

An empirical application of an expenditure diary instrument to quantify the relationships between in-store and online grocery shopping: Case study of Greater London

Postprint of:

Suel, E., Le Vine, S., Polak, J. (2015) An empirical application of an expenditure diary instrument to quantify the relationships between in-store and online grocery shopping: Case study of Greater London. *Transportation Research Record #2496.* <http://dx.doi.org/10.3141/2496-06>

Esra Suel¹

Doctoral candidate, Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
Telephone: +44 20 7594 6100
Email: e.suel10@imperial.ac.uk

Scott Le Vine

Research associate, Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
Telephone: +44 20 7594 6100
Email: slevine@imperial.ac.uk

John Polak

Professor and Chairman, Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
Telephone: +44 20 7594 6100
Email: j.polak@imperial.ac.uk

Copyright © 2015 National Academy of Sciences. All rights reserved.

¹ Corresponding author

Abstract

There is a sustained shift underway of certain types of retail activity away from in-store shopping and towards online retailing, with potentially structural consequences for shopping-related mobility. In the UK, for instance, 5.1% of spending on groceries in 2013 was transacted online, an increase from 3.8% in 2010 (the level was 1.1% in 2003). Transport researchers face serious gaps in empirical data coverage of this phenomenon, however, 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. To address this issue, we employed a well-established and nationally-representative data resource in a novel approach. The UK's Living Costs and Food Survey is traditionally used to track aggregate household expenditure patterns as well as to monitor price inflation. This study, by contrast, drew on the unique nature of its expenditure-diary instrument, in which respondents record each item they purchase during a two-week period – including, crucially, whether each shopping occasion was in-store or online. Empirically, it was found that shopping ‘basket’ characteristics are significantly linked with channel choice. Furthermore, using a two-stage modelling approach, it was found that socio-demographic factors appear to relate in different ways with *adoption* of online shopping in general and the choice of online versus in-store for individual shopping occasions. The paper closes with a brief discussion of future research needs to advance this line of enquiry.

1 Introduction

In recent years, the retail industry has been affected by advances in information and communication technologies (ICTs). As a result, the nature of shopping activity is changing, with people increasingly using a wider mix of different channels (online and in-store). In the UK, for instance, 5.1% of spending on groceries in 2013 (and 12.7% of all retail sales) was transacted online, an increase from 3.8% in 2010 and 1.1% in 2003. (Kantar World Panel, “Shopping for Groceries” Aug 6 2013). These changes are in turn having important implications for shopping related personal mobility.

In Britain, shopping’s share of total personal trips fell by 20% between 1997 and 2012, and the average distance travelled for a shopping trip increased by approximately 10%, and it is speculated [1] that these structural changes may relate, at least in part, to the impacts of certain types of retail activity shifting away from in-store purchases and towards online transactions. Transport researchers lack the evidence on which to confirm or reject this speculation, however – in part this results from gaps in empirical data coverage, as traditional travel surveys collect quite limited information with which to establish how in-store and online shopping behaviour interact with one another. The effect of this uncertainty is to limit the ability of transport policy-makers to respond effectively to the changing marketplace.

To address this issue, this study employed a well-established and nationally-representative data resource in a novel approach. The UK’s Living Costs and Food Survey (LCF) is collected continuously by the Office for National Statistics (ONS), and is traditionally used to track aggregate household expenditure patterns as well as to monitor price inflation over time. This study, by contrast, drew on the unique nature of its expenditure-diary instrument, in which respondents record each item they purchase during a two-week period -- including, crucially, whether each shopping occasion was in-store or online (see Figure 1). We report findings based on analysis of the 2011 LCF dataset, with a focus on grocery shopping in Greater London. To the authors’ knowledge, this is the first empirical study of its kind to draw on the LCF (or peer large-scale expenditure-diary datasets in other countries).

The remainder of this paper is structured as follows: Section 2 discusses the background of online shopping behaviour and its potential implications on personal travel. Section 3 describes the empirical dataset and methods employed in this study. The empirical analyses from LCF dataset is presented in Section 4. Finally, the implications of the findings from this study and future research needs are discussed in Section 5.

2 Background

2.1 Interaction between online activity and personal travel

The potential travel impacts of increasing use of information and communication technologies (ICTs) has been a consistent line of enquiry among transport researchers. At a conceptual level Mokhtarian and colleagues developed the seminal taxonomy of the types of impacts of people’s online activity on their physical mobility: (i) *substitution*, where (in the case of shopping) online activity takes place instead of conventional in-store shopping activity, (ii) *complementarity*, where online activity leads to additional trips and/or makes them more efficient by providing additional information and capabilities, (iii) *modification*, where there is a shift in behaviour and trips (e.g. timing of trips, trip chaining) associated with online activity that leads to neither a net substitution or complementarity effect, and (iv) *neutrality*, where online activity does not influence travel [2-6].

One strain of the relevant literature consists of empirical studies aimed at quantifying the net impacts of online shopping in terms of vehicle-miles travelled or shopping-journey frequency, with a range of findings. Interested readers are referred to comprehensive reviews by Rotem-Mindali and Weltevreden [7], Rotem-Mindali [8], Bhat, Sivakumar [9].

Quantifying the consequences of changing shopping activity patterns for related personal mobility is challenging due to the multiple and sometimes counter-acting relationships [10, 11]. ‘Shopping’ itself is a heterogeneous class of activities; for instance, recurring shopping for groceries may involve different behavioural mechanisms (with different mobility implications) than occasional shopping for white goods items (household appliances) [11-14].

Different stages of the shopping process (identification of need/desire for an item or service, information gathering, transaction/purchase, delivery/possession, and possibly return of an unwanted item) are quite distinct for certain types of products [11, 15]. The information gathering stage is likely to be relatively long-duration and may involve conducting unique trips for infrequently-purchased and high value items such as motor vehicles and personal computers. For routine grocery shopping, however, information-gathering likely does not involve separate visits to stores. Hence, the choice between online and offline channels for different stages of the shopping process may well vary between different product categories [13, 15, 16].

2.2 Choice of shopping ‘channel’ (online vs. in-store)

Given the theoretical ambiguity and empirical challenges, research is needed to better understand people’s participation in in-store and online shopping activity and the consequences for their personal travel [9, 17].

Researchers have developed a number of conceptual frameworks of people’s choice of shopping *channel* (online vs. in-store) [18-24]. In this segment of the literature, researchers have identified the distinctive benefits (e.g. reducing uncertainty, value of physical assessment, recreational aspects, time savings, additional information) and costs (e.g. travel to store, delivery costs) associated with in-store and online shopping. Attributes of the available shopping channels (e.g. prices, quality, intrapersonal service quality, delivery service quality etc.) also appear to influence choice behaviour. The degree to which shopping-activity outcomes are impacted is also mediated by people’s preferences.

Chang, Cheung [25] presented a comprehensive review of empirical studies that aim to identify the determinants of adoption of online-shopping behaviour. Variables frequently considered include the internet’s distinctive characteristics as a retailing channel (e.g. the degree of privacy, security, convenience, delivery service, prices etc.), characteristics of individual retailer-websites/apps and products (e.g. brands, reputation, website/app design), and consumer socio-demographics. The focus is in general to model whether an individual is an online shopper (i.e. does any shopping online at all) or not, hence studies in this section of the literature are not intended to identify how online shopping behaviour interacts with in-store shopping. Furthermore, findings regarding the impacts of specific co-variates are frequently mixed or inconclusive (cf. Figure 1 in [25]). Interested readers are also referred to Soopramanien and Robertson [26], [27-34].

Cao, Xu [35] present a detailed review of the literature on the relationship between frequencies of in-store and online shopping activities that use various statistical techniques (descriptive statistics, multivariate analyse, structural equation models, etc.). They report that such studies tend to identify complementarity effects – i.e. a positive relationship between online and in-store shopping frequency. The authors conclude that online shopping is unlikely to substitute physical shopping in the aggregate.

Circella and Mokhtarian [36] also report complementarity effects and finds limited substitution effects – i.e. a negative relationship between online and in-store shopping frequency for experience (clothing/shoes) and search (books/DVDs) goods.

A separate segment of the literature has investigated the choice of shopping channel (online vs. in-store) for specific shopping occasions, using methods such as binary logistic regression [15, 37, 38]. 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) [15, 38], 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. In the latter case, additional data on behavioural co-variates (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. For instance, Mokhtarian and Tang [39] modelled the joint choice of *pre-purchase* and *purchase* channels for clothes-shopping using revealed preference data, where respondents were asked about the channel choice for different shopping stages (i.e. purchase, information gathering, trial) on their most recent purchase.

Studies such as those listed above are in general based on ad-hoc survey data, which are not always readily available or subject to rigorous data-collection and data-quality protocols. Purely cross-sectional data where a single recent shopping occasion is described (using prompted-recall methods, cf. Mokhtarian and Tang [39]) are limited in that they cannot account for possible behaviour where consumers make use of multiple shopping channels over time to replenish their stock of products. A novel approach was introduced by Chintagunta, Chu [40], who analysed channel choice for grocery shopping using a bespoke dataset of shoppers' scanner data² from a single retail chain. The underlying hypothesis that Chintagunta and colleagues reported support for was that people choose the shopping channel for when it minimizes each shopping occasion's total transaction costs (with transaction costs represented as a function of distance to store, weekday/weekend indicator, time of day, total basket cost, weather conditions, delivery costs, and in-store promotions, presence of 'heavy/bulky' and 'perishable' items in the basket).

3 Methods and data

The relevant marketing-science and transport literature suggests multiple interrelated factors influence consumers' channel choice, as outlined in previous section. Factors identified in the literature include:

- Attributes of alternative channels: prices, product and service quality, store atmosphere, website design, opening hours, travel times, delivery costs, retail chain
- Individual tastes and situational factors: socio-demographics, attitudes toward shopping and channel alternatives, variety seeking, recreational aspects of shopping activity, monetary and time-budgets, mobility attributes, storage capacity, spatial characteristics and shopping accessibility attributes of home/work locations, shopping basket characteristics, presence of other decision makers
- Stage in shopping process: searching/browsing, purchase, returns

² Scanner panel data is collected by retailers using point-of-sales systems. This particular scanner data set consists of time of shopping, channel (online/offline) and/or store visited, details of items bought for each shopping trip conducted by each household in the sample.

- Past purchases and plans for future shopping: temporal dynamics, to account for stock depletion/replenishment behaviour.
- Products categories: books, clothing, furniture, vehicles.

Few empirical datasets that address both in-store and online-shopping phenomenon adequately, and to date have been undertaken as one-off data collection efforts. Regional and national travel surveys, which form much of the empirical data base for analysing shopping-related mobility, typically collect quite limited information with which to establish how in-store and online shopping behaviour interact with one another. For instance, the 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. In contrast to in-store shopping activities, specific episodes of online-shopping activity are not recorded in NTS respondents’ weeklong diary. The US *National Household Travel Survey* (NHTS) includes similar questions; respondents are asked how often they used the internet in the past month, how many times they purchased something through the internet in the past month and how many of these purchases were delivered to their home.

On the present study, the approach we took to compile the required empirical data resource was to employ the UK’s *Living Costs and Food Survey* (LCF) [41]. The LCF (formerly the *Expenditure and Food Survey*) is a well-established and nationally-representative (with application of appropriate statistical weighting to account for sampling error) survey that economists typically use to monitor aggregate trends in consumer spending and price inflation.³ The key component of the LCF of interest to this study is the 14-day expenditure diary instrument.

Our empirical analysis focuses on channel choice for grocery shopping, which in Britain in 2010 accounted for approximately half of all shopping journeys [43]. We sought to model two linked dimensions of shopping behaviour – whether or not a person engages in any online shopping (defined by whether they are observed to purchase any item online during their two-week expenditure-diary period), and whether individual grocery shopping occasions take place in-store or online (conditional on a person doing any online shopping). We hypothesised that people may tend to shop online when purchasing bulky items that may be burdensome to transport, and also highly-standardized products (e.g. canned food, boxed cereals) where the risk of getting an undesired item is minimal in comparison to less-standardized products such as fresh fruits and vegetables. For such less-standardized items, there is high value to being able to visually inspect before purchase.

The LCF is the UK’s principal household expenditure survey undertaken on a continuous basis by the *Office of National Statistics*, with an annual sample size of approximately 6,000 households [44]. LCF data on respondents’ spending is collected via a paper expenditure diary instrument where each person records for two weeks the items they purchase each day, where they are purchased, and their price. The sampling unit is the household; all household members age 7+ complete diaries, with a separate simplified diary instrument used by children aged 7 to 15. The expenditure-diary instrument is complemented by a household interview that collects information on socio-demographics and large, infrequent purchases. The LCF dataset that we employed contains an indicator of whether each item in a person’s expenditure diary was purchased online or in-store, but more detailed characteristics of stores (such as format, e.g. big-box stores, convenience stores, etc.) are not recorded.

The scope of this analysis encompasses residents of London who took part in the LCF in 2011 and recorded any food purchases during their two-week diary period (n=452). Expenditure diaries were

³ The US’ counterpart to the British LCF is the Bureau of Labour Statistics’ *Consumer Expenditure Survey* [42].

converted into shopping basket records for the purposes of this study. All items purchased from the same store on a given day were used to generate separate shopping occasions or baskets⁴. All shopping baskets with at least one food or non-alcoholic-drink item were defined to be ‘grocery-shopping’ activities and hence were included in the analysis.

Items purchased during grocery-shopping occasions were categorized under three ordinal product categories based on our judgement of their relative attractiveness for online and in-store purchasing:

- **Online-oriented:** Product types judged to be most amenable to online purchasing (standardized products and bulky products)
- **Neutral:** All product types judged to fall within neither of the other two categories (Table 1).
- **In-store-oriented:** Product types judged to be most amenable to in-store purchasing

Table 1 shows sample product types in each of the three categories⁵.

<<Table 1 about here>>

The co-variates employed in the empirical analysis were:

1. **Socio-demographic characteristics:**

- Age
- Gender
- Household-level gross weekly income (British Pounds)
- Current Economic Status (1 if employed; 0 otherwise 1)
- Number of Adults (defined in the UK as age 16+) in household
- Presence of children in household (under age 16)
- Binary indicator of household-level car/van ownership (1 if yes, 0 if no)

2. **Online shopper:** A binary indicator variable of whether the respondent is an ‘online shopper’ or not, where ‘online shopper’ is defined as having made any purchases from the internet during their 14-day diary period

3. **Grocery shopping and basket characteristics for each grocery shopping occasion:**

- Channel (online or in-store)

⁴ The LCF diary instrument does not distinguish between single and multiple visits to a given store on a given day. This differs from travel surveys, which record multiple visits to the same store on the same as separate shopping occasions but which typically do not make full-address spatial information available to researchers, which would be required to identify which store is visited.

⁵ In the interests of space, we have placed the categorization of all 314 product types observed in the empirical dataset in an online Appendix (Appendix ‘A’), which can be accessed via the following persistent URL:

https://workspace.imperial.ac.uk/people/Public/s.le-vine/TRB_2015_Suel_et_al_Online_and_In-Store_Shopping_Appendix_A.pdf

- Basket cost (total amount paid for all products in the basket [excluding petrol⁶] converted into three bands of basket-cost: less than or equal to £25, greater than £25 and less than or equal to £50, greater than £50)
- Share of online-oriented/neutral/in-store-oriented products as a percentage of total basket cost
- Number of shopping baskets that contain at least one item from online-oriented/neutral/in-store-oriented product category
- A categorical variable in which each shopping basket is classified on the basis of the types of products it contains:
 - i. online-oriented products only,
 - ii. neutral products only
 - iii. in-store-oriented products only
 - iv. online-oriented and neutral products only
 - v. online-oriented and in-store-oriented products only
 - vi. neutral and in-store-oriented products only
 - vii. all three product categories
- Number of items in each of the three product categories (online-oriented, neutral, and in-store-oriented)
- Share (percentage) of online-oriented/neutral/in-store-oriented products as a percentage of total *number of items* in the basket

Table 2 and Table 3 contain descriptive statistics for the reported variables, including a cross-tabulation with being an online shopper.

<<Tables 2 and 3 about here>>

The analysis employs the most recent available LCF microdata, for which fieldwork was undertaken in 2011. Online grocery retailing is evolving quickly, with an increasing number of companies offering services, increased catchment areas, tighter delivery slots, and new services (e.g. click & collect from stores and other collection points, virtual stores) [45, 46]. Our results must be viewed with this fast pace of change in mind; they are a snapshot of behaviour occurring in 2011. The models we estimated were linear-in-parameters binary logit functional form; the two stages described above were estimated sequentially.

4 Results

In this section we first present bi-variate correlations between online and in-store grocery shopping to see whether the data suggest substitutions or complementarity effects at the aggregate level. These results are then followed by the estimation results from the two-stage multi-variate analysis described in the previous section. In the first stage model the dependent variable is *adoption* of online shopping, while in the second stage model the dependent variable is whether specific shopping occasions took place online or in-store

⁶ The term *gasoline* in the North American vernacular is referred to as *petrol* in British English.

Correlations between total *spending* and *number of shopping occasions* conducted online and in-store are shown in Table 4, both at the household and individual level and for each diary day and aggregated across the entire 14-day diary period. A total of eight correlations are shown ($2 \times 2 \times 2$), and it can be seen that all eight correlations between online and in-store grocery shopping activity are negative; negatively-signed correlation coefficients are consistent with the theory of substitution between online and in-store grocery shopping occasions.

We found that, at the household level, the correlations between *spending* for groceries online and in-store were not significant, either when we look at each person's 14-day diary as a single data record or when we treat each person-day as a separate data record. By contrast, the corresponding two correlations at the household level between the *number of shopping occasions* online and in-store were both statistically-significant (and negative).

From Table 4 it can also be seen that both of the two correlations at the person-level (one being spending and the other the number of grocery shopping instances) between online and in-store spending are not significant when the person-record is the unit of analysis, but are when the person-day is the unit of analysis.

Finally, a generic observations from the bi-variate correlations presented in Table 4 is that in all cases, the negative correlation between the number of grocery shopping *occasions* is stronger (i.e. has a larger absolute value) than the corresponding value for grocery *spending*.

<<Table 4 about here>>

We now turn to the estimation results from the first-stage model (Table 5), which for convenience we refer to as online-shopping *adoption*. The binary dependent variable is the observation of whether or not people that took part in the LCF did or did not purchase any items online during their expenditure diary period. Goodness-of-fit, using the standard McFadden pseudo- r^2 statistic, was 0.17.

<<Table 5 about here>>

Age was found to have a negative and statistically significant *ceteris paribus* relationship with the likelihood of being an online shopper; in other words, younger people were more likely to be online shoppers, all else equal. No significant effect was found for gender or for employment status, but income was found to be positively associated, net of confounding effects, with being an online shopper. Smaller household sizes (measured both by the number of adults and also by the presence of children) were negatively associated with adoption of online shopping. Finally, no significant effect was found for car ownership.

In the second stage of model estimation, analyses were performed at the basket-level for online shoppers only. In other words, this stage of analysis evaluated the likelihood that any given grocery shopping occurred in-store or online. Analyses were conducted using a sub-set of the dataset that consisted solely of people that were defined to have 'adopted' online shopping behaviour. We estimated five alternative specifications of the second-stage model (Labelled: 2S-1, 2S-2,...,2S-5), each of which contain both socio-demographic and shopping-basket-related variables. A range of alternative ways to characterise shopping baskets were modelled, in order to test our initial hypothesis of a relationship between online shopping activity and shopping-basket composition. The model runs vary in how the basket-related variables are specified, as follows:

Model run #2S-1 contains the *share of amount paid* for online-oriented/neutral/in-store-oriented products as a percentage of the total cost of the shopping occasion.

Model run #2S-2 includes three 0/1 binary indicator variables for whether each shopping occasion contained at least one item from each of the online-oriented, neutral, and in-store-oriented product categories.

In **Model run #2S-3**, each shopping occasion is defined by exactly one of the following seven ($2^3 - 1 = 7$) binary indicator variables, based on the products purchased during the shopping occasion:

- 1) *online-oriented* products only
- 2) *neutral* products only
- 3) *in-store-oriented* products only
- 4) *online-oriented* and *neutral* products only
- 5) *online-oriented* and *in-store-oriented* products only
- 6) *neutral* and *in-store-oriented* products only
- 7) all three product categories (*in-store-oriented*, *neutral* and *online-oriented*)

Model run #2S-4 contains three ordinal count variables (0,1,2,...) that represent the number of items in the basket from each of the three product categories.

Model run #2S-5 is similar to Model Run 2S-1, the difference being that it contains the share of the *number of items* from each of the three product categories as a percentage of the total number of items in the basket.

Goodness-of-fit was higher for the second-stage models than the first-stage model, with pseudo- r^2 values of approximately 0.55 for all five specifications of the second-stage model.

The parameter estimates for the second-stage model suggest quite different net effects for some variables that were also included in the first-stage model. Age was positively-signed (and statistically significant) in the first-stage model, but is negatively-signed and significant in the second-stage model runs. Income is not statistically significant in any of the second-stage model runs; by contrast it was positive and statistically-significant in the first-stage model. The effect of car ownership was also different for the two stages: no significant effect associated with car ownership was found in the first-stage model (whether or not a person ‘adopts’ online shopping), but in all five second-stage models owning cars/vans in one’s household is negatively associated, net of confounding effects, with using online-shopping for individual shopping occasions. Furthermore, unlike the first-stage model, the second-stage models all indicate no statistically-significant effects associated with income or the two household-structure variables (number of adults and presence of children).

We now turn to the effects associated with each shopping occasion (as opposed to the socio-demographics of the shopper). The remainder of this section describes the results from each of the five second-stage model runs, each of which have different specifications of the shopping occasion variables. First, however, we note that the three-band variable of the amount spent on each shopping occasion provides consistent results across all five models. Shopping occasions where £50 (approx.. \$85 USD) or more are spent were the most likely to have been online-shopping instances, and those where £25 of less was spent are the most likely to have occurred in-store. There are several plausible causal mechanisms for this result, though the available data do not allow us to distinguish between them. For instance, delivery charges may be more burdensome for smaller ‘basket sizes’, and/or larger

‘basket sizes’ may be more amenable to delivery as the consumer offloads onto the retailer the task of transporting the groceries to their home.

Model run #2S-1 uses monetary share of amount paid for each of the three product categories in the basket as the representation of basket characteristics. The product categories (online-oriented, neutral, and in-store-oriented) are defined (as described in Section 3) as ordinal categories based on the researcher team’s judgment of their propensity for in-store versus online purchasing. All effects are relative to the reference category of in-store-oriented (i.e. hypothesized to be least amenable to online purchasing) category. The results of Model run #2S-1 are consistent with our hypotheses; the estimated parameters are monotonic in the theorized manner. It is also worth pointing out that Model run #2S-1 is, on the basis of its adjusted pseudo r^2 statistic (which allows comparison between alternative specifications where the number of parameters varies), statistically preferred to the other four second-stage model run specifications.

Model #2S-2 has three binary variables that indicate whether or not a given shopping occasion contained items in each of the three product categories. The ordering of the three parameters is consistent with our hypotheses, but all three parameters were found to be statistically insignificant.

Model #2S-3 contains an exhaustive combinatorial set of the three binary indicator variables for the presence of at least one item from each product category. Of the seven possible combinations, we only observed four in the empirical dataset and therefore estimate only three free parameters. None of the three freely-estimated parameters were statistically-distinguishable from the reference category (online-oriented products only).

In model run #2S-4 shopping occasions are characterized by the number of items in each of the three categories (online-oriented, neutral, and in-store-oriented). We found a positive marginal effect (statistically-significant only at $p<0.1$, not $p<0.05$) between the number of online-oriented items and the likelihood that a shopping occasion took place online. A negative effect was found for ‘neutral’ products; the marginal effect associated with the number of in-store-oriented items was not statistically significant.

Model run #2S-5 is comparable to Model run #2S-1, except that it uses the percentage share of *items* in each product category within the basket rather than the percentage of the amount spent on the shopping occasion. Results were similar to Model run #2S-1 in that the effects for the three categories are ordered monotonically in the hypothesized manner (0 for in-store-oriented < 4.95 for neutral < 5.15 for online-oriented). However, the adjusted pseudo r^2 values indicate that the specification of Model run #2S-1 is preferred. In other words, more statistical information to identify whether a shopping occasion takes place online versus in-store was found in the percentage of *spending* for each product category during the shopping occasion than in the *number of items purchased* in each category.

5 Conclusions

This study reports, a novel application of a large-scale, nationally-representative expenditure diary dataset to investigate disaggregate relationships between online and in-store shopping. The empirical case study focused on grocery shopping activities by residents of London, England.

We estimated bi-variate correlations and straightforward two-stage binary logistic models (the first stage being whether a person ‘adopts’ online shopping, and the second stage being whether online-shopping-adopters shop online or in-store on specific shopping episodes).

Our results suggest, within the timescale observed by the *Living Costs and Food Survey*'s expenditure diary instrument (14 days), net *substitution* effects between purchasing groceries online and in-store. All of the set of bi-variate correlations between online and in-store grocery shopping that we analysed were negative (though not all were statistically significant).

The multivariate statistical analysis identified different patterns in the statistical correlates in the two stages (*adoption* and choice of *in-store* versus *online*). For instance, no statistically-significant *ceteris paribus* relationship was found between car ownership and whether or not a person *adopts* online shopping, but among online-shopping-adopters car ownership was found to correlate negatively with choice of online shopping (versus in-store shopping). Beyond socio-demographics, total amount spent and the distribution of spending on 'online-oriented', 'neutral', and 'in-store-oriented' products were found to be statistically significant correlates of choice of online shopping for specific shopping occasions.

This line of research has important implications for the future research agenda. The finding that the correlation between online-shopping and in-store shopping activity varies temporally suggests that travel diaries (such as the British *National Travel Survey* and the U.S. *National Household Travel Survey*) that do not record online-shopping activity are neglecting relevant data that could potentially be collected. Survey methods research is needed, however, to understand the trade-offs that may be involved (e.g. decreased response rates, increased under-reporting, etc.) in adding complexity to the established instrument packages for surveys such as these.

Future efforts on this line of research will test hypothesized relationships between features of the built-environment and propensity to take part in online versus in-store shopping activities. Retailing is changing rapidly at the time of writing; it is the authors' hope that the novel use of background data resources will prove to be fruitful in advancing the state-of-knowledge during this period of flux.

6 Acknowledgments

The authors wish to thank the *Office of National Statistics* and the *UK Data Archive's Secure Data Service* for providing access to *Living Costs and Food Survey* microdata. The usual disclaimer applies: any errors are solely the authors' responsibility.

References

1. Department for Transport, *National Travel Survey: 2012 Statistical Release*. 2013.
2. Salomon, I., *Telecommunications and travel: substitution or modified mobility?* Journal of transport economics and policy, 1985: p. 219-235.
3. Salomon, I., *Telecommunications and travel relationships: a review*. Transportation Research Part A: General, 1986. **20**(3): p. 223-238.
4. Mokhtarian, P.L., *A typology of relationships between telecommunications and transportation*. Transportation Research Part A: General, 1990. **24**(3): p. 231-242.
5. Mokhtarian, P.L., *Telecommunications and travel: The case for complementarity*. Journal of Industrial Ecology, 2002. **6**(2): p. 43-57.

6. Hjorthol, R.J., *The relation between daily travel and use of the home computer*. Transportation Research Part A: Policy and Practice, 2002. **36**(5): p. 437-452.
7. Rotem-Mindali, O. and J.W. Weltevreden, *Transport effects of e-commerce: what can be learned after years of research?* Transportation, 2013. **40**(5): p. 867-885.
8. Rotem-Mindali, O., *E-Commerce: Implications for Travel and the Environment*, in *Handbook of Sustainable Travel*. 2014, Springer. p. 293-305.
9. Bhat, C.R., A. Sivakumar, and K.W. Axhausen, *An analysis of the impact of information and communication technologies on non-maintenance shopping activities*. Transportation Research Part B: Methodological, 2003. **37**(10): p. 857-881.
10. Golob, T.F. and A.C. Regan, *Impacts of information technology on personal travel and commercial vehicle operations: research challenges and opportunities*. Transportation Research Part C: Emerging Technologies, 2001. **9**(2): p. 87-121.
11. Mokhtarian, P.L., *A conceptual analysis of the transportation impacts of B2C e-commerce*. Transportation, 2004. **31**(3): p. 257-284.
12. Visser, E.J. and M. Lanzendorf, *Mobility and accessibility effects of B2C e-commerce: a literature review*. Tijdschrift voor Economische en Sociale Geografie, 2004. **95**(2): p. 189-205.
13. Rotem-Mindali, O. and I. Salomon, *The impacts of E-retail on the choice of shopping trips and delivery: Some preliminary findings*. Transportation Research Part A: Policy and Practice, 2007. **41**(2): p. 176-189.
14. Girard, T., P. Korgaonkar, and R. Silverblatt, *Relationship of type of product, shopping orientations, and demographics with preference for shopping on the Internet*. Journal of Business and Psychology, 2003. **18**(1): p. 101-120.
15. Hsiao, M.-H., *Shopping mode choice: Physical store shopping versus e-shopping*. Transportation Research Part E: Logistics and Transportation Review, 2009. **45**(1): p. 86-95.
16. Farag, S., *E-shopping and its interactions with in-store shopping*. 2006: Utrecht University.
17. Mokhtarian, P.L. and I. Salomon, *Emerging travel patterns: Do telecommunications make a difference? In: In perpetual motion: Travel behavior research opportunities and application challenges*, in *Presenation Series*. 1997, Institute of Transportation Studies, University of California, Davis: UCD-ITS-RP-97-39.
18. Balasubramanian, S., R. Raghunathan, and V. Mahajan, *Consumers in a multichannel environment: Product utility, process utility, and channel choice*. Journal of Interactive Marketing, 2005. **19**(2): p. 12-30.
19. Schoenbachler, D.D. and G.L. Gordon, *Multi-channel shopping: understanding what drives channel choice*. Journal of Consumer Marketing, 2002. **19**(1): p. 42-53.
20. Salomon, I. and F. Koppelman, *A framework for studying teleshopping versus store shopping*. Transportation Research Part A: General, 1988. **22**(4): p. 247-255.
21. Nagurney, A., J. Dong, and P.L. Mokhtarian, *Teleshopping versus shopping: a multicriteria network equilibrium framework*. Mathematical and Computer Modelling, 2001. **34**(7): p. 783-798.

22. Bosnjak, M., M. Galesic, and T. Tuten, *Personality determinants of online shopping: Explaining online purchase intentions using a hierarchical approach*. Journal of Business Research, 2007. **60**(6): p. 597-605.
23. Dholakia, U.M., et al., *Consumer behavior in a multichannel, multimedia retailing environment*. Journal of Interactive Marketing, 2010. **24**(2): p. 86-95.
24. Couclelis, H., *Rethinking time geography in the information age*. Environment and planning A, 2009. **41**(7): p. 1556.
25. Chang, M.K., W. Cheung, and V.S. Lai, *Literature derived reference models for the adoption of online shopping*. Information & Management, 2005. **42**(4): p. 543-559.
26. Soopramanien, D.G. and A. Robertson, *Adoption and usage of online shopping: An empirical analysis of the characteristics of "buyers" "browsers" and "non-internet shoppers"*. Journal of Retailing and Consumer Services, 2007. **14**(1): p. 73-82.
27. Teltzrow, M., B. Meyer, and H.-J. Lenz, *Multi-channel consumer perceptions*. Journal of Electronic Commerce Research, 2007. **8**(1).
28. Pauwels, K. and S.A. Neslin, *Building with bricks and mortar: The revenue impact of opening physical stores in a multichannel environment*. 2008, MSI Reports, 08-101, 25-45.
29. Ha, S. and L. Stoel, *Consumer e-shopping acceptance: Antecedents in a technology acceptance model*. Journal of Business Research, 2009. **62**(5): p. 565-571.
30. Liu, C. and S. Forsythe, *Examining drivers of online purchase intensity: Moderating role of adoption duration in sustaining post-adoption online shopping*. Journal of Retailing and Consumer Services, 2011. **18**(1): p. 101-109.
31. Bellman, S., G.L. Lohse, and E.J. Johnson, *Predictors of online buying behavior*. Communications of the ACM, 1999. **42**(12): p. 32-38.
32. Rhee, H.-S., F.J. Riggins, and C. Kim, *The impact of product type and perceived characteristics of the web on multifaceted online shopping behavior*. Journal of Organizational Computing and Electronic Commerce, 2009. **19**(1): p. 1-29.
33. Schröder, H. and S. Zaharia, *Linking multi-channel customer behavior with shopping motives: An empirical investigation of a German retailer*. Journal of Retailing and Consumer Services, 2008. **15**(6): p. 452-468.
34. Huang, Y. and H. Oppewal, *Why consumers hesitate to shop online: An experimental choice analysis of grocery shopping and the role of delivery fees*. International Journal of Retail & Distribution Management, 2006. **34**(4/5): p. 334-353.
35. Cao, X.J., Z. Xu, and F. Douma, *The interactions between e-shopping and traditional in-store shopping: an application of structural equations model*. Transportation, 2012. **39**(5): p. 957-974.
36. Circella, G. and P.L. Mokhtarian. *Complementarity or Substitution of Online and In-Store Shopping: an Empirical Analysis from Northern California*. in *Transportation Research Board 89th Annual Meeting*. 2010.
37. Cao, X.J., *The relationships between e-shopping and store shopping in the shopping process of search goods*. Transportation Research Part A: Policy and Practice, 2012. **46**(7): p. 993-1002.

38. Chocarro, R., M. Cortiñas, and M.-L. Villanueva, *Situational variables in online versus offline channel choice*. Electronic Commerce Research and Applications, 2013. **12**(5): p. 347-361.
39. Mokhtarian, P.L. and W.L. Tang, *Trivariate probit models of pre-purchase/purchase shopping channel choice: Clothing purchases in northern California*. 2011, Institute of Transportation Studies, University of California, Davis: Research Report - UCD-ITS-RR-11-15.
40. Chintagunta, P.K., J. Chu, and J. Cebollada, *Quantifying transaction costs in online/off-line grocery channel choice*. Marketing Science, 2012. **31**(1): p. 96-114.
41. Office for National Statistics and Department for Environment - Food and Rural Affairs, *Living Costs and Food Survey, 2011*. August 2013, Colchester, Essex: UK Data Archive: SN:7272, <http://dx.doi.org/10.5255/UKDA-SN-7272-2>.
42. Bureau of Labor Statistics, *2012 consumer expenditure interview survey public use microdata documentation*. 2013: Available online at: <http://www.bls.gov/cex/2012/csxitvw.pdf>.
43. Department for Transport, *National Travel Survey 2002-2010 [computer file]*. 7th Edition. Colchester, Essex: UK Data Archive [distributor], April 2012. SN: 5340, <http://dx.doi.org/10.5255/UKDA-SN-5340-3>. 2012.
44. Office for National Statistics and Department for Environment, *2011 Living Costs and Food Survey: User Guide*. 2012: Available online at: <http://discover.ukdataservice.ac.uk/catalogue/?sn=7272&type=Data%20catalogue>.
45. Kantar Worldpanel, *Shopping for Groceries: What if online retail was bigger than Tesco?* 2013: Available online at: <http://www.kantarworldpanel.com/global/News/What-if-online-retail-was-bigger-than-Tesco>.
46. Bamfield, J.A.N., *Retail Futures 2018: Shop Numbers, Online and The High Street*. 2013, Centre for Retail Research.

Table 1: Product categorization used for characterizing shopping baskets

Online-oriented (bulky and/or highly-standardized)

Sweetened breakfast cereals e.g. Sugar Puffs, Coco Pops
All yoghurt (including twinpots, yoghurt drinks) except frozen yoghurt
Tinned peaches, pears & pineapples
Frozen beans
Tomatoes, canned or bottled
Canned & carton soups
Soft drinks
Beers
Cleaning equipment e.g. hoover, steam carpet cleaner
Toilet paper

Neutral

White Bread
Pizza - frozen or non-frozen
Eggs
Olive Oil
Hard cheese - cheddar type
Dried fruit eg sultanas, raisins, currants, dates, apricots
Tomatoes, fresh
Chocolate bars
Vegetable juices e.g. tomato juice, carrot juice
Wine

In-store-oriented (less-standardized & physical inspection is highly valuable)

Sandwiches
Beef steak
White Fish
Salmon
Fresh Oranges
Haberdashery
Outerwear
Carpets and rugs
Fancy & decorative goods including mirrors, vases
Medicines

a. The full list of product categories is available from: https://workspace.imperial.ac.uk/people/Public/s.l-evine/TRB_2015_Suel_et_al_Online_and_In-Store_Shopping_Appendix_A.pdf

Table 2: LCF 2011 - Grocery Shoppers in London: Descriptive statistics for reported variables in the empirical analysis at the individual level

	All Grocery Shoppers	Online shoppers
Household-level variables	n = 292	n = 70
Average gross normal weekly household income	£984	£1,668
Adults per household	1.78	1.83
% of households with at least one child	32.5%	35.7%
% of households with at least one car or van	55.5%	67.1%
Individual-level variables	n = 452	n = 82
Average age (Mean)	44.3	
Average total personal gross income per week	£570	£7912
% of male respondents	45.4%	37.8%
% of full-time or part-time employees	63.7%	80.5%
Average number of grocery shopping instances over the 2-week diary period	6.22	6.35
Average number of online grocery shopping occasions over the 2-week diary period	0.07	0.39
Average number of in-store grocery shopping occasions over the 2-week diary period	6.15	5.96

Table 3: LCF 2011 - Grocery shopping baskets: Descriptive statistics for reported variables in the empirical analysis at the basket level

	All Grocery Shoppers' all baskets	Online shoppers' all baskets	Online baskets
	n = 2811	n = 521	n = 32
Average basket cost	£16.44	£20.69	£79.78
Average number of items in a Basket	8.3	9.2	29.0
% baskets in £0<x<£25 range	83%	76.6%	- a
% baskets in £25= q<x<£50 range	10.6%	13.2%	- a
% baskets in >£50 range	6.4%	10.2%	- a
Average share of amount paid for online-oriented products in the basket	31.1%	31.2%	39.5%
Average share of amount paid for neutral products in the basket	39.5%	42.4%	42.9%
Average share of amount paid for in-store-oriented products in the basket	29.4%	26.4%	17.7%
% baskets that has at least one online-oriented	68.9%	71.4%	87.5%
% baskets that has at least neutral product	79.2%	81.8%	81.3%
% baskets that has at least one in-store-oriented product	62.8%	61.0%	96.9%
Average number of online-oriented products in a basket	2.8	3.2	12.3
Average number of neutral products in a basket	3.5	4.1	12.0
Average number of in-store-oriented products in a basket	2.0	2.0	4.6
Average share of (number of) online-oriented products in the basket - in terms of	32.6%36	32.8%	38.5%
Average share of (number of) neutral products in the basket	42.4%	45.3%	46.6%
Average share of (number of) in-store-oriented products in the basket	25.0%	21.8%	14.9%
% baskets: online-oriented products only	9%	9%	- a
% baskets: in-store-oriented products only	5.3%	- a	- a
% baskets: neutral products only	12.6%	14.4%	- a
% baskets: online- & in-store-oriented products	6.5%	5%	- a
% baskets: online-oriented & neutral products	15.5%	15.5%	- a
% baskets: in-store-oriented & neutral products	13.1%	- a	- a
% baskets: all product categories	37.9%	41.8%	81.3%

a. Value cannot be published due to privacy issues; applies where cell counts are below n = 10

Table 4: Correlations between online and in-store shopping (p-value in brackets)

	Correlation between spending (aggregated across shopping occasions) for groceries online and in- store	Correlations between <i>number of shopping occasions</i> for groceries online and in-store
Household-level		
HH with internet shoppers, aggregated over full 14-day diary period	-0.187	-0.226**
HH with internet shoppers, where each day that any grocery shopping activity is performed is treated as a separate observation	-0.083	-0.252***
Person-level		
Internet shoppers, aggregated over full 14-day diary period	-0.025	-0.057
Internet shoppers, where each day that any grocery shopping activity is performed is treated as a separate observation	-0.115**	-0.392***

a. Values marked with *, **, or *** are statistically significant at p<0.10, p<0.05, and p<0.01 respectively

Table 5: Results from binary logistic regression models of *adoption* of online shopping (first-stage model)

Sample size	n = 452
Null log-likelihood	-213.34
Final log-likelihood	-176.61
McFadden's pseudo- r^2	0.172
McFadden's adjusted pseudo- r^2	0.139
<i>Parameter estimates</i>	
Intercept	-5.60***
Age (years)	-0.03***
Respondent is male (binary indicator)	-0.63
Respondent is employed (binary indicator)	-0.38
Household Income - ln(gross weekly income)	1.13***
Number of adults in the household	-0.79***
Presence of children in the household (binary indicator)	-0.63**
Car/van ownership (binary indicator)	0.10

a. Reference category: not an online shopper

b. Values marked with *, **, or *** are statistically significant at p<0.10, p<0.05, and p<0.01 respectively

Table 6: Results from binary logistic regression models of shopping-channel choice for *specific shopping occasions* (reference category is *in-store* shopping) (second-stage model)

Model Run	#2S-1	#2S-2	#2S-3	#2S-4	#2S-5
Sample size	521	521	521	521	521
Null log-likelihood	-117.28	-111.62	-111.44	-120.28	-117.28
Final log-likelihood	-51.53	-50.81	-50.33	-54.30	-53.57
McFadden's pseudo-r ²	0.561	0.545	0.548	0.549	0.543
McFadden's adjusted pseudo-r ²	0.467	0.437	0.441	0.449	0.450
<i>Parameter estimates</i>					
Intercept	-6.07*	-1.42	-3.32	0.33	-5.93*
Age (years)	0.05**	0.05***	0.05*	0.05***	0.05**
Respondent is male (binary indicator)	-0.84	-0.69	-0.51	-0.93	-0.66
Respondent is employed (binary indicator)	1.03*	0.92	0.93	1.04*	0.96
Household Income - ln(gross weekly income)	0.06	0.04	0.07	-0.22	0.02
Number of adults in the household	0.43	0.41	0.38	0.66	0.29
Presence of children in the household (binary indicator)	0.6	0.52	0.59	0.49	0.64
Car/van ownership (binary indicator)	-1.96***	-1.93***	-1.852***	-1.81***	-1.78***
Basket cost indicator (<=£25)	-6.10***	-5.91***	-5.266***	-6.38***	-5.82***
Basket cost indicator (>£25 and <= 50)	-3.96***	-4.00***	-3.826***	-4.70***	-3.87***
Basket cost indicator (>=£50)	Fixed at 0.				
Amount paid for online-oriented products: % of total basket cost	5.06***	-	-	-	-
Amount paid for neutral products: % of total basket cost	4.65***	-	-	-	-
Amount paid for in-store-oriented products: % of total basket cost	Fixed at 0.	-	-	-	-
Online-oriented products indicator	-	0.37	-	-	-
Neutral-oriented products indicator	-	0.2	-	-	-

Table 6: Results from binary logistic regression models of shopping-channel choice for *specific shopping occasions* (reference category is *in-store* shopping) (second-stage model)

Model Run	#2S-1	#2S-2	#2S-3	#2S-4	#2S-5
In-store-oriented products indicator	-	-1.27	-	-	-
Basket class indicator: online-oriented only	-	-	Fixed at 0.	-	-
Basket class indicator: in-store oriented only	-	-	-	-	-
Basket class indicator: neutral only	-	-	1.92	-	-
Basket class indicator: online-oriented and in-store oriented only	-	-	-	-	-
Basket class indicator: online-oriented and neutral only	-	-	0.79	-	-
Basket class indicator: neutral and in-store oriented only	-	-	-	-	-
Basket class indicator: all product categories in the basket	-	-	1.01	-	-
# of online-oriented products	-	-	-	0.08*	-
# of neutral products	-	-	-	-0.26***	-
# of in-store-oriented items	-	-	-	-0.02	-
# of online-oriented items - % share of total number of products in the basket	-	-	-	-	5.15**
# of neutral items- % share of total number of products in the basket	-	-	-	-	4.95**
# of in-store-oriented items- % share of total number of products in the basket	-	-	-	-	Fixed at 0.

a. Reference category: in-store shopping

b. Values marked with *, **, or *** are statistically significant at p<0.10, p<0.05, and p<0.01 respectively

Figure 1: Instructions presented to LCF respondents for completing the expenditure diary

Please unfold the side flaps to refer to the notes						
3 Meals, snacks and drinks CONSUMED AWAY FROM HOME						
Include canned drinks, crisps, sweets etc.						
		Where bought	Consumed in or off premises or outside home	Amount paid		
		at restaurant, pub, shop, workplace, place of work, cinema	at	£	p	
Coffee - white		station break	✓	49		
Milk tea		on train	✓	40		
Ham salad sandwich, white bread		work canteen	✓	1 25		
Quavers (crisps)		work canteen	✓	30		
Can of Diet coke		work canteen	✓	50		
Meal for two:						
Lamb, roast potatoes, carrots, peas		restaurant	✓	8 50		
Pizza - ham and mushroom, with green salad		restaurant	✓	7 00		
2x chocolate fudge cake with cream		restaurant	✓	6 00		
2 glasses of coke (not diet)		restaurant	✓	1 50		
Round of drinks: 1 pint bitter, 1 pint lager (£2.50 each)		pub	✓	5 00		
1 pint and lager (£2.50), 1 orange juice (£1)		pub	✓	3 50		
4 Clothing and footwear						
		Name of shop where bought or internet, market, car boot sale etc	Please tick male female (only if under 16)	Age	Amount paid	£ p
Shirt (RETURNED FOR REFUND)		BHS	✓	6 16 44		
Trainers		Dove Sports	✓	49 44		
Shoes		Clarks	✓	29 44		
Jumpers		Internet	✓	34 44		
5 Other payments and purchases today						
		Name of shop where bought		Amount paid		
✓ DO remember to include purchases such as newspapers, cigarettes, stamps, petrol, leisure, National Lottery tickets (state if for Saturday or Wednesday draw).				£	p	
✓ DO include payments for services like childcare, window cleaner etc						
Dinner money (for David)				6 25		
Cloud download				1 00		
Oldbikes		Oldbikes		25 00		
Taxis		ABC cars		5 00		
Admission to cinema x2		Odeon		4 40		
National Lottery Tickets (Saturday) x2		Corner shop		2 00		
Playphone				20		
Bank		Internet		12 44		
6 Winnings from lottery, bingo, betting shops, football pools, raffles etc						
		Amount won		£	p	
National Lottery		10 00				
Betting shop (Horse)		5 23				
Any of today's items to be refunded or reclaimed? → go to Section 9						
p3						

Please attach till receipts when ever possible						
Attaching a till receipt saves you from writing down all the items, but they may not give all the information required						
We need the weight or volume of every food and drink and a full description of it if this includes items which were sold by weight, e.g. a can of beer or a bag of cereal. If you do not have a set of scales to weigh these items yourself your interviewer can give you some.						
The 6 point guide below tells you what kind of descriptions you should write on the till receipt						
1. Bread						
We need to know if bread is sliced or unsliced and whether it is white, wholemeal, soya/rice, etc. You may need to weigh items such as individually sold bread rolls.						
2. Butter, margarine and spreads						
We need to know the brand name of butter (e.g. Lurpak), margarines (e.g. Flora) and reduced or low fat spreads (e.g. Gold Lowest).						
3. Meat						
We need to know if meat is cooked, fresh or frozen and what it is e.g. lamb, pork, chicken. We also need to know if fish is filleted or not.						
4. Milk						
We need to know if milk is semi-skimmed, skimmed or whole milk and whether it is fresh, sterilised or UHT.						
5. Fruit and vegetables						
We need to know if fruit and vegetables are fresh, frozen, tinned or dried, and whether potatoes are old or new.						
6. Drinks (non-alcoholic)						
We need to know if soft drinks are pure fruit juice or fruit drinks and whether they are concentrated (e.g. Kit Kats) or not concentrated and whether low calorie or not.						
Orders made over the Internet						
If you ordered your shopping over the internet please write "internet" on the receipt clearly						
BALANCE DUE 34.34						
AMOUNT TENDERED 40.00						
CHANGE DUE 5.66						
Thank you for shopping with us Please retain for a refund						