Behavioural differences between consumers attracted to shopping online versus traditional supermarkets: implications for enterprise design and marketing strategy

Rick L. Andrews

Ourso College of Business Administration, Louisiana State University,
Baton Rouge, LA 70803, USA
E-mail: andrewsr@lerner.udel.edu

Imran S. Currim*

Graduate School of Management, University of California, Irvine,
CA 92697-3125, USA
E-mail: ISCurrim@uci.edu
*Corresponding author

Abstract: Despite the dot.com shakeout, online revenues continue to increase and are projected to impose greater pressure on traditional distribution channels. However, there is a striking absence of published empirical work on how consumers attracted to shopping online behave relative to consumers shopping in a traditional store. Such behavioural differences, if they exist, could guide online enterprise design and marketing strategy. This study uses data from both traditional supermarket scanners and an online supermarket to test expected differences in choice behaviours of such consumers. For two product categories, statistically significant differences are found between consumers attracted to shopping online versus traditional supermarkets with regard to the parameters describing the choice process. Compared to traditional supermarket consumers, online consumers are less price sensitive, prefer larger sizes to smaller sizes (or at least have weaker preferences for small sizes), have stronger size loyalty, do more screening on the basis of brand names but less screening on the basis of sizes, and have stronger choice set effects. Many of these differences are found to be prevalent among the majority of online consumers rather than due to the substantially unique behaviour of a minority. Indeed, 11 to 39% of traditional supermarket consumers (depending on the product category) are found to behave like the majority of online consumers whilst 0 to 31% of online consumers are found to behave like the majority of traditional supermarket consumers. Implications of both sets of results for online enterprise design, marketing, and evolution are outlined.

Keywords: online shopping; latent class model; logit.

Reference to this paper should be made as follows: Andrews, R.L. and Currim, I.S. (2004) 'Behavioural differences between consumers attracted to shopping online versus traditional supermarkets: implications for enterprise design and marketing strategy', *Int. J. Internet Marketing and Advertising*, Vol. 1, No. 1, pp.38–61.

Biographical notes: Rick L. Andrews is the Robert S. Greer, Snr. Alumni Endowed Chair of Business Administration and Professor in the E.J. Ourso College of Business Administration at Louisiana State University. His 1992 PhD in Marketing is from Virginia Tech. Primary research interests include topics in choice modelling, the analysis of scanner data, and forecasting. His research has appeared in *Journal of Marketing Research, Marketing Science, Journal of Business & Economic Statistics, International Journal of Research in Marketing, Journal of Forecasting*, and *European Journal of Operational Research*, among others.

Imran S. Currim is Corporate Partners Research Scholar and Professor at the Graduate School of Management, University of California, Irvine. He received MS and PhD degrees from Stanford in 1980. He is the recipient of two awards from the American Marketing Association, the William O'Dell Award for five year contribution to research and the AMA/Houghton Mifflin Distinguished Teaching in Marketing Award. His research on customer choice has appeared in *Journal of Marketing Research, Marketing Science, Management Science, Journal of Consumer Research, Journal of Marketing, International Journal of Research in Marketing, Marketing Letters, European Journal of Operational Research, Inquiry*, and *Organizational Behaviour and Human Decision Processes*, among others.

1 Introduction

Despite the dot.com shakeout, online retail revenues (excluding travel, prescription drugs and automobiles) in the USA increased from $12.3 billion in 1999 to $34.1 billion in 2001 and are projected to reach $130.3 billion in 2006 (Jupiter Internet Shopping Model 2001 on www.jmm.com). As a result, it is not surprising that there is increased interest in understanding consumer choice behaviour in computer mediated shopping environments [1,2] and how a new generation of shopping infrastructures can be constructed based on these choice behaviours [3].

For example, consider a manager who seeks to provide a new internet-based grocery shopping and delivery service in a certain geographical area. S/he begins thinking about the design and marketing strategy of the service on aspects such as product assortment (brands and sizes to offer in different product categories), prices, and usage of promotions. S/he ponders whether product assortment, prices, and promotions should be similar to traditional supermarkets in the geographical area or whether these aspects should be different. S/he asks a number of central questions. What kinds of consumers are attracted to shopping for groceries online versus in a traditional supermarket? Are there behavioural differences (on price sensitivity, preference for sizes, brand loyalty, etc.) between consumers attracted to shopping for groceries online versus in a traditional supermarket, and if so, what are the behavioural differences? What are the implications of any such behavioural differences for the design and marketing strategy of the online grocery service? For example, if consumers attracted to shopping for groceries online are likely to be more brand loyal, s/he may be more likely to stock high market share national brands. In contrast, if these consumers were more price sensitive, then s/he may be more likely to stock lower priced alternatives or offer more promotions. Consequently, identifying differences in choice behaviour between consumers attracted to shopping online versus in a traditional store is critical to determining the proper design and marketing strategy of the online shopping environment.

Unfortunately, despite the growth in online sales, and recognition that online sales are likely to impose more pressure on traditional distribution channels, there is still only very limited academic research in this area. Largely due to a lack of usable data, there is a striking absence of published empirical work on what kind of customers, from a behaviourally oriented perspective, are attracted to shopping online versus in a traditional store [4,5]. Many managerial decisions regarding enterprise design and marketing strategy for online as well as traditional stores such as product assortment, pricing, and promotional strategy will rely on whether and how choice behaviours differ in the two environments.

The objective of this study is to establish whether there are behavioural differences between consumers attracted to online shopping and traditional supermarket shopping using actual choice data from an online supermarket and traditional scanner panel data from Information Resources Inc. Logit choice models are fitted to data from the two environments, and attention is focused on whether the descriptions of the choice processes provided by the models differ in the two environments. Given this objective, we address two main research questions. Our first research question is:

RQ1 Do the parameters describing the choice process for online supermarket consumers differ from the parameters describing the choice process for traditional supermarket consumers, and if they do differ, what are the differences?

This research question is addressed by using a procedure for multinomial logit parameter comparisons between different data sets [6]. If the test rejects the null hypothesis of parameter equality, then we would conclude that choice behaviours differ between consumers attracted to shopping online versus in a traditional supermarket.

Though we may reject the null hypothesis that online consumers behave in the same way as traditional store consumers, there could be some segment of online consumers that behaves like traditional consumers. What proportion of online consumers behave like many traditional supermarket consumers and vice versa? Is rejection of the common parameters hypothesis due to behaviours common to many online consumers or to the substantially unique behaviour of fewer online consumers? This is an important question because the online entrepreneur's product assortment, pricing, promotion, and targeting strategies could vary greatly depending on whether 80% of online consumers behave differently from traditional consumers or only 10% of online consumers behave very differently from traditional consumers. Information on such proportions could also provide insight into the future growth potential or evolution of online grocery shopping. For example, the larger the proportion of traditional store consumers behaving like most online consumers, the higher the likelihood that the online service will attract more such consumers over time. Thus, our second research question is:

RQ2 If there are differences in the parameters describing choice behaviour of online and traditional supermarket consumers, do overlapping segments exist? What proportion of online (traditional supermarket) consumers have choice behaviours that are shared with consumers in the traditional (online) supermarket environment?

To address this research question, this study formulates a model that recognises differences in error variance between online and offline consumers [6] and simultaneously allows for heterogeneity in choice behaviours. The choice behaviours of interest in this study are:

- preferences for brand names and sizes
- responses to marketing mix variables such as price and promotion
- state dependence, also known as brand loyalty, which is defined as the effect of previous choice outcomes on current choice behaviour
- choice set effects, defined as the consideration of fewer alternatives than are available at the time of choice.

If these choice behaviours of customers attracted to shopping online versus in traditional supermarkets are truly different, the enterprise design and marketing strategies for the two retailing environments should be tailored to the respective choice behaviours of their consumers. Otherwise, similar design and marketing strategies can be used.

Given the projected hyper-growth in trade of some goods over the internet, it is crucial that businesses understand whether there are differences in choice behaviour between customers attracted to shopping online versus in a traditional store, and if so, the nature of those differences. Understanding differences in preferences and choice sets, the effects of state dependence and marketing mix variables such as price and promotion on choices could be a critical component in businesses' adaptation to internet commerce.

The next section employs the current state of knowledge to develop expectations about differences in choice behaviour between consumers attracted to shopping for groceries online versus in traditional supermarkets. We then describe a choice modelling approach designed to reveal the extent and nature of differences between the two types of consumers, the data used, results, and implications of the results for manufacturers and online and traditional retailers.

2 Expected effects

We focus primarily on the marketing literature because that literature comprises works than focus on differences in consumer behaviour. The Management Information Systems literature, in contrast, comprises works that focus on differences in prices in electronic versus conventional channels [7–10]. This marketing literature is employed to develop general expectations about differences in choice behaviours – price sensitivity, preference for sizes, brand and size loyalty, and consideration sets – between consumers attracted to online versus traditional supermarket shopping.

2.1 Price sensitivity

Purchasing online requires the individual to change his/her behaviour. Behavioural change is difficult and requires incentive, such as explicit monetary savings or an increase in convenience. Firstly, whilst it may be possible to achieve monetary savings by shopping for supermarket products online, we expect that individuals who choose to shop online are likely to do so primarily because of convenience. In contrast to the general population, internet users are characterised as having higher education and income and a greater tendency to hold jobs in upper and middle management or work as trained professionals [11,12]. This suggests a higher opportunity cost of time spent visiting the store, which impacts on the importance they place on convenience and price. Degeratu *et al.* [4] report that consumers shopping online have higher incomes than those shopping in traditional supermarkets and suggest that their higher opportunity cost of time is likely to make them more convenience sensitive and less price sensitive. Bellman, Lohse, and Johnson [2] find that online buyers have "wired lifestyles" (heavy use of the internet for e-mail, work, reading news at home, etc.), are "time starved" (less discretionary time, for example due to a working spouse), and "seek new ways to find information and buy things that are faster and more convenient". Further, if delivery fees are charged and the household agrees to pay such fees it indicates that the household is willing to pay a premium for convenience [13].

Secondly, Alba *et al.* [14] theorise that one key difference between online and offline shopping lies in the ability of the online environment to provide more readily accessible information about price and non-price attributes. More information about price could increase price sensitivity for undifferentiated products. On the other hand more information about non-price attributes could reduce price sensitivity for differentiated products. For example, Peapod allows shoppers almost instantaneously to generate a listing of brands rank ordered on various nutritional aspects. Specifically, in the margarine category consumers can instantaneously rank order products based on calories, fat, and sodium content [15] For non-food products such as detergent, consumers can rank order products based on their popularity. Degeratu *et al.* [4] suggest that the lower search costs online for listed non-price attributes such as nutritional information and popularity may shift consumer focus from price to non-price attributes and as a result lower price sensitivity. Lynch and Ariely [16] use laboratory experiments on the sale of wines through electronic channels to show that providing product information to customers can soften price competition and increase customer loyalty. Shankar, Rangaswamy, and Pusateri [17] use survey data from travellers to show that prior experience with a brand in the physical world can decrease price sensitivity online.

Consequently, our expectation is that price sensitivity for supermarket products will be lower online than in traditional stores. This is particularly likely in the early stages of the life cycle for online shopping. This may change over time as electronic shopping matures and more mainstream consumers adopt the online shopping medium. Further, if and when there is an increase in the number of such online grocery services available to a household, price sensitivity may increase if households compare prices of products offered by alternative online grocery services.

2.2 Size preference

Internet users, who have higher income [11,12], higher opportunity cost of time [4] and are more 'time starved' [2] than the general population, are expected to attempt to reduce time expended acquiring groceries more than the general population. One strategy to reduce time spent acquiring groceries is to reduce the number of orders by purchasing larger sizes. Higher income consumers can also afford to purchase larger sizes. Time starved households may also have more children and hence require larger sizes. Consequently we expect that online purchasers will have a stronger preference for larger sizes. Comparing actual supermarket purchases made by consumers over a seven month period with choice decisions collected in one sitting of a laboratory simulation, Burke *et al.* [18] found that larger sizes are purchased more frequently online.

2.3 Brand and size loyalty and screening

Alba *et al.* [14] theorise that a particular advantage of interactive home shopping over other retail formats is that consumers can more efficiently screen alternatives so that they can focus on alternatives that match their preferences. Interactive home shopping enables such screening almost instantaneously using electronic agents that use information about an individual consumer's specific preferences and the alternatives available [19]. For example, for first time shoppers Peapod offers an 'Express Shop' and 'Personal List'. Consumers enter one or more words, such as the manufacturer or brand name (e.g., Nabisco, Cheerios) and category (e.g., cereal, shampoo) and the electronic agent quickly locates all of the items on the shopping list. For repeat shoppers, Peapod offers 'Previous Orders'. Beginning with the second order, consumers can save time by shopping from a list of their previously ordered items. We expect wired, time starved, convenience driven consumers to use such electronic agents, resulting in greater brand and size loyalty. If more than a single brand is acceptable and certain brands have been purchased frequently, electronic agents that list previous purchases will facilitate screening effects based on acceptable brands, which suggests higher prevalence of brand-based choice set formation online versus offline. Electronic agents may also shift consumer focus from price to non-price attributes such as brand name, resulting in greater brand loyalty or usage of brand name for screening purposes. In addition, less price sensitive consumers (due to having higher income levels) will have their choices determined to a relatively greater extent by non-price attributes such as brand name [5], resulting in greater brand loyalty or usage of brand name for screening purposes.

2.4 How this study is different

Both Lynch and Ariely [16] and Burke *et al.* [18] employ a laboratory approach involving experimentation and simulation. In contrast, our study employs market data on actual purchases made by customers. We view these approaches as complementary. Whilst experiments allow greater control and permit causal explanation for behavioural differences observed, market data on purchases permit investigation of whether behavioural differences actually exist between customers shopping online versus in a traditional store.

Two separate pioneering studies employ Peapod purchase data. Degeratu *et al.* [4] fitted conditional logit models separately to IRI and Peapod data but do not formally test for equivalence of the choice process in the two environments. Degeratu *et al.* [4] focus on comparing the effects of search attributes such as scent and bleach content (liquid detergent) and printed design (paper towels) in online and traditional supermarkets, whereas we focus on comparing choice set effects and state dependence effects in online and traditional supermarkets. Like Degeratu *et al.*, we examine differences in price sensitivity and intrinsic brand preferences in online and traditional supermarkets. Although her proposals call for a comparison of online versus traditional shopping, Zhang [5] reports preliminary results on just Peapod data for two product categories, butter and liquid detergent.

In the next section we propose a latent class model formulation that can be fitted to data pooled from online and traditional shoppers. Within each latent class, online and traditional shoppers are assumed to have common choice behaviours but possibly different error variances in their choices. The allocation of consumers to segments via the model's posterior segment probabilities allows us to investigate the question of what percentage of online (traditional) customers have choice behaviours that are shared with consumers in the traditional (online) environment. No previous research has addressed this research question.

In summary, the research reviewed above suggests that online and traditional supermarket consumers may differ in price sensitivity, size preference, brand and size loyalty (state dependence), and brand screening to form choice sets. These expectations are tested using the model specification described next.

3 Model specification

For the models estimated in this study, the basic utility specification for brand-size k at purchase occasion t (consumer subscripts are suppressed for now) is

$$U_t^k = \alpha_i + \delta_j + x_t^k \cdot \beta + BL_t^i \cdot \beta_{BL} + SL_t^j \cdot \beta_{SL} + \varepsilon_t^k \, , \tag{1}$$

where:

α_i	=	the effect of brand [20] i (i=1,…,I), given brand-size k has brand name i;
δ_j	=	the effect of size j (j=1,….J), given brand-size k has size j;
x_t^k	=	a vector of marketing mix variables, including price and promotion;
β	=	a vector of market response parameters;
BL_t^i	=	brand loyalty of brand i, defined as

$$BL_t^i = \lambda_B BL_{t-1}^i + (1-\lambda_B) \cdot \begin{cases} 1 \text{ if brand } i \text{ was purchased at } t\text{-1;} \\ 0 \text{ otherwise;} \end{cases}$$

λ_B = the brand loyalty smoothing parameter, estimated empirically;

β_{BL} = the brand loyalty coefficient;

SL_t^j = size loyalty of size *j*, defined as

$$SL_t^j = \lambda_S SL_{t-1}^j + (1-\lambda_S)\cdot\begin{cases}1 \text{ if size } j \text{ was purchased at } t\text{-1;}\\ 0 \text{ otherwise;}\end{cases}$$

λ_S = the size loyalty smoothing parameter, estimated empirically;

β_{SL} = the size loyalty coefficient;

ε_t^k = random error term, distributed double exponential.

If there are no choice set effects or parameter heterogeneity, the choice probability for alternative *k* at occasion *t* is formed by

$$P_t(k) = \frac{\exp(U_t^k)}{\sum_{m=1}^{K}\exp(U_t^m)}, \tag{2}$$

where *K* is the number of brand-size alternatives.

There are a variety of ways to incorporate choice set effects into the above model. The method we use, feature-based screening [21], assumes that consumers choose from choice sets created by screening according to brand names, sizes, or current promotions. Consumers may also choose from the full set of *K* available alternatives with no screening. Choice set formation must be probabilistic since we cannot know for certain which choice set a consumer is using. If a consumer only considers previously purchased brands, for example, then the choice probability is

$$\begin{aligned} P_t^B(k) &= \frac{\exp(U_t^k)}{\sum_{m\in\{CS_B\}}\exp(U_t^m)} \quad \text{if } k \in \{CS_B\} \\ &= 0 \text{ otherwise,} \end{aligned} \tag{3}$$

where CS_B contains all alternatives whose brand names have been previously purchased by that consumer. The overall choice probability when there are choice set effects is then

$$P_t(k) = \theta_B P_t^B(k) + \theta_S P_t^S(k) + \theta_P P_t^P(k) + \theta_F P_t^F(k) \tag{4}$$

where the B, S, P, and F super- and sub-scripts indicate choice sets formed by screening on Brands, Sizes, Promotions, and the Full set (no screening), respectively. Since the restriction $\theta_B + \theta_S + \theta_P + \theta_F = 1$ is required, only three of the four parameters need be estimated. The choice set model described by (4) therefore requires only three more parameters than the choice set model described by (2).

The probability of observing consumer c's choice at purchase occasion t is

$$P_{ct} = \sum_k Y_{ct}(k) \cdot P_{ct}(k) \tag{5}$$

where $Y_{ct}(k)$=1 if consumer c purchased brand-size k at purchase occasion t, and 0 otherwise, and a consumer subscript c has been added to $P_t(k)$ in (4). The probability or likelihood of observing consumer c's entire purchase history of T_c purchases is

$$L_c = \prod_{t=1}^{T_c} P_{ct} \ . \tag{6}$$

Given the above model specification, the first research question addresses the issue of whether choice model parameters are different for customers attracted to shopping online (henceforth Peapod data) versus traditional supermarkets (henceforth IRI data). To test formally for differences in the parameters for IRI and Peapod data sets, we use the procedure suggested by Swait and Louviere [6] for comparing parameters of different data sets. Because of scale factor differences between data sets, choice model parameters can appear to be different when they actually are not (the scale factor is inversely related to error variance). Thus, one should allow for differences in scale factor when testing parameter equality across data sets. Essentially, the procedure is as follows:

1 Estimate separate choice models for IRI and Peapod data sets, obtaining the log likelihood values for each (call them L_{IRI} and L_{PEAPOD}).

2 Pool the IRI and Peapod data sets, assuming common parameters but allowing for different scale factors. Only the relative scale factor of the two data sets can be estimated. We fix the scale factor for the IRI data set (μ_{IRI}) to one and empirically estimate the scale factor for the Peapod data set (μ_{PEAPOD}). Obtain the log likelihood value (call it L_μ).

3 The likelihood ratio test statistic $-2[L_\mu - (L_{IRI} + L_{PEAPOD})]$ is distributed χ^2 under the null hypothesis of parameter equality. If the null hypothesis is rejected, there are differences between the choice behaviours of customers attracted to shopping online versus traditional supermarkets.

If there are such differences, the second research question addresses whether there are some traditional supermarket consumers whose behaviours are similar to those of most online consumers (and vice versa), and if so, the sizes of those segments. At one extreme, the differences between the choice behaviours could be driven by the unique behaviours of a small proportion of online consumers. At the other extreme, almost all online consumers could have behaviours different from consumers attracted to traditional supermarkets. An online entrepreneur's product assortment, pricing, promotion, and targeting strategies could vary significantly depending on where reality lies between such extremes.

To address this research question, we fitted models allowing heterogeneity in choice behaviours to pooled online and traditional supermarket consumers and observed the resulting clustering of these consumers into segments. Within each segment, online and traditional supermarket consumers' behaviours differ only by an empirically-estimated scaling constant. Also, independent sets of segment weights are estimated for Peapod and IRI consumers, which reflect the very likely possibility that the allocation of consumers to segments will be different for Peapod and IRI consumers. We are interested in the proportion of online consumers who are allocated to segments containing traditional supermarket consumers and vice versa.

We allow consumer heterogeneity in preferences for brands [22] and sizes, responses to marketing mix [23], brand and size loyalty, and choice set effects using a latent class formulation. The latent class approach assumes that there are latent classes or segments of consumers whose behaviours are described by different sets of parameters. Each segment *s* would have its own $\alpha_i, \delta_j, \beta, \lambda_B, \lambda_S, \beta_{BL}, \beta_{SL}, \theta_B, \theta_S, \theta_P$, and θ_F Since segment membership is unobservable, the overall likelihood of consumer c's behaviour, given that s/he shopped at Peapod, is

$$L_{c|Peapod} = \sum_{s=1}^{S} f_{Peapod}^{s} \cdot L_{c|Peapod}^{s} \tag{7}$$

where f_{Peapod}^{s} is the *a priori* probability that Peapod consumer c belongs to segment s and $L_{c|Peapod}^{s}$ is the likelihood of observing consumer *c*'s purchase history given membership in segment s, found using equation (6). If a consumer c instead shopped at a traditional supermarket and is included in the IRI panel, the likelihood of *c*'s behaviour is

$$L_{c|IRI} = \sum_{s=1}^{S} f_{IRI}^{s} \cdot L_{c|IRI}^{s} \tag{8}$$

and the terms $\left(f_{Peapod}^{s}, f_{IRI}^{s}\right)$ are defined analogously. Notice that the segment weights are estimated separately for Peapod and IRI consumers, and it is possible that some of these values could be zero. This would indicate that consumers from *either* IRI or Peapod (but not both) are assigned to some segment and that the behaviours of those consumers were unique to that environment. As usual, $\sum_{s=1}^{S} f_{Peapod}^{s} = \sum_{s=1}^{S} f_{IRI}^{s} = 1$.

As an example, consider a situation in which a 2-segment latent class model was estimated for pooled IRI and Peapod data. If the purchase behaviours of IRI and Peapod consumers were radically different, we might observe that $f_{Peapod}^{1} = 1,\ f_{Peapod}^{2} = 0,\ f_{IRI}^{1} = 0,$ and $f_{IRI}^{2} = 1$. This means that all the Peapod consumers are allocated to segment one, and all the IRI consumers are allocated to segment two.

Later, we call this a 1x1 (IRI x Peapod) model because there is really only one segment each for IRI and Peapod consumers. If half of the Peapod consumers are like traditional IRI consumers, and all IRI consumers are homogeneous, then we might observe $f^1_{Peapod} = .5$, $f^2_{Peapod} = .5$, $f^1_{IRI} = 0$, and $f^2_{IRI} = 1$. We would call this a 1x2 model.

The overall log likelihood function maximised to estimate the parameters is

$$L = \sum_c \delta^c_{IRI} \ln\left(L_{c|IRI}\right) + (1-\delta^c_{IRI}) \ln\left(L_{c|Peapod}\right) \tag{9}$$

where $\delta^c_{IRI} = 1$ if consumer c shopped at a traditional IRI supermarket and 0 otherwise.

Once the parameters are estimated, posterior probabilities are used to assign consumers to segments. The posterior probability that IRI consumer c is a member of segment s is given by

$$p\left(c \in s \middle| IRI\right) = \frac{f^s_{IRI} \cdot L^s_{c|IRI}}{\sum_{m=1}^{S} f^m_{IRI} \cdot L^m_{c|IRI}}, \tag{10}$$

with the posterior probabilities for Peapod consumers computed analogously. The consumer is assigned to the segment with the highest posterior assignment probability.

To summarise, in contrast to other applications, the latent class model proposed here is fitted to pooled Peapod and IRI data, allowing for different within-segment error variances for Peapod and IRI consumers. With this model, we can determine what proportion of Peapod (IRI) consumers behave like IRI (Peapod) consumers.

4 Data

For this study, we use two of the most comprehensive data sets currently available, which have also been employed in earlier research [4,5].

The first data set is from Peapod, Inc., an internet grocer based in Skokie, Illinois, wherein purchases of 279 subscribers in a Chicago suburban area are tracked from May 1996 to July 1997. The second data set is from IRI and is composed of the purchases of 1,614 panellists over the 112 week period from September 1995 to November 1997. These panellists shopped in the same grocery chain and in the same geographical area, although not in the same supermarket as the Peapod subscribers (the supermarket from which Peapod customers are served during the period is not included in the sample of supermarkets that IRI employs to collect panel data). Two of the three stores included in the IRI data are located in the same part of the metropolitan area as the Peapod store and all three stores are in relatively affluent areas.

The analysis in this study focuses on two product categories: liquid laundry detergent and margarine. For the liquid laundry detergent category, we use all Peapod consumers making laundry detergent purchases, 148 in number. These consumers account for 1,034 purchases. Our goal was to keep the number of IRI and Peapod consumers approximately equal. Random selection of IRI consumers produced 147 consumers making 1,579 observations. The fact that IRI consumers make more purchases than Peapod consumers is not surprising, given that the data collection period for IRI is about one year longer than that for Peapod.

Brand names and sizes with 3% or greater market share were retained. The brand names retained were Ajax, All, Arm & Hammer, Cheer, Era, Surf, Tide, Wisk, and Yes, whilst the sizes (in ounces) were 50, 64 90, 100, 128, 150 (Peapod only), and 200. The IRI data has store feature and aisle display variables available as promotion indicator variables, whilst the Peapod data has a single promotion indicator variable that takes a value of zero when the item is not on sale and a value of one when the item is on sale. In the online shopping environment, the promotion indicator 'P' informed the shopper that the item was on sale but did not provide the amount of the price discount. We do not constrain the effects of promotions within segments to be the same for IRI and Peapod data since the promotion variables are not the same (there is obviously no aisle display in an electronic shopping environment). The prices for IRI and Peapod data are standardised so that they have the same scales. Thus, within segments, we can constrain the price coefficients to be the same for IRI and Peapod consumers. Brand and size loyalty variables are created from the data, as in the pioneering work of Guadagni and Little [24].

Demographic or lifestyle data is not included in the models since such data for online grocery consumers is currently incomplete or does not exist. Given the behaviour-based segments derived in our study, one could easily profile the segments according to demographic (income, education, etc.) and lifestyle (internet usage, whether time starved, etc.) characteristics in order to identify such differences.

For the margarine category, only 110 Peapod consumers made purchases, resulting in 808 observations. The randomly-selected sample of 110 IRI consumers made 2,308 purchases. The brand names retained include Blue Bonnet (IRI only), Brummel & Brown, Fleischman, I Can't Believe It's Not Butter, Imperial, Land O'Lakes, Parkay, Promise, and Shedd's Country Crock. The sizes retained (in ounces) are 16 and 48. The price, promotion, and loyalty variables are defined exactly as they were for the liquid laundry detergent data.

In summary, there is good overlap between the time periods represented in both Peapod and IRI data. Across both product categories almost all brands and sizes (17 of 18 brands and 8 of 9 sizes) are represented in both online and traditional store environments. Because the types of promotions in the two environments vary (e.g., there is no aisle display in an electronic shopping environment), we do not constrain the effects of promotions within segments to be the same for IRI and Peapod data. However, we do correct for these differences by including promotions in the model specifications.

6 Results

6.1 Tests for common model parameters

Research question 1 asks whether the parameters describing the choice process for online supermarket consumers differ from the parameters describing the choice process for traditional supermarket consumers. The results of the test for common model parameters for the laundry detergent category are shown below:

H_0: $\beta_{IRI}=\beta_{PEAPOD}$

$L_{IRI}=-2701$

$L_{PEAPOD}=-1681$

$L_{\mu}=-4485$ ($\mu_{IRI}=1.00$, $\mu_{PEAPOD}=1.21$)

$\chi^2=-2[L_{\mu}-(L_{IRI}+L_{PEAPOD})]=-2[-4485-(-2701-1681)]=206$, d.f.=13, $p<0.0001$

Thus, the null hypothesis is rejected for the laundry detergent category. For the margarine category, the test for common model parameters shows the following:

H_0: $\beta_{IRI}=\beta_{PEAPOD}$

$L_{IRI}=-2414$

$L_{PEAPOD}=-983$

$L_{\mu}= -3409$ ($\mu_{IRI}=1.00$, $\mu_{PEAPOD}=0.58$)

$\chi^2=-2[L_{\mu}-(L_{IRI}+L_{PEAPOD})]=-2[-3409 -(-2414-983)]=24$, d.f.=10, $p<0.01$

For both product categories, the null hypothesis that choice behaviours of consumers attracted to shopping online versus traditional supermarkets are the same is rejected, though it is more overwhelmingly rejected in the liquid detergent product category. We note that the scalability restriction (i.e., that IRI parameters are different from Peapod parameters only by a constant) cannot be imposed on six parameters (three choice set probabilities, two smoothing parameters for loyalty variables, and the promotion coefficient).

In the following sections, we present detailed analysis of the differences between customers attracted to shopping online versus traditional supermarkets, firstly for laundry detergent, then for margarine.

6.2 Laundry detergent results

Table 1 shows the parameter estimates for independent choice models for the Peapod and IRI data sets. The parameter estimates for Peapod have been adjusted by the scale factor from the test of parameter equality (1.21) to make them comparable to the IRI parameter estimates.

Peapod consumers prefer All and Ajax, whilst IRI consumers prefer Tide and Wisk. With regard to sizes, Peapod consumers prefer the largest sizes (150 oz., and 200 oz.), whilst IRI consumers prefer the smallest size (50 oz.), which is consistent with expectations developed earlier.

Table 1 Parameters for independent choice models for Peapod and IRI data – laundry detergent data (t-statistics in parentheses)

Parameter Estimates	*Peapod*[b]		*IRI*	
Brand effects:				
Ajax	2.2367	(4.824)	-0.1935	(-0.626)
All	2.8655	(11.119)	0.5061	(1.793)
Arm & Hammer	-0.8014	(-2.092)	-0.3908	(-1.886)
Cheer	0.6154	(2.884)	2.0037	(9.397)
Era	-1.0136	(-1.705)	0.6704	(3.695)
Surf	-0.7999	(-1.402)	1.7202	(7.938)
Tide	0.1509	(1.137)	2.8992	(12.737)
Wisk	-0.0944	(-0.578)	2.6057	(11.269)
Yes	-0.0000-		-0.0000-	
Size Effects:				
50 oz.	-0.2755	(-1.822)	2.4464	(11.728)
64 oz.	-2.0352	(-6.721)	1.6322	(6.470)
90 oz.	-0.4114	(-3.005)	0.6849	(2.857)
100 oz.	-0.2930	(-2.972)	1.6311	(8.616)
128 oz.	-1.3678	(-5.729)	1.5260	(6.069)
150 oz.	-0.1987	(-1.052)	--	
200 oz.	-0.0000-		-0.0000-	
Price	--[c]		-1.3409	(-14.785)
Store feature (IRI)	--		0.5328	(4.023)
Aisle display (IRI)	--		0.4974	(3.767)
Promotion (Peapod)	0.6759	(5.400)	--	
λ_{Brand}	0.7297[a]		0.7511[a]	
λ_{Size}	0.6975[a]		0.5965[a]	
Brand loyalty	2.3010	(12.334)	3.3761	(29.527)
Size loyalty	3.2706	(26.998)	1.1852	(7.282)
Choice set effects:				
Brand screening	0.7617[a]		0.0000[a]	
Size screening	0.0000[a]		0.5096[a]	
Promotion screening	0.0264[a]		0.1462[a]	
No screening	0.2119[a]		0.3442[a]	
Log likelihood	-1681		-2701	
Descriptive Statistics				
# Purchases/year	6.01		4.95	
Ave. size purchased	105 oz.		101 oz.	
Ave. price paid/oz.	$0.0627		$0.0483	
Purchases on promotion	25.44%		42.94%	
Ave. brands purchased	1.36		3.23	

[a] Transformed from estimated parameter, so t-statistic does not test any meaningful hypothesis

[b] Peapod parameters adjusted for scale difference so that they are comparable to IRI parameters

[c] Price is highly correlated with promotion in the Peapod data, so it was necessary to remove the price variable in order to obtain reasonable parameter estimates

The price coefficient was removed from the Peapod model because collinearity between price and promotion resulted in counterintuitive results. For 20 of 38 brand sizes analysed in the Peapod laundry data, the correlations between price and promotion are dramatically high (greater than 0.7), whilst in the IRI data set none of the correlations is that high. Thus, we removed the price variable to obtain credible results.

The promotion coefficients are positive and significant for both models. The smoothing parameters λ for the loyalty variables are quite close for the IRI and Peapod data sets. The coefficients for the loyalty variables are not as similar, however. The Peapod consumers have a higher size loyalty coefficient, whilst the IRI consumers have a higher brand loyalty coefficient.

As per expectations developed earlier, Peapod consumers screen using brand names (i.e., consider only brands they have purchased before) more than IRI consumers do, screening on about 76% of purchase occasions. IRI consumers screen using sizes on nearly 51% of purchase occasions. One could interpret Peapod consumers' 76% screening probability for brands as being roughly equivalent to an infinitely large brand loyalty coefficient for 76% of Peapod consumers. Likewise, 51% of IRI consumers have the rough equivalent of an infinite size loyalty coefficient. IRI consumers tend to screen on the basis of promotions more frequently than Peapod consumers do (15% versus 3%).

The descriptive statistics at the bottom of Table 1 support and augment the interpretation of the models. The Peapod consumers are heavier users who buy more frequently, and they also buy more per purchase occasion, on average. They pay more per ounce and make fewer purchases on promotion, consistent with the finding that they are less price sensitive. Peapod consumers buy only 1.36 different brands, on average, whereas IRI consumers buy 3.23 brands. This finding is consistent with Peapod consumers' more prevalent formation of choice sets on the basis of brand names.

Research question 1 asks whether the parameters describing the choice process for consumers attracted to shopping online versus traditional supermarkets are different. For the laundry detergent category, there are significant differences between IRI and Peapod data sets in terms of preferences for brands and sizes, responses to marketing mix variables, state dependence, and choice set effects.

Research question 2 asks what proportion of traditional supermarket consumers have choice behaviours that are shared with most online supermarket consumers and vice versa. This question provides insight into whether rejection of the common parameters hypothesis is due to behaviours common to most online consumers or due to the substantially unique behaviour of fewer online consumers.

We address this research question by fitting the latent class model to pooled IRI and Peapod data, restricting behaviours of IRI and Peapod consumers to be the same within segments, but allowing for different numbers of segments as well as different error variances (scale factors) within segments for IRI and Peapod consumers. This procedure produced a latent class model with TWO segments for both IRI and Peapod consumers. We summarise the general estimation results for laundry detergent as follows:

IRI x Peapod Segments	*Log1*	*P*	*BIC*
1x1	-4501	31	9246
2x2	-4224	54	8872
1x2	-4280	50	8954
1x3	-4247	74	9077

Logl =Value of maximized log likelihood function
P =number of parameters required
BIC =-2 Logl+P ln(n), where n=sample size

We began by estimating a model with one segment for both IRI and Peapod data sets (1x1), in which IRI and Peapod parameters are different only by a scale factor. [25] A model with two segments for each of the IRI and Peapod data sets (2x2) is much improved according to BIC. Within each segment, IRI and Peapod parameters differ only by a scale factor. As we will see, there does not appear to be a strong need for two IRI segments in this model since only 11% of IRI consumers are assigned to one of the segments – perhaps one segment is sufficient for IRI consumers but not Peapod consumers. Thus, we also estimate a 1x2 model and even a 1x3 model, but neither is preferred over the 2x2 model. A 3x3 model (results not reported) was also estimated, but the extra parameters were not justified according to BIC. Note that the fit of the 2x2 model is better than the combined fit of the independent logit choice models fitted separately to IRI and Peapod data sets (Table 1). This is because the independent models assume that all IRI consumers are alike and that all Peapod consumers are alike, whereas the 2x2 model allows for overlapping segments – it acknowledges that some Peapod consumers behave like some IRI consumers. The 2x2 model represents an alternative overlapping segmentation scheme that recognises common choice behaviours by the two types of consumers.

Consumers are assigned to segments in the laundry detergent category as follows:

	Peapod Consumers	*IRI Consumers*
Segment 1	69%	11%
Segment 2	31%	89%

The consumer allocation Table above shows that 89% of IRI consumers are assigned to the second segment, with only 11% assigned to the first segment, which is indicative of only weak heterogeneity in IRI consumer behaviour. However, the Peapod consumers are allocated to the two segments in a 69%-31% split, suggesting greater heterogeneity. If behaviours of consumers attracted to shopping online versus traditional supermarkets were *extremely* different, we would observe nearly 100% of IRI consumers assigned to one segment, with nearly 100% of Peapod consumers assigned to the other. The fit would be the same as that of the combined models in Table 1. If behaviours of consumers attracted to shopping online versus traditional supermarkets were extremely homogeneous, we may observe the same percentage of IRI consumers and Peapod consumers assigned to any given segment. Our findings suggest that about 31% of Peapod consumers behave like the majority of IRI consumers, with the other 69% behaving differently (preference for larger sizes, stronger size loyalty, etc.). Thus many of the differences between customers attracted to shopping online versus in traditional supermarkets (e.g., online consumers' preference for larger sizes, higher size loyalty) are driven by a majority of online consumers (69%) rather than the substantially unique behaviour of a minority.

6.3 Margarine results

Table 2 shows the parameter estimates for independent choice models for IRI and Peapod data sets. The parameter estimates for Peapod have been adjusted by the scale factor from the test of common parameters (0.58) to make them comparable to the IRI parameter estimates. The smaller scale factor for Peapod data indicates that the error variance is higher in the Peapod data than in the IRI data. This was not the case in the laundry detergent data, with the scale factor value of 1.21 indicating lower error variance in the Peapod data. Thus, there is no systematic difference in the error variation in choices across Peapod and IRI datasets.

Table 2 Parameters for independent choice models for Peapod and IRI data – margarine data (t-statistics in parentheses)

Parameter Estimates	*Peapod*[b]		*IRI*	
Brand effects:				
Blue Bonnet	--		-1.2202	(-8.095)
Brummel & Brown	3.2316	(2.946)	1.3486	(7.900)
Fleischman	2.9740	(2.828)	1.4372	(8.678)
I Can't Believe It's	2.8366	(2.911)	2.3981	(16.660)
Imperial	2.6947	(2.062)	-0.7949	(-6.570)
Land O' Lakes	1.7703	(1.489)	0.9152	(6.020)
Parkay	3.3277	(1.935)	0.2208	(1.481)
Promise	2.3154	(2.220)	1.6988	(11.797)
Shedd's Country Crock	-0.0000-		-0.0000-	
Size Effects:				
16 oz.	0.9301	(1.148)	2.1850	(10.214)
48 oz.	-0.0000-		-0.0000-	
Price	-1.2391	(-4.588)	-1.7821	(-23.581)
Store feature (IRI)	--		-0.0020	(-0.029)
Aisle display (IRI)	--		0.6242	(5.938)
Promotion (Peapod)	--		--	
λ_{Brand}	0.6984[a]		0.7700[a]	
λ_{Size}	0.7028[a]		0.7416[a]	
Brand loyalty	6.3523	(4.920)	3.1897	(22.119)
Size loyalty	5.3876	(3.978)	3.0716	(10.682)
Choice set effects:				
Brand screening	0.8666[a]		0.5792[a]	
Size screening	0.0001[a]		0.1023[a]	
Promotion screening	0.0000[a]		0.0000[a]	
No screening	0.1333[a]		0.3185[a]	
Log likelihood	-983		-2414	
Descriptive Statistics				
# Purchases/year	6.31		9.74	
Ave. size purchased	17.07 oz.		17.45 oz.	
Ave. price paid/oz.	$0.0908		$0.0603	
Purchases on promotion	18.44%		24.39%	
Ave. brands purchased	1.23		3.31	

[a] Transformed from estimated parameter, so t-statistic does not test any meaningful hypothesis

[b] Peapod parameters adjusted for scale difference so that they are comparable to IRI parameters

[c] Price is highly correlated with promotion in the Peapod data, so it was necessary to remove the promotion variable in order to obtain reasonable parameter estimates

Both Peapod and IRI consumers have strong preferences for I Can't Believe It's Not Butter, though Peapod consumers have slightly stronger preferences for Brummel & Brown. Both types of consumers have highly significant preferences for Promise as well. Though Peapod consumers' preferences for the small size are not statistically different from their preferences for the large size, IRI consumers prefer the small size of margarine to the large one. This is consistent with expectations and the laundry detergent results.

Also consistent with expectations, Peapod consumers are less price sensitive than IRI consumers. The aisle display coefficient is positive and significant for the IRI data, but the promotion variable was removed from the Peapod model because of the same collinearity that plagued the Peapod laundry detergent data. There are fewer temporal price variations in the Peapod data than in the IRI data – the prices typically change only when there is a promotion, making it difficult to disentangle the effects of price and promotion.

Like the results for laundry detergent, the smoothing parameters for loyalty variables are very similar for Peapod and IRI data sets. Brand and size loyalty are strong for Peapod and IRI data sets, with both loyalty coefficients larger for Peapod consumers. In the Peapod data, over 86% of consumers screen using brand names, implying an infinite brand loyalty coefficient for these consumers. In the IRI data, 58% screen using brand names. As expected, and consistent with the laundry detergent results, brand screening is more prevalent with the Peapod data than with the IRI data. Also consistent with the laundry detergent results, size screening is more prevalent with the IRI data than with the Peapod data, with about 10% of IRI consumers screening on the basis of sizes. Neither the online nor the traditional consumers screen on the basis of promotions, which could be due to the way that promotion-based screening is operationalised. Promotion-based screening occurs when consumers consider only the promoted brands. Perhaps consumers do consider the promoted brands, but in addition to other favourite brands. Promotion-based screening was also uncommon in the laundry detergent category (3% for Peapod, 15% for IRI).

Descriptive statistics for the margarine category appear at the bottom of Table 2. Though Peapod consumers make about the same number of purchases per year in the margarine category as they do in the laundry detergent category (6.01 versus 6.31), IRI consumers make substantially more margarine purchases (9.74/yr.). Perhaps Peapod consumers cannot stock up on margarine since it is perishable, and they make fill-in purchases at local supermarkets between major Peapod orders. Consistent with the laundry detergent findings, Peapod consumers pay more for margarine and make fewer purchases on promotion. Also consistent with the laundry detergent category is the finding that Peapod consumers purchase far fewer different brands (1.23 versus 3.31 for IRI consumers), most likely the result of Peapod consumers' stronger brand screening to form choice sets.

Overall, there are significant differences between IRI and Peapod consumers' margarine category behaviours, but not as many differences as for the laundry detergent category. For example, the brand preference coefficients are much more similar (strong preferences for I Can't Believe It's Not Butter, Promise), as is brand screening to form choice sets. This finding is also consistent with the tests for common β parameters – the test solidly rejected the null hypothesis of common parameters for the laundry detergent category ($p < .0001$) but much less so for the margarine category ($p < .01$). Thus, there appear to be greater differences between IRI and Peapod consumer behaviours in the laundry detergent category than in the margarine category.

The segmentation results for the pooled IRI and Peapod margarine data are shown below:

IRI x Peapod Segments	*Log1*	*P*	*BIC*
1x1	-3401	26	7006
2x2	-3133	44	6613
2x1	-3145	41	6612
3x1	-3102	61	6683

A model with two segments for IRI consumers and one segment for Peapod consumers is optimal, though it is only ever so slightly better than the model with two segments for both IRI and Peapod consumers. We also fitted a 3x1 model, but it is not preferred over the 2x1 model.

The allocation of consumers by the 2x1 model is as follows:

	Peapod Consumers	*IRI Consumers*
Segment 1	100%	39%
Segment 2	0%	61%

IRI consumers are assigned to the two segments in a 39%-61% split, with all Peapod consumers assigned to the first segment. This indicates that many of the differences between customers attracted to shopping online versus traditional supermarkets (e.g., lower price sensitivity, higher brand but lower size screening) are driven by a majority of online consumers rather than the substantially unique behaviour of a minority. In addition, across the two product categories studied, between 11 and 39% of traditional supermarket consumers are found to behave like the majority of online consumers whilst 0 to 31% of online consumers are found to behave like the majority of traditional supermarket consumers. These proportions indicate that given the much larger number of traditional supermarket consumers in the country today, traditional supermarkets may be more at risk of losing their customers to online shopping than online supermarkets may be of losing their customers back to traditional supermarkets.

7 Summary of results

We summarise the results from both categories as follows:

1 For both categories, there are statistically significant differences between consumers attracted to shopping online versus in traditional stores with regard to the parameters describing the choice process. However, there are more similarities between choice behaviours of the two types of consumers in the margarine category than in the laundry detergent category.

2 Compared to traditional supermarket consumers, online consumers:
- are less price sensitive
- prefer larger sizes to smaller sizes
- do more screening on the basis of brand names
- do less screening on the basis of sizes
- have stronger choice set effects (i.e., a lower percentage of customers who do no screening).

3 Many of the differences identified in 2 above are driven by the majority of online and traditional supermarket consumers rather than the substantially unique behaviour of a minority. Across the two product categories studied, between 11 and 39 % of traditional supermarket consumers are found to behave like the majority of online consumers whilst 0 to 31% of online consumers are found to behave like the majority of traditional supermarket consumers.

4 There are no systematic differences between the IRI and Peapod environments in terms of the amount of noise or error variance in observed choices. For laundry detergent purchases, the Peapod data had lower error variance than the IRI data, but for margarine purchases, Peapod data had higher error variance than the IRI data.

8 Implications and future research

Despite the growth in online sales, and projections that online shopping environments are likely to impose more pressure on traditional distribution channels, there is a striking absence of published empirical work on how consumers attracted to shopping online behave relative to consumers attracted to shopping in traditional stores. The answer to this question guides online decisions on product assortment, prices, promotions, and targeting. We seek to establish whether there are behavioural differences between consumers attracted to shopping online versus in traditional supermarkets.

For both product categories studied, the common parameters hypothesis for online and traditional consumers is rejected. This finding suggests that enterprise design and marketing strategies perhaps should differ in the two purchasing environments, depending on whether the parameter differences are the result of the behaviours of most online consumers or only a small minority of online consumers. Across the two product categories studied, between 11 and 39% of consumers shopping in traditional supermarkets are found to behave like the majority of online consumers. Online supermarkets may want to target this minority of traditional supermarket consumers whose behaviours are different from those of the majority of traditional supermarket shoppers since they may be more likely to switch to the online shopping environment. Online grocery services could analyse the personal characteristics (e.g., household income, occupation, internet use, whether one spouse is not employed outside the home, time starvation, etc.) of this subset of consumers shopping in traditional supermarkets and target this segment with efforts (e.g., to increase awareness of the online service, convenience features such as electronic agents which facilitate buying of standard or repeat purchase items, and availability of larger sizes of popular brands) to induce them to switch to the online grocery service. Given the large number of traditional supermarket consumers today, a proportion between 11 and 39% of such consumers behaving like the majority of online consumers is indicative of good potential for growth of the online grocery service.

Likewise, 0 to 31% of online consumers across the two product categories studied are found to behave like the majority of traditional supermarket consumers. Traditional supermarkets may want to target such online consumers whose behaviours are very similar to those of their regular customers since very little behavioural change is required for these consumers to switch back and forth between the two environments. They could employ appeals based on lower prices for smaller sizes. Because of substantial differences in the current sizes of the two populations of customers attracted to shopping online versus traditional supermarkets, there will be large asymmetries in the potential sizes of customer groups that retailers can hope to switch from one environment to the other, though there is some similarity in the proportion overlap (11-39% and 0-31%).

Notwithstanding infrastructure costs of online versus traditional stores, the finding that online supermarket customers prefer larger sizes and are less price sensitive implies potential for higher profits for manufacturers and retailers of such products. A preference for larger sizes could be due to higher consumption rates, building inventory because of higher importance for convenience, or minimising delivery costs. If consumption rates are higher, then the fact that online heavy users of grocery items are less price sensitive potentially means higher expectations of profits for manufacturers and retailers marketing supermarket products and services online. This is more likely in earlier stages of the life cycle for such services, which are characterised by absence of online competitors for a certain geographical region. In later stages, with online consumers becoming more mainstream and several online firms for each geographical region competing on price, promotion, etc., then it is possible that online consumers will mirror their traditional store counterparts.

If a preference for larger sizes is due to building inventory for convenience purposes, then the presence of such inventory results in online consumers being less susceptible to the marketing practices of other manufacturers and retailers, resulting in potentially even higher expected profits. The finding that online consumers have lower price sensitivity indicates that these consumers may place more importance on convenience. The online supermarket product retailer may wish to offer higher service levels to reward and keep such consumers, such as electronic confirmation of orders received, quality control on accuracy of orders filled, guaranteed undamaged item packages, careful bagging of items ordered, on-time delivery, electronic communications thanking customers for orders placed, and elicitation of problems encountered for quick resolution.

The finding that online grocery consumers do more screening on the basis of brand names implies greater potential profitability for large share, reputable brands. Since these brands are more likely to have a larger number of loyal consumers [26], these brands will be able to leverage electronic agents that remember and recall past purchases to make their loyal consumers less vulnerable to the promotion of competitors. As a result competitive brands may find it more difficult to achieve even short-term incremental sales from promotion. Of course this is likely to change if online grocery shopping attracts more price sensitive consumers over time. However, handling and delivery fees may inhibit such customers from shopping online. Brand loyalty could also be affected by electronic agents being able to search for the lowest price or for products currently on promotion. In contrast brand loyalty may be enhanced by use of electronic agents to order or sort products based on nutritional information. Consequently, it may be interesting to run an experiment with online customers to study the effects of different types of electronic search on brand loyalty, price and promotion sensitivity, and choice set effects.

The finding that online customers engage in (more prevalent) brand screening to form choice sets indicates that it is important for manufacturers to get their products purchased by new customers. An alternative to promotions is to use targeted advertising to specific online customers to build brand preference based on differentiation. Advertising in the online context can take the form of internet links to the manufacturer's website, where additional information can be presented to differentiate the brand from other major brands in the market in a way that the customer is persuaded to purchase the product for trial purposes. It may be interesting to run experiments with online consumers to study the effects of such communications on brand loyalty, price and promotion sensitivity, and choice set effects. Web advertising communication could be combined with virtual point of sale displays and targeting technology to customise products and promotions and to determine the appropriate mix of online and offline marketing to build sales. Study of choice sets across online consumers can also provide input into optimal online product assortments.

9 Summary and limitations

This study uses data from both traditional supermarket scanners and an online supermarket to test for expected differences in the parameters describing the choice process of consumers attracted to shopping online versus traditional supermarkets. Results of analyses conducted for two product categories indicate that compared to traditional supermarket consumers, online consumers are less price sensitive, prefer larger sizes to smaller sizes (or at least have weaker preferences for small sizes), have stronger size loyalty, do more screening on the basis of brand names but less screening on the basis of sizes, and have stronger choice set effects. Many of these differences are found to be prevalent among the majority of online consumers rather than due to the substantially unique behaviour of a minority. One potential limitation is that the data collected during 1996-97 may reflect the behaviour of earlier adopters. Future research needs to ascertain the extent to which such behaviour is representative of the behaviour of later adopters. The managerial implication, at least for earlier adopters, is that online retailers should focus on the convenience needs of such customers by stocking the high market share national brands and the larger sizes that they prefer, and offering convenience in ordering and delivery. In addition, 11 to 39% of traditional supermarket consumers (depending on the product category) are found to behave like the majority of online consumers, whilst 0 to 31% of online consumers are found to behave like the majority of traditional supermarket consumers. These results suggest that traditional supermarkets may be able to win back customers who have migrated to making at least some purchases online by offering internet based ordering and pick-up or delivery of purchased items. These traditional supermarkets may be also able to keep customers who have a propensity to become online shoppers by offering price promotions on smaller sizes at their store-based websites.

Acknowledgements

The authors thank Professor Arvind Rangaswamy at Pennsylvania State University for transferring the data files, Marci Thompson and Sally Martin of Peapod, Inc., and Rob Stevens of IRI, Inc. for extracting, organising and allowing the use of the data sets, and Professors Donna Hoffman, Aradhna Krishna, and Rajeev Tyagi at Vanderbilt University, University of Michigan, and University of California, Irvine, respectively, for assistance and comments. The first author gratefully acknowledges a research grant from the University of Delaware. The authors contributed equally to the article and are listed in alphabetical order.

References and Notes

1 Hoffman, D.L. and Novak, T.P. (1996) 'Marketing in hypermedia computer-mediated environments: conceptual foundations', *Journal of Marketing*, Vol. 60, No. 3, pp.50–68.

2 Bellman, S., Lohse, G.L. and Johnson, E.J. (1999) 'Predictors of online buying behavior', *Communications of the ACM*, Vol. 42, No. 12, December, pp.32–38.

3 Baty, J.B. and Lee, R.M. (1995) 'Intershop: enhancing the vendor/customer dialectic in electronic shopping', *Journal of Management Information Systems*, Vol. 11, No. 4, pp.9–31.

4 Degeratu, A., Rangaswamy, A. and Wu, J. (1999) 'Consumer choice behavior in online and traditional supermarkets: the effects of brand name, price, and other search attributes', *Working Paper*, Smeal College of Business Administration, Pennsylvania State University.

5 Zhang, J. (1998) *Investigating the Dynamic Brand Choice Process: A Comparison of On-line Shopping Environments and Store Shopping Environments*, Marketing Department, Kellogg Graduate School of Management, Northwestern University, October.

6 Swait, J. and Louviere, J. (1993) 'The role of the scale parameter in the estimation and comparison of multinomial logit models', *Journal of Marketing Research*, Vol. 30, August, pp.305–314.

7 Lee, H.G. (1997) 'Do electronic marketplaces lower the price of goods?', *Communications of the ACM*, Vol. 41, January, No. 12.

8 Bailey, J.P. (1998) 'Intermediation and electronic markets: aggregation and pricing in internet commerce', *PhD Thesis, Technology, Management, and Policy*, Massachusetts Institute of Technology, Cambridge, MA.

9 Clemons, E.K., Hann, I-H. and Hitt, L. (1998) 'The nature of competition in electronic markets: an empirical investigation of online travel agent offerings', *Working Paper*, The Wharton School of the University of Pennsylvania, June.

10 Brynjolfsson, E. and Smith, M. (1999) 'Frictionless commerce? A comparison of internet and conventional retailers', *Working Paper*, MIT Sloan School of Management, Cambridge, MA.

11 Kehoe, C., Pitkow, J. and Rogers, J. (1998) *GVU's Ninth WWW User Survey Report*., Office of Technology Licensing, Georgia Tech Research Corporation, Atlanta.

12 Hoffman, D.L., Kalsbeek, W.D. and Novak, T.P. (1996) 'Internet and web use in the United States: baselines for commercial development', Special Section on 'Internet in the Home', *Communications of the ACM*, December.

13 Peapod's delivery charge varies by market. Most typically, it offers the customer a choice between three price plans: 1) $19.95 per month, unlimited orders, 2) $9.95 per order or 3) $4.95 per month plus $4.95 per order. Customers can choose the plan that best meets their usage patterns. To promote convenience (i) delivery time is selected by the consumer and (ii) in certain cities Peapod offers unattended delivery service.

14 Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A. and Wood, S. (1997) 'Interactive home shopping: consumer, retailer, and manufacturer incentives to participate in electronic marketplaces', *Journal of Marketing*, Vol. 61, July, pp.38–53.

15 Whilst some traditional supermarkets may arrange their dairy cases so that reduced fat items are separated from their regular counterparts, items also vary on sodium content. Whilst fat and calories may be correlated in margarine products, consumers may not be aware of such correlations especially due to the number of high calorie products which are labelled and advertised as non-fat.

16 Lynch, Jr., J.G. and Ariely, D. (1998) 'Interactive home shopping: effects of search goods for price and quality information on consumer price sensitivity, satisfaction with merchandise, and retention', *Presented at Marketing Science and the Internet*, INFORM College of Marketing Mini-Conference, Cambridge, MA, 6-8 March.

17 Shankar, V., Rangaswamy, A. and Pusateri, M. (1998) 'The impact of internet marketing on price sensitivity and price competition', *Presented at Marketing Science and the Internet*, INFORM College of Marketing Mini-Conference, Cambridge, MA, 6-8 March.

18 Burke, R.R., Harlam, B.A., Kahn, B. and Lodish, L.M. (1992) 'Comparing dynamic consumer choice in real and computer simulated environments', *Journal of Consumer Research*, June, Vol. 19, No. 1, pp.71–82.

19 Maes, P. (1994) 'Agents that reduce work and information overload', *Communications of the ACM*, Vol. 37, July, pp.31–40.

20 Technically, effect of brand includes the effects of brand attributes and the value of brand name (see [4]).

21 Andrews, R.L. and Manrai, A.K. (1998) 'Feature-based elimination: model and empirical comparison', *European Journal of Operational Research*, Vol. 111, 1 December, pp.248–267.

22 Chintagunta, P.K., Jain, D.C. and Vilcassim, N.J. (1991) 'Investigating heterogeneity in brand preferences in logit models for panel data', *Journal of Marketing Research*, Vol. 28, November, pp.417–528.

23 Gönül, F. and Srinivasan, K. (1993) 'Modeling multiple sources of heterogeneity in multinomial logit models: methodological and managerial issues', *Marketing Science*, Vol. 12, Summer, pp.213–29.

24 Guadagni, P. and Little, J.D.C. (1983) 'A logit model of brand choice calibrated on scanner data', *Marketing Science*, No. 2, Summer, pp.203–38.

25 We did not remove the price variable from the Peapod utility functions for estimation of the latent class models.

26 McAllister, L. (1983) 'A theory of consumer promotions', *SSM Working Paper No. 1457-83*, Sloan School of Management, Massachusetts Institute of Technology.