alternatives. Alternatives can be aggregated based on major selected characteristics suitable for the research context. Most studies in transportation research use spatial aggregates of alternatives (e.g. shopping centres, transportation analysis zones) when modelling shopping location choice consistent with research aims and data availability (Cadwallader, 1975; Koppelman and Hauser, 1978; Miller and O'Kelly, 1983; Thill and Wheeler, 2000). Studies in marketing, retailing and competition economics, on the other hand, typically use aggregates of alternatives based on store types, retail chains, or formats (Aaker and Jones, 1971; Burnett, 1978; Rust and Donthu, 1995; Messinger and Narasimhan, 1997; Bell et al., 1998; Gonzalez-Benito, 2002; Bhatnagar and Ratchford, 2004; Fox et al., 2004; Reutterer and Teller, 2009; Schiraldi et al., 2011). Models that focus on the choice between online and offline channels use aggregates of store alternatives based on channel as summarised in the previous paragraph. Aggregation will significantly reduce the number of alternatives resulting in computational gains and reducing data requirements. The disadvantage of using aggregated alternatives, however, is that the accuracy of the choice model will be reduced due to loss of information in aggregation and also potential aggregation bias. Aggregated alternatives will include a heterogeneous set of elemental alternatives and measuring attributes at the aggregate will inevitably contain measurement errors (e.g. travel time to each store vs. travel time to an aggregated set of stores), which can only be partially accounted for in model specification (Ben-Akiva and Lerman, 1985). Additionally, potential effects of future scenarios on a particular alternative cannot be predicted since the unit of analysis is not individual but aggregated alternatives. For example, effects of increased online shopping on a particular out-of-town superstore cannot be assessed, which might be of interest for policy making and business planning. Hence, it might be desirable to use elemental alternatives depending on research aims for developing choice models where data is available and use techniques to overcome computational issues.

3. Data sources

Finding an appropriate data source for studying channel choice poses a challenge. Regional and national travel surveys form much of the empirical data base for analysing shopping related mobility, yet they do not adequately capture online shopping. For instance, British National Travel Survey (NTS) asks respondents whether they have internet access at home, whether they usually do their main food shopping online or in-store, and whether they ever order groceries online.¹ In contrast to in-store shopping activities, however, specifics of online shopping activity are not recorded. Detailed information that is potentially relevant such as shopping basket characteristics and store attributes are not collected. Consumer panel data collected continuously by market research companies is a potentially valuable data source suited for this context as they provide information on shopping baskets and store attributes for both online and in-store shopping occasions. Consumer panel data is also attractive as market research companies operate consumer panels in many countries and make it available to researchers either free of charge or for a fee. For instance, Nielsen's consumer panel data from the US is available freely to accredited academic researchers through Chicago Booth School of Business ([Kilts Center, 2017](#)). Such wide availability is important for generalisability of our proposed method. For the present study, we have collaborated with Kantar, a large market research company with multinational operations, for data collection. An augmented version of their consumer panel data (Worldpanel) to capture additional travel related aspects was used for the modelling work in this study. Data enrichment involved investigating data sources available from websites and API-based tools, national data archives, and relevant companies.

3.1. Data collection: augmented consumer panel

Data from consumer panels are widely used in marketing literature ([Leicester, 2013](#)) mostly for studying product level decision-making and category management, but also for store choice modelling ([Siddarth et al., 1995](#); [Bell et al., 1998](#); [Bell and Lattin, 1998](#); [Gonzalez-Benito, 2002](#)), yet to our best knowledge, are not utilized in transport behaviour research. Consumer panels are run by market research companies where participating households use hand-held optical scanners to record daily purchases of fast moving consumer goods that are brought home. These are the products that are typically bought frequently and at relatively low prices including groceries, toiletries, health and beauty items you are likely to find in supermarkets. For our analysis here the focus is on grocery shopping and all shopping occasions where the basket contains at least one food or drink item was included. Panellists are also asked to report information on visited stores and channels (e.g., retail chain, online vs. in-store), provide household attributes (e.g., socio-demographics and home location), and record shopping basket characteristics (e.g., basket value in terms of monetary value, basket size in terms of number of items). Consumer panels are often continuous where households can participate for as long as they wish and receive incentives in the form of redeemable points. The collected data, however, were found to have some shortcomings for this line of work. First, information on the retail chain and channel is available, but the specific store is often not known and no data on store attributes is available. Second, no travel related information is recorded. To overcome these issues, panellists were asked to fill in an add-on survey to report the exact location of the stores they visited and travel modes used for each of the logged shopping trips for one week. Few additional trip-level recall questions were also included to capture relevant information on availability of choosing the car mode, travel costs, trip chaining, presence of other individuals, and their online shopping experience.

A city region like London was ideal for data collection where a variety of alternatives and online stores are available to households. Limiting the number of alternatives in the universal choice set was also important from a practical perspective as data gathering for store attributes required significant time. Hence it was decided to select two boroughs that are geographically closer to each other and that would maximize the sample size within the budget limitations. Barnet and Enfield was identified as being suitable when combined. Barnet and Enfield are boroughs in north London, with population sizes of 356,400 and 312,500 ([Office for National Statistics, 2011](#)). These boroughs are not among the most densely populated central regions with many supermarkets in London. Also, they are neighbouring boroughs hence households share many stores in their choice sets. Therefore, the burden of data collection for store attributes are reduced as selected.

Our sample consists of 121 main shoppers from Kantar's panel who responded to the add-on travel survey. Kantar Worldpanel defines main shoppers as household members that are responsible for the bulk of grocery shopping in their households. The dataset that we were provided with had an indicator for main shopper only for each logged trip and does not distinguish between other household members. Further, it was not possible to include other household members in the add-on survey due to company policies. Hence, for the present study, we use shopping occasions recorded by main shoppers only and the decision maker in our models is assumed to be the main shopper. This particular population segment will share certain demographic attributes in terms of age, gender, employment status, etc., which are potentially also influential on their preferences for online shopping and other store attributes. Also, main shoppers will have a higher frequency of grocery shopping, we observe that more than 85% of logged shopping occasions were conducted by main shoppers in our dataset. This might be related to responsibility sharing dynamics within the household and also reduced reporting quality for non-main shopper members. It might be argued that using data from main shoppers only is acceptable as they predominantly account for the highest share in spending and shopping trips generated and hence are most relevant for studying

¹ In the NTS questionnaire, under the section 'Home deliveries and food shopping', the question text is as follows: 'Nowadays, do you/does anyone in your household ever order any of these things [with a list of things to include and exclude] on the internet? We ask this question because we want to see to what extent shopping by internet reduces people's need to travel.'

shifts in observed patterns. However, if developed models are to be incorporated into activity-generation or trip-generation models for predicting demand at the aggregate, the estimations will not be valid. Estimated parameters should not be used for predictions using the whole population or a random sample drawn from the population. Table 1 presents a comparison of demographics for our sample and Census data from Enfield and Barnet in 2011 (Office for National Statistics, 2011). We note that no information is available from the Census on main shoppers, hence age of household reference person is used as proxy for comparison. The older age groups and highest social classes are overrepresented in our study sample.

It was not possible to collect additional location information on stores that do not belong to major supermarket chains in the UK due to panel restrictions. The retail chains included in the final dataset are as follows: 99P, Aldi, Asda, Co-op, Iceland, Lidl, Morrisons, M&S, Ocado, Poundland, Sainsbury's, Tesco, and Waitrose. Among these Morrisons, Sainsbury's, Tesco, Waitrose, and Asda operate both online and in-store shops. Ocado operates only through the online channel with the bulk of its fulfilment via Waitrose, and others operate via traditional physical stores (in-store) only. Chains also differ by their pricing strategies. Discounters (99P, Aldi, Lidl, Iceland, Poundland) are characterised by their low prices and the upscale retailers (MS, Waitrose, Ocado) offer higher quality products at higher prices. Stores with no location information, which accounted for 12.5% of all recorded occasions for the survey week and mostly consisted of smaller independent stores and farmers markets, were excluded from our analyses. Click&collect instances where the goods were ordered online and collected from a local branch or collection point are classified as online shopping in our analyses, which accounted for only two occasions. Ideally, the difference between click&collect and home deliveries should be accounted for in the model specification. However, it was decided to classify click&collect as online shopping due to lack of sufficient observations.

The final dataset used for model estimations consisted of 272 observations (shopping occasions) for one week (19th September 2014 and 26th September 2014) from 121 households residing in Enfield and Barnet. An overview of survey responses with a breakdown of observations by different questions are presented in Table 2. There were only 11 online shopping instances including two click&collect out of 272 recorded shopping trips. The limited number of observations mainly results from low adoption rates, as only 34 out of the 121 households have done any shopping online for the two years we have purchase data on. Data related issues for model estimation results are further discussed in Section 4. For the observed online instances, issues with deliveries seems to be minimal. In-store trips are mostly home-based where home is either the origin of the trip to the store, or destination of the trip from the store. The most common travel mode for grocery shopping trips was driving, followed by walking. Among the public transport modes, only two included tube and all others were bus trips. None of the respondents chose the taxi mode for their grocery shopping during the survey week.

3.2. Data enrichment: generating the choice set

The augmented consumer panel data included postcode level location and retail chain information for the chosen alternatives. No other store attributes are collected for visited stores. For estimating discrete choice models, detailed information on both chosen and unchosen store attributes are needed. For determining the choice set, a deterministic choice set generation process was adopted permitting the exclusion of in-store alternatives that are far from home locations. The distance threshold of 7.5 miles was used, which accounts for more than 95% of food shopping journeys for London residents in NTS (DfT, 2016). All stores from the twelve chains within the distance threshold from any of the household home locations were included in the choice set for all decision makers, totalling 1095 brick and mortar stores. For online alternatives, six chains that offered home-deliveries in London at the time of the survey (i.e., Asda, Sainsbury's, Tesco, Morrisons, Waitrose, and Ocado) were included. Collecting high quality data on unchosen alternatives for revealed preference data presents a challenge especially with large choice sets. The aim was to find publicly available sources that will enable replication of this study on a larger scale.

For attributes of in-store alternatives, we made use of several sources. The dataset of retail and leisure premises across Great Britain maintained by Local Data Company (LDC) was used to obtain information on all grocery stores in London including postcode, retail chain, and opening hours. Floor areas for each of the stores are sourced from the UK Valuation Office Agency (VOA) website, which is maintained for business tax calculation purposes. Spatial attributes can be used to account for competition and agglomeration effects, we experimented with several measures to incorporate these effects in our models. First, we make use of a measure for walkability as a proxy for accessibility to different services. Walk Score, a web based tool, computes a score with a corresponding band category based on distance to different categories of amenities and has been previously used in academic research (Brewster et al., 2009). While not officially supported in the UK, it provided useful information on spatial characteristics of store locations. For this study, we have computed walking scores for each store in the choice set using postcodes via Walk Score's API. In the final model specifications, we use a dummy for the highest score category that we label as high street dummy as a store attribute (Section 4). Secondly, using LDC data set and GIS software, we calculated the number of competitor supermarkets within a threshold distance. We experimented with different threshold values including 200 m, 300 m, 500 m, 1 mile. We note that there are different ways that have been proposed by previous studies of characterising built environment characteristics (Schiraldi et al., 2011; Borgers and Timmermans, 1987; Fotheringham, 1988).

For online alternatives, the literature survey revealed various attributes that influence choice behaviour for shopping as presented in Section 2. However, collection of data on attributes is more challenging. While the attributes of in-store alternatives are often static over time (e.g., floor area, opening hours, parking fees), virtual alternatives are often characterised by dynamic attributes. For online shopping, availability of delivery slots are determined based on demand at the time of pur-

Table 1

Household demographics: study sample vs. Census 2011 (Enfield and Barnet).

Age groups	Census (%)	Study sample (%)	Social class ^a	Census (%)	Study sample (%)
<35	19.9	14.5	AB	27.7	31.5
35–54	42.8	40.3	C1	32.4	41.7
55–64	15.5	24.2	C2	15.7	13.1
65+	21.8	21.0	DE	24.2	13.7

^a Based on approximated social grade produced by ONS. AB: Higher and intermediate managerial, administrative, professional occupations; C1: Supervisory, clerical and junior managerial, administrative, professional occupations; C2: Skilled manual occupations; DE: Semi-skilled and unskilled manual occupations, unemployed and lowest grade occupations.

Table 2

Overview of add-on travel survey responses (numbers of observations).

In-store	261		Online	11	
On my own	179	68.6%	Delivery on time	11	100.0%
Other adults	50	19.2%	Delivery not on time	0	0.0%
Children	23	8.8%	First attempted delivery	10	90.9%
Other adults and children	9	3.4%	Missed delivery	1	9.1%
Home-based	202	77.4%	No problems with delivery	9	81.8%
Non-home-based	59	22.6%	Other delivery problems	2	18.2%
Public transport	22	8.4%	Single delivery fee	5	45.5%
Car	181	69.3%	Membership fee	6	54.5%
Walking	58	22.2%			
Taxi	0	0.0%			

chase. Similarly, delivery fees are dynamically generated based on the demand for available slots, time of placing an order, basket size, personalised coupons and price matching offers. Therefore, there is a difficulty in capturing attributes of online alternatives with traditional data collection methods. This poses a significant challenge especially for unchosen alternatives (e.g., information on the next available delivery slot from the online store for an observation where a shopper visited an in-store alternative). The consumer panel surveys do not collect such information. No crowd sourced data was available to provide a measure of service quality for online retailers in London. Hence, the attributes of online alternatives for our modelling work was rather limited. We used average and median delivery fees over one week for each of the six retail chains from their websites. Additionally, we used information on delivery windows distinguishing between chains offering one-hour or two-hour delivery windows. None were found to be significant.

Retail chains compete over prices and can be distinguished by the pricing strategies they adopt. Since prices are determined weekly and at the product level, there is a challenge in quantifying price effects on store choice for a given shopping observation. We have attempted to compute a price index for each retail chain to give an average measure of price levels using the weekly pricing data using a representative shopping basket for each retail firm in the UK published by The Grocer magazine. The average prices for Tesco were normalised to unity, and all other chain price indices were computed by using Tesco average prices as reference. However, none of the price indices were found to be significant in estimated models in Section 4. This might be explained by price competition (price matching strategies) making the price difference between the same products as measured by The Grocer negligible.

Distances or travel times between the chosen and unchosen store alternatives and the decision-maker is not known. We do not observe the exact origin location for the observed shopping trip. Therefore, we use home location for computing the distances and travel times used in the model specification. Table 2 shows that most (77%) of the observed in-store trips either started or ended at home in our survey sample. Distances and travel times by different modes for a total of 132,495 home-store pairs (121 households, 1095 stores) were computed using API tools provided by mapping providers and Transport for London.

Store attributes and travel times data collected from different sources were merged with the revealed preference data from the augmented consumer panel data for the empirical modelling work.

4. Modelling analysis

4.1. Aggregation of alternatives and determining the choice set

Discrete choice models vary in the level of aggregation of the choice alternatives. In the context here, treatment of large choice sets of available stores introduce substantial challenges. There are competing advantages and disadvantages of different approaches as evidenced in literature (Section 2). Here, we report modelling analyses using two alternative approaches, using elemental store alternatives (model M1) and aggregated alternatives using store format and retail chain (model M2). A

comparison of results enables a better understanding of the extent to which the aggregation bias affects estimation results and predictive capabilities.

First, a model for channel, store, and travel mode choice was developed at the level of elemental store alternatives (model M1). Three travel modes are available to decision makers for the 1095 in-store alternatives: driving, public transport, and walking. We do not distinguish between different public transport options due to the limited number of observations and lack of variability in our dataset. Similarly, we do not distinguish between home deliveries and click&collect for the six online store alternatives. The resulting universal choice set for M1 consists of 3291 alternatives, $n = 3 * 1095 + 6$. Definition of the consideration sets is often challenging in discrete choice models as data on which alternatives were actually considered by the decision-maker for each observation is almost never available (Train, 2009; Shocker et al., 1991; Andrews and Srinivasan, 1995). Researchers' decisions are influenced by the study context and availability of data. Here, driving for in-store alternatives is only available to people who selected 'car was available for this shopping trip' option in response to the relevant question in the add-on survey. Online alternatives are only available to online shoppers (34 out of 121), which we define as individuals who have done any shopping online during the two-year period prior to our survey. In other words, if the decision maker is not observed making use of the online channel for the past two years, it is assumed that online alternatives are ignored by the individual when making a store choice decision. Inevitably, this assumption ignores certain behaviours. For instance, it can be argued that individuals who don't have a record of making online purchases are making a conscious decision about rejecting it (maybe following an unobserved experience with this channel), and online alternatives should be included in their consideration sets for parameter estimations. Similar situations where the analyst needs to make decisions regarding availability of alternatives arise in other familiar contexts such as travel mode choice. Driving licence holding often determines the availability of auto drive alternative, yet this also has the disadvantage of ignoring the decision about not getting a drivers license due to a conscious decision about mode choice. For the selected approach regarding consideration sets, implications on parameter estimations need to be considered for interpreting results. It is most likely that individuals who adopt online shopping are the ones who already have a positive attitude towards it, hence using our approach here will potentially overestimate associated taste parameters. Also, the online channel was only available for basket sizes over £25, as this was the minimum basket threshold required to qualify for home-delivery in London for the period studied. We note that this procedure implicitly assumes that the basket size is exogenous to the decision making processes for channel choice. Yet, in reality consumers might add more products to their grocery basket when shopping online just to match the minimum delivery limit. In other words, decisions regarding the basket size are endogenous to the store and channel choice. Here, we treat basket size as exogenous to simplify the model structure. To incorporate this additional choice dimension, one would need more data and use of discrete-continuous choice models. Our assumption here is a potentially important and restrictive one and needs to be addressed in future work as discussed further in Section 6.

Here, we develop a second model (M2) where we use aggregated store alternatives. We use retail firms (i.e. different chains such as Tesco, Asda, and Waitrose) and store formats (e.g. convenience, supermarkets, hypermarkets, superstores) for grouping alternatives into aggregated categories. Stores operated by the same retailer within the same format category share many characteristics that highly influence consumers' choice (e.g. product portfolio, pricing, store atmosphere). Compared to aggregation by spatial zones, grouping by chain and format is more suitable for our context and will account for some of the bias introduced by aggregation. The difficulty with this approach, however, is representing the heterogeneity in spatial characteristics for each of the aggregate alternatives such as travel times by different modes and near-by competitors. Format categories as defined by the Institute of Grocery Distribution (IGD)² are used for aggregation based on floor areas: convenience S - small stores with sales areas less than 280 m², supermarkets M - medium stores with sales areas between 280 m² and 2300 m², superstores L - large store with sales areas greater than 2300 m², and online stores (IGD, 2015). IGD definitions were largely consistent yet not always fully in line with firm definitions. We obtained 26 aggregate alternatives for channel and store choice: 20 in-store and six online. Each of these aggregate alternatives contains multiple stores that share the same main characteristics (e.g., format and firm) but are different with respect to other attributes (e.g., floor area, accessibility, local environment, travel times from home locations). Table 3 summarises the resulting set together with the number of stores within each aggregate alternative category based on the distance threshold definition used with M1. Not all retail firms operate through all channels. For instance, discounter stores only operate medium sized supermarkets and Ocado is an exclusively online retailer. The resulting universal choice set for M2 consisted of 66 alternatives, $n = 3 * 20 + 6$, and the same procedures were applied for determining the availability of driving and online alternatives as M1. Using aggregated alternatives required further specification search and a different set of attributes data was used for characterising each alternative as detailed in the following sections.

4.2. Model structure and estimation

For joint models (both M1 and M2) of channel, store location, and travel mode choice we use discrete choice models that belong to the family of random utility models. An overview of discrete choice methods, existing model structures, and estimation procedures are presented in Ben-Akiva and Lerman (1985) and Train (2009). In the context of store and channel

² IGD is a research and training charity in the UK that focuses on food and consumer goods industry with a strong community of corporate members representing leading grocery retailers in the UK. Conversations with retailers, market research companies, and academics in the field confirm the credibility of their understanding and segmentation of the grocery retailing market in the UK. More information on IGD is available from their website: <http://www.igd.com>.

Table 3

Aggregation of store alternatives using store format and retail firm.

	99P	Poundland	Aldi	Lidl	Iceland	Asda	Morrisons	Sainsbury's	Tesco	Waitrose	M&S	Co-op	Ocado
Convenience (S)								37	101		8	39	
Supermarket (M)	27	39	8	29	69	8		133	188	42	74	91	
Hypermarket (L)						18	27	62	59	36			
Online (O)							1	1	1	1	1		1

choice for instance, some stores might be similar to each other in ways that are not fully captured in the observed utility and their utilities will share many unobserved attributes. The basic multinomial logit (MNL) form will fail to account such correlation due to its assumption of independent random terms. Advanced nesting structures offer gains over the basic MNL structure in multi-dimensional choice situations where complex correlation structures arise. In this study, our initial hypothesis for jointly modelling channel, store, and travel mode choice is that applying nested, multi-level nested, and cross nested logit structures can capture more of the correlation in the unobserved utility. These formulations might capture heightened substitution, for instance, between two store alternatives that are either of the same channel (online vs. in-store) or belong to the same retail chain (Tesco's vs. Waitrose). Also, heightened correlation may exist between options that share a common travel mode where people are less likely to switch from driving to public transport when shopping. In destination and mode choice models, alternative nesting structures with destination over mode and mode over destination are often tested. We experimented with alternative nesting structures for both models. Fig. 1 presents selected structures we have tested for the model with elemental alternatives (M1).

When estimating choice models using aggregated alternatives (M2), elemental utilities are not computed, hence it is not possible to directly compute the utility associated with each aggregate alternative by taking the maximum of individual utilities. Instead, aggregated utilities need to be approximated using available attributes data. Natural logarithm of some size measure (e.g., number of stores, retail floor area, employment) is often included in the utility function to reduce aggregation bias (Mabit, 2011; Ben-Akiva and Lerman, 1985). In the context of store choice modelling for grocery shopping, Gonzalez-Benito (2002) compared different approaches for dealing with aggregation bias. When using aggregation by chains or store formats, as we propose here, using averaged utilities effectively means that the store attributes that relate to proximity and accessibility will lose their meaning leading to poor model fits. Using an alternative approach, aggregate utilities can be approximated using nearest store attributes, where the it is effectively assumed that the choice sets are composed of the nearest store from each chain category. Yet, in reality multiple stores from the same chain and format might be considered by the decision maker. Merely using nearest store attributes for computing aggregated alternative utilities will neglect intra-chain heterogeneity and size effects. Here, we explore the use of additional variables to capture some of the effects associated with the size and ease of access to stores within each category in addition to nearest store attributes.

All models presented were estimated using python BIOGEME (Bierlaire, 2003). We report empirical results for the basic MNL and NL models. No significant improvement over the more basic forms were achieved when using more complex forms (multi-level nested logit, cross-nested logit, and mixture models), potentially due to identifiability problems resulting from data limitations. We also note that the repeated nature of the data was not taken into account and a purely cross-sectional estimation was carried out.

5. Model results and discussion

Specification search was conducted using the MNL structure for both of the models M1 and M2. We start from a minimal specification with chain specific constants (i.e. chain dummies) only and add variables one by one, keeping the ones that improve the model fit and corresponding coefficients are significant. The aggregated alternatives used in M2 cannot be characterised in the same way as elemental alternatives in M1, hence there are differences in utility specifications. The decision maker in our models is assumed to be the individual (i.e., main shopper for the household). In the following two subsections we present estimation results from models M1 and M2. To further explore potential identification issues resulting from sample size and small number of online shopping instances, we conducted a Monte Carlo simulation study presented in Section 5.3.

5.1. Model M1 estimation results at the level of elemental alternatives

Table 4 shows estimation results for Model M1, both for the MNL model with the full choice set and the NL model depicted in Fig. 1b with sampling alternatives. None of the NL structures offered an improvement over MNL. We did not find the expected heightened correlation among virtual alternatives and among alternatives that share a common travel mode. As discussed further in Section 5.3, our experiments showed that while it might be possible to retrieve taste coefficients even with limited sample sizes, it is highly unlikely to identify heightened correlation among unobserved utilities (i.e. nesting parameters) using small samples. Hence, sample size limitations and low adoption rates for online alternatives might explain the lack of correlation.

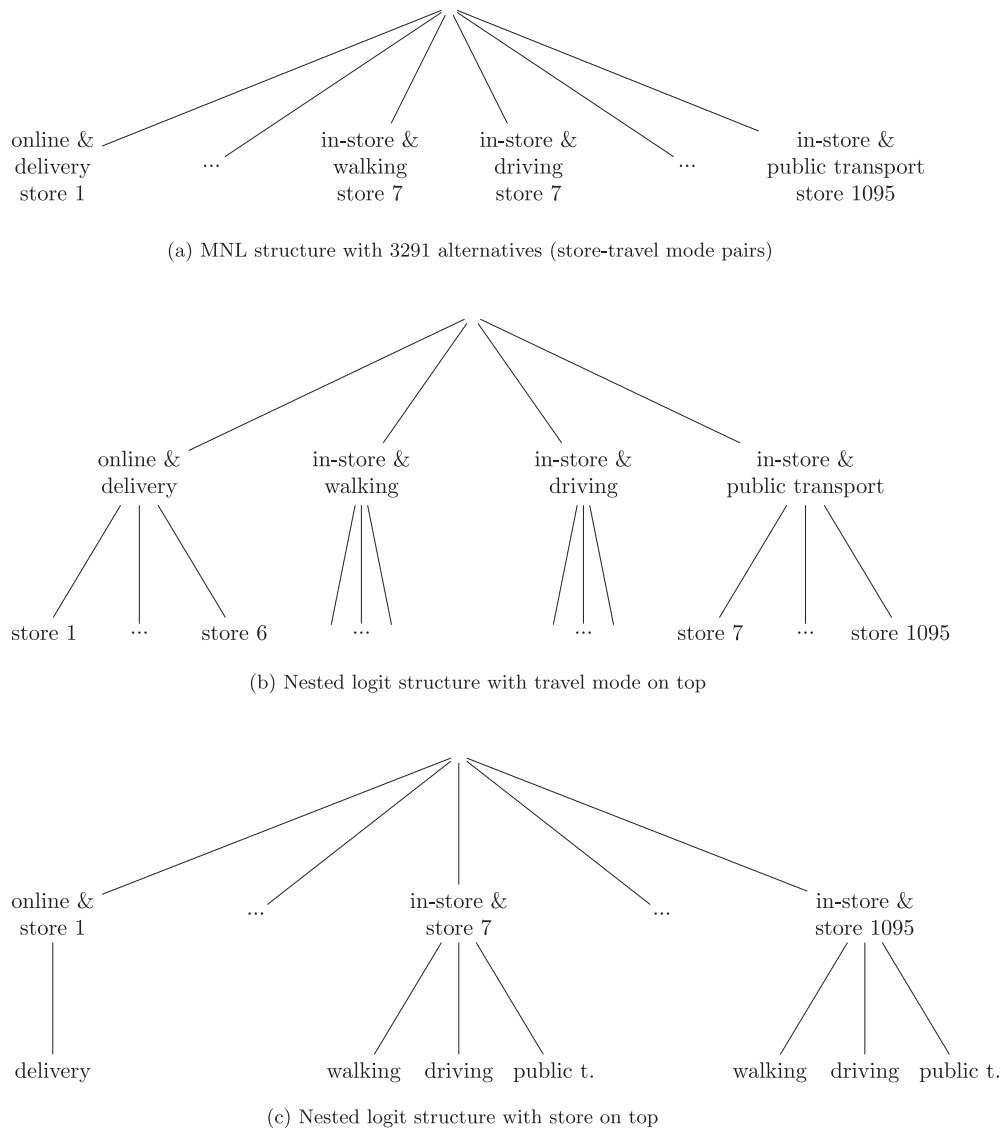


Fig. 1. Alternative logit structures for channel, store, and travel mode choice.

The socio-demographic characteristics of the decision maker were included in specification search for explaining the intra-personal heterogeneity. Income was missing for majority of the cases, social class was used as a proxy for income where higher social classes (Class AB) are assumed to be associated with higher income bands. We did not have information on whether the main shopper is male or female. Household size, number of children, age, education level, and life stage were not found to be significant when interacted with the online channel dummy. This finding might be related to online alternatives being available to households that have already adopted online shopping, hence model coefficients do not reflect preferences for adoption. As suggested by our preliminary findings, socio-demographic factors relate in different ways with adoption and the choice of online versus in-store for individual shopping occasions (Suel et al., 2015), which might explain the lack of significance.

No alternative specific constants enter the utility function due to the high number of alternatives; this might lead to endogeneity problems caused by the omission of unobserved attributes specific to each store alternative (Guevara, 2010). Instead, chain dummies to represent different firms, online channel dummy, and travel mode specific constants were used to capture mean effects of associated unobserved attributes. We also explored using additional dummy variables associated with area types, pricing strategies, and store formats. Dummies were also interacted with socio-demographic attributes and situational factors (e.g., basket size, presence of other decision makers, trip chaining). Not all chain dummy coefficients were significant potentially due to lack of enough observations for some chains. The coefficients for chain dummies that were significant all have negative signs in line with their lower market shares. The dummy for the highest social class category (AB)

Table 4

M1 estimation results: choice of store, channel and travel mode using elemental alternatives.

Description	MNL model				NL model			
	$J_{on} = 6, J_{in} = 1095 * 3 = 3285$				$J_{on} = 6, J_{in} = 200 * 3 = 600$			
	Coeff. est.	Std. err.	t-Stat	p-Value	Coeff. Est.	Std. err.	t-Stat	p-Value
99P	−1.660	0.753	−2.20	0.03	−1.710	0.731	−2.34	0.02
Aldi	0.400	0.415	0.96	0.34	0.252	0.481	0.52	0.60
Asda	−0.427	0.320	−1.33	0.18	−0.594	0.415	−1.43	0.15
Co-op	−1.550	0.349	−4.44	0.00	−2.400	0.530	−4.52	0.00
Iceland	−1.470	0.482	−3.05	0.00	−1.350	0.558	−2.42	0.02
Lidl	−0.662	0.438	−1.51	0.13	−0.774	0.466	−1.66	0.10
Morrisons	0.371	0.287	1.30	0.20	0.383	0.294	1.30	0.19
M&S	−1.340	0.437	−3.06	0.00	−0.948	0.474	−2.00	0.05
Ocado	1.260	0.669	1.88	0.06	1.200	0.691	1.74	0.08
Poundland	−0.794	0.538	−1.47	0.14	−0.601	0.558	−1.08	0.28
Sainsbury's	0.093	0.186	0.50	0.62	−0.140	0.213	−0.66	0.51
Tesco	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.
Waitrose	−0.772	0.338	−2.28	0.02	−0.643	0.446	−1.44	0.15
Upscale × Class AB	0.862	0.422	2.04	0.04	0.894	0.484	1.85	0.06
Mode online delivery	1.360	1.080	1.26	0.21	0.366	1.220	0.30	0.76
Mode driving	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.
Mode public transport	1.820	0.559	3.25	0.00	3.870	0.953	4.07	0.00
Mode walking	3.460	0.387	8.93	0.00	4.160	0.820	5.07	0.00
In-store × # store within 200 m	0.291	0.062	4.70	0.00	0.272	0.064	4.22	0.00
In-store × ln (floor area)	0.790	0.067	11.82	0.00	0.803	0.079	10.14	0.00
In-store × High-street dummy	−0.562	0.166	−3.40	0.00	−0.655	0.182	−3.60	0.00
Online × Class AB	2.180	1.030	2.12	0.03	2.320	0.966	2.40	0.02
Mode driving × ln (travel time with driving)	−2.560	0.086	−29.93	0.00	−2.790	0.119	−23.40	0.00
Mode public transport × ln (travel time with public transport)	−2.900	0.149	−19.52	0.00	−3.330	0.170	−19.54	0.00
Mode walking × ln (travel time with walking)	−2.460	0.094	−26.08	0.00	−2.690	0.136	−19.84	0.00
Mode walking × # items in basket	−0.029	0.011	−2.73	0.01	−0.031	0.011	−2.81	0.00
μ_{on}	–	–	–	–	1.000	0.000	0.00	1.00
$\mu_{driving}$	–	–	–	–	1.060	0.055	1.10	0.27
μ_{public}	–	–	–	–	1.190	0.111	1.71	0.09
$\mu_{walking}$	–	–	–	–	1.080	0.060	1.33	0.18
<i>Summary statistics</i>								
# observations	272				272			
# est. param	24				28			
$\mathcal{L}(\beta_0)$	−2197.92				−1236.13			
$\mathcal{L}(\beta)$	−1142.38				−734.792			
$\hat{\rho}^2$	0.469				0.388			

enters the utility function for upscale chains (i.e., Waitrose, M&S, Ocado) and online alternatives. Households in higher social class categories are more likely to visit upscale stores and shop online. As more people start shopping online, this would suggest that online channels would draw more aggressively from the upscale stores as higher income groups are associated with a higher preference for both online and upscale stores. It should be noted, however, these models might overestimate predicted shares of upscale retailers as taste coefficients capture preferences of early adopters. Their attitudes will not necessarily represent the preferences of the population. Furthermore, Ocado, which is an upscale retailer in our models, is among the first players in the online grocery market in the UK. Their first mover advantage may diminish over time.

For in-store alternatives, larger stores are more attractive as expected. The coefficient estimated was negative for the high street dummy and positive for number of competitors in close proximity. This finding might be seen as counter intuitive as the expectation is for agglomeration to have a positive and competition to have a negative effect on utility. This finding reveals a potentially interesting aspect of choice for grocery store choice. People are less likely to go to the city centre with many attractions (e.g., non-grocery shops, theatres, pubs) to do their grocery shopping. However, if the retail mix design includes and multiple supermarkets leading to a less dense centre design it will become more attractive.

The online dummy coefficient was positive but not significant suggesting that among basket sizes over £25, there is no significant association between online channels and larger baskets. Since online deliveries is not an option for smaller basket sizes due to supply side limitations, data was not available to study preferences for different channels for smaller baskets. It is, however, potentially an interesting area of research for retailers to understand the demand if they reduce delivery thresholds for online orders. Results show that walking trips are negatively associated with large baskets. Hence, in this model setting, online alternatives are a poorer substitute for walking trips especially when delivery is not an option for small baskets. This would also suggest driving trips will lose most with increasing adoption as their share was high to begin with and are positively associated with larger basket shopping.

Travel mode dummies were included in the final specification where the dummy for the driving mode was normalised to zero. Walking mode coefficient was positive and this positive utility effect declines with the number of items in the basket (i.e. walking mode becomes less attractive for large basket shopping). Coefficient for public transport was also positive with a smaller magnitude. This suggests walking and public transport modes are preferred to driving after other effects are accounted for, which might partially be related to lower costs associated with the walking and public transport modes as travel costs were excluded from the utility specification due to lack of data. Further, the hassles associated with the driving mode due to parking availability and fares might be influential. Public transport might also be attractive for certain routes in ways that are not captured by the utility specification. Travel time coefficients were estimated separately for each mode. As expected, all were negative and the largest effect is for public transport. Time spent in public transport is associated with more discomfort when compared to driving and walking. The travel time coefficients estimated for driving and walking were not significantly different from each other.

No price or delivery fee related variables were included in the final specification as corresponding covariates detailed in Section 3 were not significant. We note, however, the mean attractiveness captured by the chain dummies will account for some of the pricing and delivery fee policy effects. To capture trip chaining effects, we included a home-based indicator variable in the utility function (i.e., the decision maker was at home just before and just after the recorded grocery shopping activity), interacted with different travel modes and also travel times. However, none of the coefficients were significant.

5.2. Model M2 estimation results using aggregated alternatives

While some of the coefficients from M2 estimations using aggregated alternatives are in-line with M1 results using elemental alternatives, there are also important differences. This is an artefact of the differences in choice set definitions: M1 uses elemental alternatives ($n = 1101$) whereas M2 uses aggregated alternatives ($n = 26$). The latter M2 model essentially assumes that the choice sets consist of the nearest store from each aggregated category and aims to account for the potential inclusion of other alternatives in the choice set using additional proximity and size related variables. However, the actual choice set (i.e., set of all alternatives considered by the decision maker) is not observed and therefore cannot be used in choice set generation when estimating models. Hence, both models will have some bias in parameter estimations due to errors associated with choice set generation. Alternative approaches for the choice set generation in choice models can be used, yet none can eliminate the bias fully unless data on considered set of alternatives is collected. Below, we compare coefficient estimates to assess whether this bias has a significant influence on model results and interpretation.

Table 5 shows the estimation results from Model M2. For most specifications with nesting (by channel, travel modes, destinations), the nesting parameters were not significantly different from unity. The only nesting structure that offers a significant improvement over MNL was the structure presented in Fig. 2. This indicates higher correlation among unobserved utilities for in-store alternatives that belong to the 'group-of-five' (g5) nest, i.e., Asda, Tesco, Sainsbury's, Morrisons, Co-op. G5 nest consists of chains that grab the largest market share in the UK and use similar pricing strategies. We did not observe the same heightened correlation structure among the group-of-five stores in M1. It might, for instance, be related to similarities in store location strategies.

Format dummies, including the online channel dummy was insignificant. The dummy for the highest social class category (AB) enters the utility function in interaction with two formats: hypermarkets (size L) and online. Households in higher social class categories are more likely to shop from online and also large stores, with a stronger effect for the online channels. There is also a significant and positive relationship between hypermarkets and the number of items in the basket. Large stores and online stores are both associated with higher income shoppers and larger basket sizes. This suggests that an increase in online shopping adoption will potentially result in higher income groups to switch to online channels for their larger basket shopping substituting their trips to hypermarkets.

Also, the coefficient for the natural logarithm of the floor area for the nearest store was positive in line with M1 results. All chain dummies except for the Ocado dummy were included in the final specification. It was not possible to estimate both the Ocado and online dummy parameters, inclusion of both parameters for estimation leads to very high standard errors.

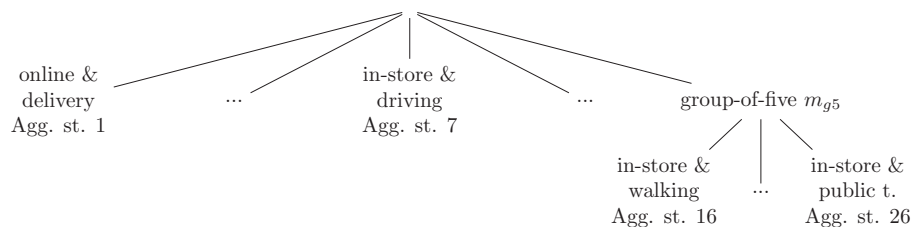
The coefficient for the walking mode, similar to M1, was positive and the positive utility effect declines with number of items in the basket. The public transport mode coefficient was negative in contrast to M1. Travel time coefficients were all negative, the largest effect is for walking followed by public transport. For public transport mode, increasing accessibility (as measured by the interaction with the variable for the number of stores within a 45 min travel time) reduces the negative effect of choosing public transport. The coefficient for the interaction term for the driving mode and number of stores within 30 min travel time is negative. This can be explained by the driving mode becoming less attractive when one has more options within easy reach from home location. When compared to M1 estimates, there are significant differences for certain coefficients but not others. For instance, it is potentially a more realistic assumption for the walking mode that only the nearest store in each category is being considered by the decision maker; hence M2's assumptions are not too restrictive leading to similar coefficient estimates for M1 and M2. For other modes, on the other hand, M2's assumptions affect coefficient estimates more significantly.

Coefficient of the natural logarithm of the size measure is significant and positive in the NL specification. Exclusion of this parameter have an effect on coefficients associated with some chain dummies and the large store format dummy. In the absence of a size variable, the large hypermarkets dummy coefficient becomes significant and remains positive. This finding

Table 5

M2 estimation results: choice of channel, store, and travel mode using aggregated alternatives.

Description	MNL Model				NL Model			
	Coeff. est.	Std. err.	t-Stat	p-Value	Coeff. Est.	Std. err.	t-Stat	p-Value
99P	−1.370	0.893	−1.53	0.13	−2.740	0.852	−3.21	0.00
Aldi	0.717	0.964	0.74	0.46	−0.669	0.743	−0.90	0.37
Asda	−0.323	0.613	−0.53	0.60	0.117	0.303	0.39	0.70
Co-op	−1.240	0.386	−3.20	0.00	−0.526	0.223	−2.36	0.02
Iceland	−1.030	0.536	−1.92	0.06	−2.330	0.577	−4.04	0.00
Lidl	−0.376	0.670	−0.56	0.57	−1.720	0.624	−2.75	0.01
Morrisons	0.048	0.407	0.12	0.91	0.268	0.209	1.28	0.20
M&S	−0.702	0.469	−1.50	0.13	−1.920	0.436	−4.39	0.00
Poundland	−0.810	0.651	−1.24	0.21	−2.180	0.668	−3.26	0.00
Sainsbury's	0.076	0.189	0.40	0.69	0.089	0.101	0.88	0.38
Tesco	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.
Waitrose	−0.735	0.363	−2.02	0.04	−1.760	0.363	−4.85	0.00
Format L (hypermarkets)	0.059	0.415	0.14	0.89	0.016	0.266	0.06	0.95
Format M (supermarkets)	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.
Format S (convenience)	−0.962	0.416	−2.31	0.02	−0.290	0.258	−1.13	0.26
Format O (online)	1.740	2.090	0.83	0.41	1.210	1.250	0.97	0.33
Mode driving	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.	Fixed.
Mode public transport	−1.960	0.805	−2.43	0.01	−1.050	0.493	−2.14	0.03
Mode walking	3.670	0.710	5.17	0.00	2.020	0.472	4.29	0.00
Format L × Class AB	0.681	0.343	1.99	0.05	0.366	0.220	1.67	0.10
Format O × Class AB	2.450	0.985	2.49	0.01	2.090	0.868	2.41	0.02
Format L × # items in basket	0.021	0.008	2.50	0.01	0.011	0.006	2.02	0.04
Mode walking × # items in basket	−0.110	0.033	−3.33	0.00	−0.061	0.020	−3.12	0.00
Nearest store ln(travel time - driving)	−1.050	0.143	−7.38	0.00	−0.643	0.114	−5.62	0.00
Nearest store ln(travel time - public transport)	−1.110	0.238	−4.65	0.00	−0.717	0.161	−4.46	0.00
Nearest store ln(travel time - walking)	−2.190	0.188	−11.67	0.00	−1.260	0.189	−6.65	0.00
Nearest store ln (floor area)	0.370	0.133	2.79	0.01	0.315	0.083	3.78	0.00
Mode driving × # stores within 30 min.	−0.023	0.008	−2.83	0.00	−0.014	0.005	−2.74	0.01
Mode public transport × # stores within 45 min.	0.062	0.016	3.94	0.00	0.040	0.009	4.44	0.00
ln (# stores)	0.505	0.322	1.57	0.12	0.498	0.173	2.88	0.00
μ_{g5}	−	−	−	−	1.960	0.287	6.86	0.00
<i>Summary statistics</i>								
# observations		272				272		
# est. param		27				28		
$\mathcal{L}(\beta_0)$		−1100.91				−1100.91		
$\mathcal{L}(\beta)$		−692.55				−682.07		
$\hat{\rho}^2$		0.346				0.355		

**Fig. 2.** Nested logit structure with a group-of-five nest.

suggests that the number of available stores within each category (i.e., size of the aggregated variable) is influential on the overall probability of the category being chosen.

5.3. Monte Carlo analysis of potential identification issues

In this section we explore potential identifiability problems arising from sample size issues using Monte Carlo experiments. This was motivated by the empirical estimation results presented in the previous two subsections where we did not find significant indication of heightened correlation that we expected among virtual alternatives and some of the coefficients we expected to be influential were not found to be significant. The model setting for Monte Carlo experiments was designed to be similar to the case of channel and store choice in this study, characterized by one small and one large nest, low levels of adoption, and small number of online shopping instances. The idea is to simulate observations based on a

known (true) underlying model, estimate these parameters using maximum likelihood estimation routines and perform hypothesis testing at 0.05 level, i.e. perform the same analysis as in the previous section. We implemented a nested logit with two nests ($K = 2$) and 26 alternatives, among which the first 6 belong to the small (online) nest ($J_{small} = 6$) and the other 20 to the large (in-store) nest ($J_{large} = 20$). We generated $N = 300$ and $N = 1500$ observations data using the true model where the alternatives in the small nest were only made available in 15% of the N observations, so that the number of online shopping observations were very small similar to the real case (e.g., always less than 5%). A total of $M = 5$ attributes were generated: two nest specific variables for each nest and one shared variable. The number of attributes also dictated the number of taste parameters that were estimated from the generated sample, one for each explanatory variable. Attributes x_m for each of the observations were randomly drawn from an iid uniform distribution in the range $(-1, 1)$. We set the true values of the taste coefficients to $\beta_m = 1$ and $\mu_{small} = \mu_{large} = 2$. To generate data files, 100 simulations were run. Note that a more comprehensive set of experiments were carried out where we also examine the effects of using different sample sizes, number of parameters, magnitudes of true parameters, numbers of shared and nest-specific parameters, and availability levels. In the interest of space, we only present a summary of most relevant findings here and refer the interested readers to Suel (2016). Dataset generation was completed using Python and estimations were carried out in MATLAB.

The summary results from Monte Carlo experiments are presented in Fig. 3. The first two columns in Fig. 3 show two plots for estimated taste parameters $\hat{\beta}_m$, one for the coefficient estimates that were significant at the $p < 0.05$ level (black dots) and the second for the ones that were not significant at the $p < 0.05$ level (white dots). Note that β_1 and β_2 are small-nest-specific, β_3 and β_4 are large-nest-specific, and β_5 are shared taste coefficients. Third and fourth columns show nesting parameter estimates associated with the small and the large nests, $\hat{\mu}_{small}$ and $\hat{\mu}_{large}$, across 100 simulation runs as stacked histogram plots. Black bars correspond to parameter estimates that were significantly different from one at the $p < 0.05$ level, and white bars correspond to the ones that were not significant at that level. Finally, in the fifth column, the absolute size of t-statistics are compared across different simulation estimations for $\hat{\beta}_m$ and $\hat{\mu}_k$ where larger values suggest higher statistical significance. The dashed horizontal lines represent the critical values for the one-tailed and two-tailed t-tests for the significance level $\alpha = 0.05$ with the corresponding degrees of freedom. We use the one-tailed t-test (which has a lower critical value) for the nesting parameters as μ_k is always greater than one to be consistent with utility maximization.

The estimation retrieved the taste parameters even for small sample sizes and low adoption rates, similar to the case in the real data. Estimation of nest-specific taste parameters associated with the rarely observed nest was more problematic (β_1 and β_2). For $N = 300$, taste coefficients associated with shared and large (in-store) nest-specific parameters were significantly different from zero for 72% of the generated experiments. For small (online) nest-specific coefficients, however, this was only 25%. When larger sample sizes are used $N = 1500$, average error in parameter estimates is substantially reduced and taste parameters are recovered with higher confidence. It was not possible to retrieve the nesting parameters for any of the experiments with $N = 300$. Even with higher sample sizes at 1500, identification is challenging for nesting parameters: 63% for μ_{small} and 96% for μ_{large} . This is also in line with the literature, which suggests parametric tests (e.g., t-test and like-

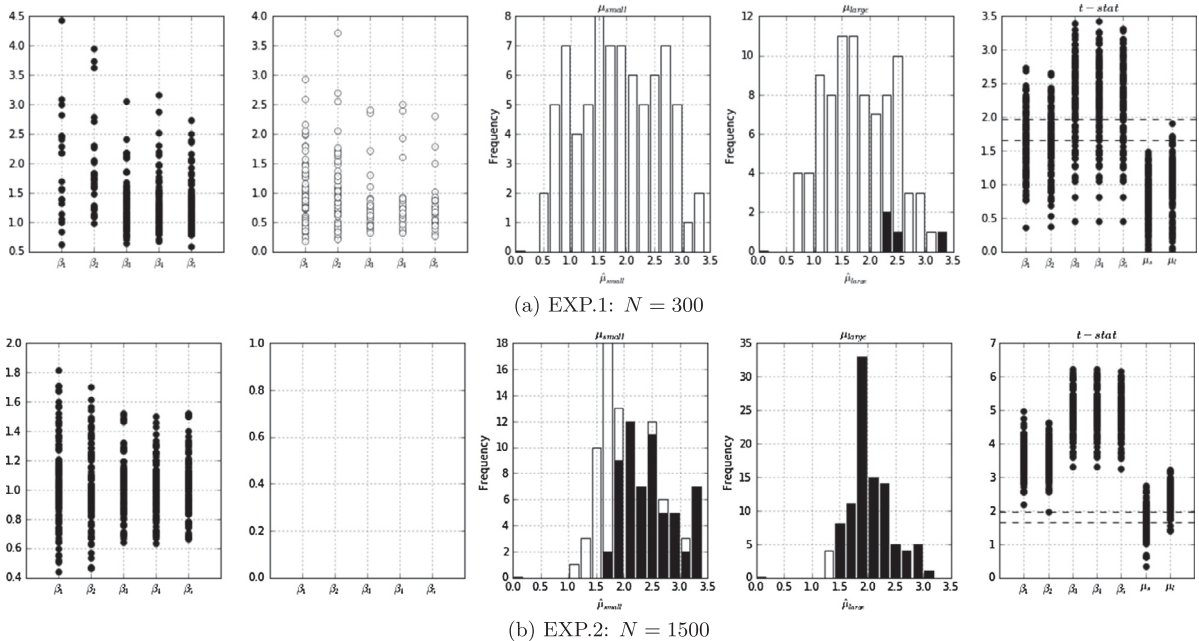


Fig. 3. Monte Carlo experiments: results summary.

likelihood ratio test) might be especially problematic for parameters where the null value is on the boundary of the parameter space such as the nesting parameter in the context here (Fitzmaurice et al., 2007; Anderson and Braak, 2003; Lee and Braun, 2012).

These findings have important implications for interpreting our results. First, results clearly showed that retrieving nesting parameters is highly unlikely with smaller sample sizes and low adoption rates. This is precisely the case for the estimation sample we use for the models developed in the previous sections. Hence, we need more data to confirm or reject the hypotheses regarding heightened correlation among alternatives (e.g., for nesting by channel, by different travel modes, different chains). Second, lack of significance for some of the taste coefficients that we expected to be influential (e.g., price expectations from each chain's store, delivery fees, trip chaining, presence of other adults or children when shopping) might be related to sample size issues. Monte-Carlo simulations suggest that this would especially affect online specific parameters. Third, our results where we report significant coefficient estimates are also in line with experiment results as power of tests are often sufficient to identify taste coefficients especially for shared and large-nest-specific parameters.

In addition to the Monte-Carlo simulations, we further performed validation experiments to quantify the model accuracy for prediction. For model validation, no additional dataset was available which contains information on store locations and mode choice. Therefore, we use a holdout sample consisting of observations from the same households as the estimation sample for the one month period prior to the survey week. Given the discrepancy in level of detail in the estimation and holdout samples, predicted and observed shares were compared using channel and retail chain data. We quantified the discrepancy between the observed and predicted market shares for the 18 aggregated categories. Both M1 and M2 were able to estimate the market shares with prediction error lower than 2.5% in root mean squared error (RSME). We observe that both M1 and M2 halve the RMSE prediction error compared to the naive method. In the naive method, individuals randomly choose one store among all available alternatives, each store has the same probability of being chosen. Similar behaviour can also be seen for the cross-entropy. This improvement in prediction demonstrates the power of the proposed methods in capturing choice behaviour despite not being able to identify the parameters significantly. Additional details and discussion of model validation is available from Suel (2016).

6. Summary and conclusions

In this paper, we develop discrete choice models for the joint choice of channel, shopping destination, and travel mode. This is a critical extension to the existing transport literature on destination choice which largely ignores online activity. The contribution here is that the online alternatives are incorporated within a unified modelling framework based on discrete choice methods. This allows for a realistic representation of choice behaviour among available shopping options that consist of both in-store and online alternatives, and multiple travel modes. This extension provides the tools to quantify the effects of increased online shopping on traditional physical stores and shares of different travel modes used for shopping. Our second contribution here is that we identify a well-known source of data widely used in other disciplines and demonstrate its potential in a new context with modest modifications to its data collection protocol. There was also a significant challenge in collecting high quality data on attributes for each of the stores in the choice set, which is a common issue for revealed preference data. This was an especially demanding task due to the large number of grocery stores available to London residents. We took this opportunity to investigate web based data sources that are becoming increasingly accessible with the advances in API tools. We show that these tools have the potential to significantly reduce data collection efforts for estimating discrete choice models with revealed preference data and also make it easier to repeat estimations for different time periods or samples.

We successfully estimated joint choice models using the enriched version of the consumer panel data. Our results provided evidence that virtual alternatives currently offer an attractive substitute among early adopters for large basket shopping mostly for high income groups. This might suggest a significant reduction in shopping trips to hypermarkets and supermarkets that are often associated with large basket shopping which might lead to store closures. Further, online deliveries mostly draw from driving trips and less so from walking; as walking mode is mainly attractive for small basket shopping and online channels are not yet available for small baskets. The present study also confirmed previous findings related to smaller stores and longer travel distances being associated with declining utilities. Agglomeration and competition effects also play a significant role in store choice.

During our study we also identified substantial limitations in empirical applications of discrete choice models, largely due to sample size issues. Crucially, we did not find any statistically significant improvement over the basic MNL structure when using nesting over channels or travel modes. Further, some taste coefficients associated with attributes that we expected to be influential on choice behaviour such as pricing strategies, delivery prices, types of products in the basket were not found to be significant. Our experiments presented in Section 5.3 show that lack of significance both for nesting and taste parameters might well be related to data limitations. Secondly, what became evident when comparing the empirical results from elemental and aggregated models is that the assumptions made on choice set generation influence parameter estimates and predictions. The bias introduced due to errors in choice set generation need to be explored in future work and to better assess the most suitable approach based on research aims. Third, the models developed treat basket related choices as exogenous, which is potentially an important restriction. In reality, consumers' decisions regarding what to buy will be highly influenced by the choice of channel and store. Further, treating shopping observations separately as independent instances is a limita-

tion as in reality consumers use a mix use of channels and stores over time to satisfy different needs. Future research is needed to explore methods to address these issues and overcome identified challenges.

Notwithstanding the limitations, results from model estimations provided new insights and contribute to existing capabilities to quantify preferences in the choice context studied. We have successfully demonstrated that our proposed framework for modelling joint choice of channel, store, and mode can be implemented in practice using existing data sources. Such demonstration is critical in this context where data is scarce and studies often rely on ad-hoc data collection that are hard to replicate. Even with our limited sample, we were able to quantify effects of different retail firms, income, agglomeration and competition, travel attributes, and shopping basket characteristics. Based on this framework, additional data collection and higher budgets will be required for reliably quantifying hypothesized effects and ensuring sample representativeness for real life implementations of policy or business strategies. Such effort will also allow to reveal additional variables and dimensions that affect behaviour. The challenges faced here also highlight the inadequacies in empirical data sources in allowing the use of advanced methodologies developed in discrete choice contexts to answer emerging questions in practice. We believe future research should also focus on developing data collection techniques to enable empirical applications of advanced modelling methods.

Acknowledgements

We thank Rees Jeffreys Road Fund for supporting this study and UK Engineering and Physical Sciences Research Council for supporting this work under the award EP/I038837/1. We would also like to thank Kantar Worldpanel and Local Data Company for providing the data used in this research.

References

- Aaker, D.A., Jones, J.M., 1971. Modeling store choice behavior. *J. Market. Res. (JMR)* 8, 38–42.
- Anderson, M., Braak, C.T., 2003. Permutation tests for multi-factorial analysis of variance. *J. Statist. Comput. Simul.* 73, 85–113.
- Andrews, R.L., Srinivasan, T., 1995. Studying consideration effects in empirical choice models using scanner panel data. *J. Mark. Res.*, 30–41.
- Balasubramanian, S., Raghunathan, R., Mahajan, V., 2005. Consumers in a multichannel environment: product utility, process utility, and channel choice. *J. Interactive Market.* 19, 12–30. <http://dx.doi.org/10.1002/dir.20032>.
- Bell, D.R., Ho, T.H., Tang, C.S., 1998. Determining where to shop: fixed and variable costs of shopping. *J. Market. Res. (JMR)*, 352–369.
- Bell, D.R., Lattin, J.M., 1998. Shopping behavior and consumer preference for store price format: why large basket shoppers prefer EDLP. *Market. Sci.* 17, 66–88.
- Bellman, S., Lohse, G.L., Johnson, E.J., 1999. Predictors of online buying behavior. *Commun. ACM* 42, 32–38.
- Ben-Akiva, M., Lerman, S.R., 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*, vol. 6. MIT Press. <http://dx.doi.org/10.2307/1391567>.
- Bhat, C.R., Sivakumar, A., Axhausen, K.W., 2003. An analysis of the impact of information and communication technologies on non-maintenance shopping activities. *Transp. Res. Part B: Meth.* 37, 857–881.
- Bhatnagar, A., Ratchford, B.T., 2004. A model of retail format competition for non-durable goods. *Int. J. Res. Mark.* 21, 39–59.
- Bierlaire, M., 2003. BIOGEME: A Free Package for the Estimation of Discrete Choice Models. Swiss Transport Research Conference.
- Borgers, A., Timmermans, H.J., 1987. Choice model specification, substitution and spatial structure effects: a simulation experiment. *Regional Sci. Urban Econ.* 17, 29–47. [http://dx.doi.org/10.1016/0166-0462\(87\)90067-6](http://dx.doi.org/10.1016/0166-0462(87)90067-6).
- Bosnjak, M., Galesic, M., Tuten, T., 2007. Personality determinants of online shopping: explaining online purchase intentions using a hierarchical approach. *J. Bus. Res.* 60, 597–605.
- Brewster, M., Hurtado, D., Olson, S., Yen, J., 2009. Walkscore. Com: A New Methodology to Explore Associations between Neighborhood Resources, Race, and Health. APHA Poster Presentation.
- Bucchioni, P., Liu, X., Weidenhamer, D., 2015. Retail E-commerce Sales.
- Burnett, P., 1978. Markovian models of movement within urban spatial structures. *Geographical Anal.* 10, 142–153.
- Butler, S., 2016. Amazon Fresh food deliveries 'to start this month in UK'. <https://www.theguardian.com/technology/2016/may/03/amazon-fresh-food-deliveries-understood-to-start-this-month>. [online].
- Cadwallader, M., 1975. A behavioral model of consumer spatial decision making. *Econ. Geography*, 339–349.
- Cao, X.J., Xu, Z., Douma, F., 2012. The interactions between e-shopping and traditional in-store shopping: an application of structural equations model. *Transportation* 39, 957–974.
- Centre for Retail Research, 2016. Online Retailing: Britain, Europe, US and Canada 2016. <http://www.retailresearch.org/online/retailing.php>. [online].
- Chang, M.K., Cheung, W., Lai, V.S., 2005. Literature derived reference models for the adoption of online shopping. *Inform. Manage.* 42, 543–559. <http://dx.doi.org/10.1016/j.im.2004.02.006>.
- Chintagunta, P.K., Chu, J., Cebollada, J., 2012. Quantifying transaction costs in online/off-line grocery channel choice. *Market. Sci.* 31, 96–114. <http://dx.doi.org/10.1287/mksc.1110.0678>.
- Chocarro, R., Cortiñas, M., Villanueva, M.L., 2013. Situational variables in online versus offline channel choice. *Electron. Commer. Res. Appl.* 12, 347–361.
- Circella, G., Mokhtarian, P.L., 2010. Complementarity or substitution of online and in-store shopping: an empirical analysis from Northern California. In: Transportation Research Board 89th Annual Meeting.
- Couclelis, H., 2009. Rethinking time geography in the information age. *Environ. Plan. A* 41, 1556.
- Crocco, F., Eboli, L., Mazzulla, G., 2013. Individual attitudes and shopping mode characteristics affecting the use of e-shopping and related travel. *Transp. Telecommun.* 14, 45–56.
- DfT, 2015. National Travel Survey: England 2014.
- DfT, 2016. National Travel Survey, 2002–2014. [data collection].
- Dholakia, U.M., Kahn, B.E., Reeves, R., Rindfleisch, A., Stewart, D., Taylor, E., 2010. Consumer behavior in a multichannel, multimedia retailing environment. *J. Interactive Market.* 24, 86–95.
- Dijst, M., Farag, S., Schwanen, T., 2008. A comparative study of attitude theory and other theoretical models for understanding travel behaviour. *Environ. Plan. A* 40, 831–847.
- Farag, S., 2006. E-Shopping and its Interactions with In-Store Shopping (Ph.D. Thesis). Utrecht University.
- Farag, S., Schwanen, T., Dijst, M., Faber, J., 2007. Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping. *Transp. Res. Part A: Policy Pract.* 41, 125–141.
- Fitzmaurice, G.M., Lipsitz, S.R., Ibrahim, J.G., 2007. A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics* 63, 942–946.

- Fotheringham, A.S., 1988. Note–consumer store choice and choice set definition. *Market. Sci.* 7, 299–310. <http://dx.doi.org/10.1287/mksc.7.3.299>.
- Fox, E.J., Montgomery, A.L., Lodish, L.M., 2004. Consumer shopping and spending across retail formats. *J. Bus.* 77, S25–S60. <http://dx.doi.org/10.1086/381518>.
- Girard, T., Korgaonkar, P., Silverblatt, R., 2003. Relationship of type of product, shopping orientations, and demographics with preference for shopping on the Internet. *J. Bus. Psychol.* 18, 101–120.
- Gladding, N., 2016. Five year forecast: the grocery market by 2021. <<http://www.igd.com/Research/Retail/Five-year-forecast-the-grocery-market-by-2021/>>. [online].
- Gonzalez-Benito, O., 2002. Overcoming data limitations for store choice modelling. Exploiting retail chain choice data by means of aggregate logit models. *J. Retailing Consumer Serv.* 9, 259–268.
- Guevara, C.A., 2010. Endogeneity and Sampling of Alternatives in Spatial Choice Models (Ph.D. Thesis). Massachusetts Institute of Technology.
- Ha, S., Stoel, L., 2009. Consumer e-shopping acceptance: antecedents in a technology acceptance model. *J. Bus. Res.* 62, 565–571.
- Hjorthol, R.J., 2002. The relation between daily travel and use of the home computer. *Transp. Res. Part A: Policy Pract.* 36, 437–452.
- Hsiao, M.H.H., 2009. Shopping mode choice: physical store shopping versus e-shopping. *Transp. Res. Part E: Logist. Transp. Rev.* 45, 86–95. <http://dx.doi.org/10.1016/j.tre.2008.06.002>.
- IGD, 2015. UK Grocery Retailing. <<http://www.igd.com/our-expertise/Retail/retail-outlook/3371/UK-Grocery-Retailing/>>. [online].
- Kilts Center, 2017. Nielsen Datasets. <<https://research.chicagobooth.edu/nielsen/datasets/>>. [online].
- Koppelman, F.S., Hauser, J.R., 1978. Destination choice behavior for non-grocery-shopping trips. *Transp. Res. Rec.: J. Transp. Res. Board.*
- Lee, O.E., Braun, T.M., 2012. Permutation tests for random effects in linear mixed models. *Biometrics* 68, 486–493.
- Leicester, A., 2013. The Potential Use of In-Home Scanner Technology for Budget Surveys.
- Liu, C., Forsythe, S., 2011. Examining drivers of online purchase intensity: moderating role of adoption duration in sustaining post-adoption online shopping. *J. Retailing Consumer Serv.* 18, 101–109. <http://dx.doi.org/10.1016/j.jretconser.2010.10.004>.
- Mabit, S., 2011. Vehicle Type Choice and Differentiated Registration Taxes. ETC 2011.
- Messinger, P.R., Narasimhan, C., 1997. A model of retail formats based on consumers' economizing on shopping time. *Market. Sci.* 16, 1–23. <http://dx.doi.org/10.1287/mksc.16.1.1>.
- Miller, E.J., O'Kelly, M.E., 1983. Estimating shopping destination models from travel diary data. *Prof. Geographer* 35, 440–449.
- Mokhtarian, P.L., 2003. Telecommunications and Travel. The case for complementarity. *J. Ind. Ecol.* 6, 43–57. <http://dx.doi.org/10.1162/108819802763471771>.
- Mokhtarian, P.L., 2004. A conceptual analysis of the transportation impacts of B2C e-commerce. *Transportation* 31, 257–284.
- Mokhtarian, P.L., 2009. If telecommunication is such a good substitute for travel, why does congestion continue to get worse? *Transp. Lett.*
- Mokhtarian, P.L., Salomon, I., Handy, S.L., 2006. The impacts of ICT on leisure activities and travel: a conceptual exploration. *Transportation* 33, 263–289.
- Mokhtarian, P.L., Tang, W.L., 2013. Trivariate probit models of pre-purchase/purchase shopping channel choice: clothing purchases in Northern California. *Choice Modelling: The State of the Art and the State of Practice*, 243–273.
- Nagurney, A., Dong, J., Mokhtarian, P.L., 2001. Teleshopping versus shopping: a multicriteria network equilibrium framework. *Math. Comput. Model.* 34, 783–798.
- Office for National Statistics, 2011. 2011 Census: Aggregate data (England and Wales). <<http://infuse.ukdataservice.ac.uk>> [computer file].
- Office for National Statistics, 2015. Retail Sales 2014.
- Pauwels, K., Neslin, S.A., 2011. Building with Bricks and Mortar: The Revenue Impact of Opening Physical Stores in a Multichannel Environment. Working Paper, pp. 1–29. [interrefdoi:10.1016/j.jretai.2015.02.001](http://dx.doi.org/10.1016/j.jretai.2015.02.001)<http://dx.doi.org/10.1016/j.jretai.2015.02.001>.
- Recker, W.W., Schuler, H.J., 1981. Destination choice and processing spatial information: some empirical tests with alternative constructs. *Econ. Geography* 57, 373–383. <http://dx.doi.org/10.2307/144216>.
- Reutterer, T., Teller, C., 2009. Store format choice and shopping trip types. *Int. J. Retail Distrib. Manage.* 37, 695–710.
- Rhee, H.S., Riggins, F.J., Kim, C., 2009. The impact of product type and perceived characteristics of the web on multifaceted online shopping behavior. *J. Organiz. Comput. Electr. Comm.* 19, 1–29.
- Rotem-Mindali, O., 2014. E-commerce: implications for travel and the environment. In: *Handbook of Sustainable Travel*. Springer, pp. 293–305.
- Rotem-Mindali, O., Salomon, I., 2007. The impacts of E-retail on the choice of shopping trips and delivery: some preliminary findings. *Transp. Res. Part A: Policy Pract.* 41, 176–189.
- Rotem-Mindali, O., Weltevreden, J.W., 2013. Transport effects of e-commerce: what can be learned after years of research? *Transportation* 40, 867–885.
- Rust, R.T., Donthu, N., 1995. Capturing geographically localized misspecification error in retail store choice models. *J. Market. Res. (JMR)*, 32.
- Salomon, I., 1985. Telecommunications and travel: substitution or modified mobility? *J. Transp. Econ. Policy*, 219–235. <http://dx.doi.org/10.1080/09595238100185051>.
- Salomon, I., 1986. Telecommunications and travel relationships: a review. *Transp. Res. Part A: General* 20, 223–238.
- Salomon, I., Koppelman, F.S., 1988. A framework for studying teleshopping versus store shopping. *Transp. Res. Part A: General* 22, 247–255.
- Schiraldi, P., Seiler, S., Smith, H., 2011. Supermarket Choice with Multi-Store Shopping: Measuring The effect of Format Regulation. Working paper.
- Schoenbachler, D.D., Gordon, G.L., 2002. Multi-channel shopping: understanding what drives channel choice. *J. Consumer Market.* 19, 42–53.
- Schröder, H., Zaharia, S., 2008. Linking multi-channel customer behavior with shopping motives: an empirical investigation of a German retailer. *J. Retailing Consumer Serv.* 15, 452–468. <http://dx.doi.org/10.1016/j.jretconser.2008.01.001>.
- Shocker, A.D., Ben-Akiva, M., Boccara, B., Nedungadi, P., 1991. Consideration set influences on consumer decision-making and choice: issues, models, and suggestions. *Market. Lett.* 2, 181–197.
- Siddarth, S., Bucklin, R.E., Morrison, D.G., 1995. Making the cut: modeling and analyzing choice set restriction in scanner panel data. *J. Market. Res. (JMR)* 32, 255–266. <http://dx.doi.org/10.2307/3151979>.
- Soopramanien, D.G.R., Robertson, A., 2007. Adoption and usage of online shopping: an empirical analysis of the characteristics of buyers browsers and non-internet shoppers. *J. Retailing Consumer Serv.* 14, 73–82. <http://dx.doi.org/10.1016/j.jretconser.2006.04.002>.
- Suel, E., 2016. Modelling the Relationship between Multi-Channel Retail and Personal Mobility Behaviour (Ph.D. Thesis). Imperial College London.
- Suel, E., Le Vine, S., Polak, J.W., 2015. Empirical application of expenditure diary instrument to quantify relationships between in-store and online grocery shopping: case study of greater London. *Transp. Res. Rec.: J. Transp. Res. Board* 2496, 45–54.
- Teltzrow, M., Meyer, B., Lenz, H.J., 2007. Multi-channel consumer perceptions. *J. Electr. Comm. Res.* 8, 18–31.
- Thielman, S., 2016. Google Gets Into the Home-Delivered Grocery Game with Express. <<https://www.theguardian.com/technology/2015/sep/08/google-express-home-grocery-delivery>>. [online].
- Thill, J.C., Wheeler, A., 2000. Knowledge discovery and induction of decision trees in spatial decision problems. In: *Spatial Economic Science*. Springer, pp. 188–205.
- Train, K.E., 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Visser, E.J., Lanzendorf, M., 2004. Mobility and accessibility effects of B2C ecommerce: a literature review. *Tijdschrift voor Economische en Sociale Geografie* 95, 189–205.
- Weltevreden, J.W.J., 2007. Substitution or complementarity? How the Internet changes city centre shopping. *J. Retailing Consumer Serv.* 14, 192–207. <http://dx.doi.org/10.1016/j.jretconser.2006.09.001>.
- Whipp, L., 2016. Walmart Partners with Uber and Lyft to Deliver Groceries. <<http://www.ft.com/cms/s/0/d6c6e844-2979-11e6-8b18-91555f2f4fde.html#axzz4CISmgeWF>>. [online].
- Zhou, Y., Wang, X.C., 2014. Explore the relationship between online shopping and shopping trips: an analysis with the 2009 NHTS data. *Transp. Res. Part A: Policy Pract.* 70, 1–9.