Beatrice Biondi, Laura Cornelsen

Reference price effect on consumer choice in online and traditional supermarkets: An application of discrete choice model on home scan data

Quaderni di Dipartimento

Serie Ricerche 2020, n. 2
ISSN 1973-9346

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

Mod.01P.5.5
Rev.01 24.5.08

Reference price effect on consumer choice in online and traditional supermarkets: An application of discrete choice model on home scan data

Beatrice Biondi[1] and Laura Cornelsen[2]

[1] *Department of Statistical Sciences, University of Bologna, b.biondi@unibo.it*
[2] *London School of Hygiene and Tropical Medicine, laura.cornelsen@lshtm.ac.uk*

July 6, 2020

Abstract

Reference price effect on consumer choices is an acknowledged phenomenon in marketing literature. Several studies have explored this issue using both observational studies and choice experiments. Furthermore, evidence exists on different consumer behaviour when shopping in traditional stores and online. Hence, question arises on whether and how consumers take into account reference price when shopping online. This study aims to analyze and compare consumer behaviour in online stores and traditional stores. A definition of reference price based on past prices is adopted and a discrete choice model is proposed, which includes gain and losses as additional product attributes, with individual-specific coefficients. The model is applied on observed cola purchases (home scan data) in traditional supermarkets and in an online store. Results indicate that loss aversion does play a role in online choices, but the effect is smaller if compared to choices in traditional stores.

JEL classification: D11, D12, Q11

Keywords: Discrete Choice Model, Mixed Logit Model, Control Function, Reference Price, Food Choice Behaviour.

1 Introduction

The term *reference price* refers to the idea that consumers adopt or follow a point of reference for price evaluation when choosing. More specifically, the choice process is assumed to involve a comparison between the actual price of a good and a price expectation, called reference price.

Reference-dependance in individual evaluations has been first theorized by Kahneman and Tversky in their Prospect Theory. The authors theorized and empirically proved that losses have greater impact on preferences than gains ("losses loom larger than correspondent gains"). This happens because individuals are averse to loss, and therefore the pain provoked by a loss is greater than the joy induced by a gain (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991).

The existence of a reference price effect on consumer choices is well documented in the economic and marketing literature (Kalwani et al., 1990; Kalyanaram and Winer, 1995; Winer, 1986). In particular, some recent studies explored this issue using choice experiment (Caputo et al., 2020, 2018; Hu, 2007), while other studies used observed purchases to assess reference price effect (Bell and Lattin, 2000; Cornelsen et al., 2018).

Beyond the general interest on reference price effect, the rising consumer adoption of online stores for groceries shopping poses new research questions. Does consumer behaviour differ online and offline? Do consumers make different choices when shopping online because of the different environment and shopping process?

Some evidences of changing consumer behaviour online and in traditional stores emerged in the literature. In particular, Degeratu et al. (2000) analyzed online and offline choices of three grocery products (one food: margarine) made by different samples, and found that brand names are more valuable online to the extent that lower information on other attributes is available. Accordingly, the brand of margarine is less important online because of the easier access to and comparison of nutritional information online than offline. Thus, factual information such as sugar or fat content has higher impact on online choices.

One further study compared differences in brand loyalty, finding that online consumers are more loyal to main brands but less loyal to small brands, compared to consumers buying in traditional stores (Danaher et al., 2003). On the other hand, Anesbury et al. (2016) found no differences in terms of time spent or effort expended when shopping online and in traditional stores.

Some studies also analyzed possible differences in price sensitivity when shopping online: when different samples for online and offline choices are condisidered, lower price sensitivity for online shoppers is found (Andrews and Currim, 2004; Degeratu et al., 2000, estimated on the same data on margarine and laundry detergent purchases). At the same time, Chu et al. (2010; 2008) analyze how price sensitivity on the same individuals varies when shifting between online and offline purchases and found that consumers are in general less price sensitive when shopping online; they found also differences in price sensitivity between heavy and light online shoppers, the latter being less price sensitive.

Based on the evidence summarized above, we are interested in exploring whether reference price has different effects on online and offline choices.

The aim of the present study is therefore to analyze reference price effect on consumer choices when shopping for groceries online, compared to the effect on consumers shopping in traditional stores. A definition of reference price based on past prices is adopted and a discrete choice model is proposed, which includes gain and losses as additional attributes, accounts for price endogeneity, and estimates individual-specific coefficients. The model is applied on observed cola purchases (home scan data) in traditional supermarkets and in the online store of the same chain.

Section 2 presents the methodology used, i.e. discrete choice model, control function approach to correct for endogeneity and the definition of reference price adopted. Section 3 presents the data and the empirical model. Then, results are outlined and discussed in Section 4.

2 Methodology

We analyze whether a reference price effect exists when shopping for food online in a discrete choice framework. The discrete choice model used is a logit model with individual specific parameters, also called mixed logit. In the words of McFadden and Train (2000) "Mixed logit is a highly flexible model that can approximate any random utility model". Hence, this model estimates the probability of choice based on utility maximization theory (Marschak, 1959; McFadden et al., 1973); the utility function for the individual i and j-th alternative is $U_{ij} = f(\mathbf{x_j}, \beta_\mathbf{i}) + \varepsilon_{ij}$ where $\mathbf{x_j}$ is a vector of M alternative-specific attributes including price; $\beta_\mathbf{i}$ is a vector of individual-specific coefficients for each attribute and for price; ε_{ij} is the random component. The utility function can be expressed as a weighted sum of alternatives' attributes, with weights equal to the taste parameters β (Lancaster, 1966). The probability of choosing the j-th alternative among J available alternatives, conditional on individual parameters β_i is:

$$P_{ij}|(\beta_i) = \frac{exp(\beta_i x_j)}{\sum_{l=1}^{J} exp(\beta_i x_l)} \tag{1}$$

The conditional probability of a sequence of K choices for an individual is the product of (1) over the set of choices, called $S_i(\beta)$ (Hole, 2007). Finally, the unconditional probability of the observed sequence of choices for each individual is:

$$P_i(\theta) = \int S_i(\beta)\ f(\beta|\theta)\ d\beta \tag{2}$$

This integral is approximated using maximum simulated likelihood.

Coefficients have a probability distribution given a priori and the error component is assumed to be iid extreme value distributed. The model returns an estimate of the average value of coefficients with an associated standard error and a standard deviation coefficient (with an associated standard error as well) for each random parameter.

Of paramount importance in discrete choice model specification is the inclusion of all product's attributes that contribute to utility, and in turn final choice. Ideally, all the relevant attributes should be included in the model and the error should capture only the random component, under the assumption that explanatory variables are independent of unobserved factors. In practice, there are a number of reasons why this could fail, leading to endogeneity problem and consequent inconsistency of estimates.

The most general cause for endogeneity in choice models is that the researcher cannot include all the factors that are related to price in the utility function, because sometimes they are not measurable (see Train 2009, Chapter 13 for an overview of endogeneity in discrete choice models). In the case of drinks, examples include the status of the brand, the design of the bottle/can and label, their tastiness, or the texture of bubbles. These attributes are not easily measurable and products likely incorporate them in their prices.

Control function approach to correct for endogeneity in discrete choice models has been proposed by Petrin and Train (2010) and is widely used in consumer research. This approach consists in adding a new variable to the model, which captures the part of error that is correlated with the endogenous variable.

This means that the error in the choice model, ε_j, can be split into two parts: one error correlated with the endogenous variable – captured by the control function, CF – and one *true* error ϵ_j: $\varepsilon_j = CF_j + \epsilon_j$ When price is endogenous, we assume that it is linear in instruments, with an error separable from the error term in the model. Therefore, the price of the alternative j (faced by a consumer in a given period and geographical area) can be written as:

$$p_j = \gamma z_{sj} + \mu_j \tag{3}$$

where z_{sj} with $s = 1, \ldots, S$ are a set of instruments that do not enter the utility function directly but affect price, and μ_j is the unobserved part.

The control function approach comprises two steps: first, equation 3 is estimated with OLS, obtaining the residuals $\hat{\mu_j}$; these residuals are then added to the choice model as a new variable, and a coefficient is estimated for this residual price. The control function is: $CF_j = \lambda\ \mu_j$; and the modified utility controlled for endogeneity of prices is therefore:

$$U_j = f(p_j, x_j, \beta) + \lambda\mu_j + \epsilon_j \tag{4}$$

As additional characteristics of each product, the choice model includes the possible loss or gain to evaluate whether and how they affect choice probability. Assuming the existence of a reference price in consumers' mind, losses and gains for each product are calculated as the difference between its current price and the reference price (Putler, 1992).

The loss for alternative j at time t, $Loss_{jt}$, is defined as:

$$Loss_{jt} = \begin{cases} p_{jt} - RP_{jt}, & \text{if } p_{jt} > RP_{jt} \\ 0, & \text{Otherwise} \end{cases} \tag{5}$$

The gain, $Gain_{jt}$, is:

$$Gain_{jt} = \begin{cases} RP_{jt} - p_{jt}, & \text{if } p_{jt} < RP_{jt} \\ 0, & \text{Otherwise} \end{cases} \tag{6}$$

In the present study, the reference price is defined based on past prices[1]; it is the price at time $t-1$ of the product chosen at time t (Krishnamurthi et al., 1992), that is $RP_{j,t} = p_{j,t-1}$. Accordingly, the consumer experiences a loss (gain) if the price of the product he/she chooses has increased (decreased) with respect to

[1] Other specifications of reference price have been proposed; they differ based on time – current vs past prices – and whether they are memory- or external-driven. See (Briesch et al., 1997) for an overview of reference price definition used in the literature

the previous purchase occasion. If the price is constant over time, the consumer experience no loss nor gain.

3 Data and empirical model

Discrete choice models are estimated on revealed preferences (real market purchases). We use Great Britain Kantar Fast Moving Consumer Good (FMCG) panel. The GB Kantar FMCG panel is a representative consumer panel of food and beverages purchased by households in GB (i.e. England, Wales and Scotland) and brought into their home. The data cover purchases of households in different types of retailers (e.g. supermarkets, independent market, specialized shops like butchers and greengrocers, online shops, etc.). Participants scan take-home purchases using hand-held barcode scanners[2]; for each household shopping trip, we have the detail of purchased products at the Universal Product Code (UPC) level. Ethical approval was not required as the data were obtained in anonymised format. Upon joining the panel, participants agree to the terms and conditions of GB Kantar FMCG[3].

The raw dataset consists in transaction-level data reported by more than 30 thousand households each year. The statistical unit is the UPC transaction: for each product purchased, we have information on volume purchased, number of item purchased, and amount spent. The price can be obtained by dividing expenditure by volume (or number of items); the dataset contains information about promotional status. Kantar provides also nutritional data on products

[2] That's why this kind of data are called *home scan* or *household scanner* data, as opposed to retail scanner data, which reports the purchases made at one point of sale with few information about households and are used in marketing research since the eighties (Guadagni and Little, 1983).

[3] see www.kantarworldpanel.com/en for contact details

through direct measurement in outlets, or using product images supplied by Brandbank, a third-party supplier. Furthermore, information on the day of purchase and outlet are recorded.

Self-reported socio-demographic data are collected by Kantar and describe household size and composition, age, ethnicity and highest qualification of the main shopper. It also includes information on the geographical location (postcode district), income group, social class and tenure of the household.

For the present study, we use data in the period 2014-2016.

Choice Set

The choice model takes into account the choice of a cola bottle with volume higher than 1 liter, among similar products (within-shelf choice). We consider the most disaggregated product level, identified by the UPC code. To avoid price and promotion differences between stores, which could obfuscate the reference price effect, only purchases made in one supermarket chain were taken. We consider one of the main retailers in GB and estimate models on purchases made in traditional, brick and mortar stores and on purchases made in online stores.

Modelling real market choices permits to avoid any potential bias due to the experimental condition of choice experiments, such as hypothetical bias. On the other hand, when estimating discrete choice models on actual choices, the choice set is unobservable by the researcher and must be defined ex-post. We restrict the choice set to several options of colas (following Carson and Louviere 2014), but do not include other beverages, under the assumption that consumers maximize utility of cola among available colas. The choice set contains the market leader products, i.e. six products that together account for 57% of total volume sales of cola in the retailer selected[4]. The choice set includes branded products as well as private labelled colas; three products are diet colas; sizes vary between 1.75L and 2L. Table 1 displays main characteristics of products in the choice set.

Table 1. Choice set descriptives: Products attributes.

	Brand	Sugar (g/100ml)
Product1	Branded	0.0
Product2	Branded	10.6
Product3	Branded	0.0
Product4	Private label	0.0
Product5	Branded	11.0
Product6	Private label	10.7

Price of the product is also a relevant attribute in the choice of colas. The price of the product chosen in each choice occasion can be obtained by dividing amount spent by the number of items bought; however, the estimation of discrete choice models requires prices and reference prices of non-chosen alternatives as well. Ideally we would like to have the price of alternatives on the shelves of the same store in the same day, but it is highly unlikely that every day at least one sampled household purchased those specific products in that specific outlet. Therefore, for each non-purchased product we define its price as the most frequent price observed in a specific region and week; i.e. the weeekly modal price in each of the 10 region.

Table 2 shows average prices separately for products purchased in traditional stores and online store, along with the number of observed choices that enter the model. The table displays that cola prices in the two types of store are similar, meaning that the retailer maintains the same pricing strategy in traditional and

[4]Proportion based on total sales in the traditional and online stores, among 79 products in total.

online stores.

Table 2. Choice set descriptives: Average prices and number of choices.

	Traditional purchases			Online purchases		
	Price (SD)	N	Share (%)	Price (SD)	N	Share (%)
Product1	1.26 (0.41)	1,254	19	1.24 (0.41)	837	31
Product2	1.12 (0.26)	1,215	19	1.10 (0.23)	379	14
Product3	1.11 (0.25)	1,151	18	1.10 (0.23)	679	25
Product4	0.48 (0.05)	1,331	21	0.46 (0.04)	342	13
Product5	1.23 (0.41)	724	11	1.21 (0.39)	259	10
Product6	0.48 (0.05)	785	12	0.50 (0.04)	175	7
Total		**6,460**	**100**	**0.50 (0.04)**	**2,671**	**100**

Notes: Prices are per bottle: prices are calculated as the average of modal unit values by region and week.

Interestingly, market shares are different in the two stores, highlighting the existence of different preferences when shopping online and in traditional stores. For example, branded products together account for 70 percent of shares of cola sales in brick-and-mortar stores, while accounting for 80 percent of shares in the online market, meaning that brands provide more utility online than offline. Moreover, diet products are more frequently purchased online than offline (69% of share versus 58%). These descriptive findings are in line with Degeratu et al. (2000) results on consumer choice behaviour online and offline.

Sample

The sample include purchases of one of the alternatives in the choice set made by households in the period 2014-2016; in order to include an individual IRP, we retain only households that purchased a cola for at least two consecutive weeks; the choice in week t is included in the model, while the price of the choice in week $t-1$ is used for determination of the reference price[5].

The final sample for traditional store purchases includes $1,067$ households, that

[5]Other data manipulations are: (1) households that purchase different products in the choice set over a week are excluded from the sample, i.e. household are only included if they buy only one of the six products in the choice set over a week; (2) if more than one pack of the same cola is bought, this is considered as a single choice of the product in the choice set..

make a total of $6,460$ cola purchases over 3 years. For purchases made on internet, the sample is made of 397 households, that make a total of $2,689$ cola purchases. 80 households are included in both samples. Table 3 presents descriptive statistics for the two samples.

Table 3. Samples Descriptive Statistics.

	Traditional sample		**Online sample**	
	Mean	St. Dev.	Mean	St. Dev.
Household size	3.18	1.32	3.50	1.29
Number of children	0.87	1.08	1.17	1.16
Number of children if have children	1.75	0.88	1.90	0.89
	Percent of Households		Percent of Households	
Households with children	49.8		61.5	
Income				
£70,000 +	6.2		5.8	
£60,000 - £69,999 pa	4.4		8.1	
£50,000 - £59,999 pa	9.1		9.3	
£40,000 - £49,999 pa	13.3		12.6	
£30,000 - £39,999 pa	20.7		19.9	
£20,000 - £29,999 pa	21.0		20.9	
£10,000 - £19,999 pa	19.9		18.6	
£0 - £9,999 pa	5.4		4.8	
ONS Social Grade				
Class AB	20.6		21.2	
Class C1	35.6		38.0	
Class C2	20.3		18.4	
Class D	15.7		17.4	
Class E	7.7		5.0	
Education of RP (highest qualification)				
Degree or higher	26.9		29.2	
Higher education	17.3		18.4	
A Level	16.5		17.6	
GCSE	25.7		25.4	
Other	6.8		5.8	
None	6.8		3.5	
Number of households	1,067		397	
Number of observations	6,460		2,689	

Reference price

The reference price is the one period lagged price, i.e. price in week $t-1$ of the product purchased in week t, as explained in Section 2.

Losses and gains are calculated as the difference between price and reference price and are region-specific. For the chosen alternative – for which price is available – losses and gains are calculated with respect to the actual price paid. Whereas, losses and gains of the other alternatives in the choice set are calculated based on the modal price for the same region in week t and week $t-1$. Losses and gains only occur when the distance between the actual price and the reference price exceeds 0.05 £. Trimming observations within this range reduces their occurrence and increases the average magnitude of losses and gains, to better discern their effect. Table 4 displays descriptive statistics of losses and gains for each product in the choice set, in the three years estimation period.

Table 4. Proportion of Losses and Gains, and Average Distance from Reference Price.

	Prop. of losses	Average loss	Prop. of gains	Average gain
Traditional stores				
Product1	12.9%	0.83 (0.23)	13.2%	0.82 (0.25)
Product2	10.6%	0.52 (0.26)	11.8%	0.51 (0.25)
Product3	10.8%	0.50 (0.27)	11.2%	0.49 (0.25)
Product4	15.1%	0.10 (0.03)	15.3%	0.10 (0.03)
Product5	14.5%	0.85 (0.23)	15.3%	0.81 (0.25)
Product6	13.5%	0.10 (0.03)	14.0%	0.10 (0.03)
Online store				
Product1	14.1%	0.80 (0.26)	14.3%	0.78 (0.28)
Product2	10.4%	0.50 (0.27)	11.0%	0.49 (0.25)
Product3	9.5%	0.47 (0.26)	10.6%	0.44 (0.22)
Product4	6.8%	0.10 (0.05)	7.4%	0.10 (0.05)
Product5	13.4%	0.87 (0.22)	14.9%	0.78 (0.27)
Product6	4.8%	0.11 (0.08)	5.0%	0.11 (0.07)

Notes: Average losses (gains) are the average distance between the actual price and the reference price conditional on losses (gains). Standard deviations in parentheses.

Empirical model

The attributes included in the empirical model are price, gain, loss, size of the bottle, whether the product has a private label or it is branded, and sugar content in g/100m; for all attributes, coefficients in the model are assumed to be normally distributed. Interactions between alternative specific constants and demographic variables are included in the model. Demographics considered are income, household size, number of children living in the house, social class (AB class versus others) and highest level of education (whether degree or higher, higher education or other types).

Control function approach is used to correct for price endogeneity in the empirical model, as discussed in Section 2. A set of linear regressions on prices are estimated using Hausman type instrumental variables Hausman (1996). For each region, the instruments are the prices in the other GB regions. The residuals are then used as a new variable in the discrete choice model.

The choice model estimates a parameter for the observed price and separate paramenters for gain and loss (Kalwani et al., 1990). Therefore, the deterministic utility V_{jit} of alternative j with explicit attributes is:

$$\begin{aligned} V_{jit} = {} & \beta_1 p_{jit} + \beta_2 res\_price_{jit} + \beta_3 Gain_{jit} + \beta_4 Loss_{jit} + \beta_5 sugar_{jit} \\ & + \beta_6 size_{jit} + \beta_7 priv\_lab_{jit} + \beta_8 inc_i * A_k + \beta_9 hh\_size_i * A_k + \beta_{10} child_i * A_k \\ & + \beta_{11} class\_AB_i * A_k + \beta_{12} degree_i * A_k + \beta_{13} high\_educ_i * A_k \end{aligned} \tag{7}$$

where A_k is the alternative specific constant included for $K - 1$ alternatives, and each interaction has $K - 1$ associated parameters. The final number of parameters, with 6 alternatives in the choice set is therefore $7 + (5 * 6) = 37$. All parameters except for the interactions are assumed to follow a normal distribution.

A mixed logit model is estimated with attributes and interactions specified in (7). Maximum simulated likelihood is used for estimation.

4 Results and discussion

Table 5 displays estimated parameters on traditional stores and online store purchases. The standard deviations of some coefficients are significant, indicating that there is variation in the taste for attributes among individuals, which is captured by the model. Table 6 displays model diagnostics. The complete model, with added loss and gain parameters, fits the data significantly better than the reduced one[6], according to likelihood ratio test results, both for online and offline models.

In general, higher prices, higher sugar content and private label (with respect to main brands) decrease choice probability in both types of purchase, online and offline. However, since coefficients are normally distributed, not all consumers place negative value on those attributes, and we can retrieve the share of consumers that places a positive value on each attribute (Train, 2009, Ch. 6). More specifically, our results highlight that in traditional stores almost all consumers (98%) prefer lower prices, even if some consumers more than others. In online stores, all consumer prefer lower prices and there is no variation (price estimated standard deviation not significant). These results are in line with our expectations. In the same fashion, despite on average consumers prefer low sugar, for 26% of traditional consumers and 41% of online consumers higher sugar content increases product utility; between 20% and 25% of consumers prefer private labelled colas both online and offline.

We have seen that on average branded products cost almost three times as private labelled products. Despite the fact that some consumers prefer private label, willingness to pay results indicate that on average online consumers

[6]Estimated coefficients of the reduced model, without loss and gain parameters, are displayed in the Appendix.

Table 5. Model results.

	Traditional store		Online store	
	Coefficients	S.E.	Coefficients	S.E.
Income*Product1	0.12*	0.03	0.08	0.05
Hh size*Product1	-3.52*	0.42	2.17	1.17
Child*Product1	-0.61	0.84	-1.36	3.08
Class AB*Product1	7.54*	1.38	7.23*	2.92
Degree*Product1	2.95*	1.15	2.03	1.32
High educ.*Product1	-1.58	1.21	-3.18	1.77
Income*Product2	0.08*	0.02	0.12*	0.04
Hh size*Product2	-1.48*	0.24	1.31	1.14
Child*Product2	-1.17	0.72	-1.95	2.92
Class AB*Product2	7.50*	1.38	3.47	2.69
Degree*Product2	2.90*	1.02	-3.08*	1.39
High educ.*Product2	2.43*	0.91	-1.11	1.60
Income*Product3	0.12*	0.03	0.18*	0.05
Hh size*Product3	-4.24*	0.50	-0.39	1.21
Child*Product3	0.52	0.79	5.10	3.24
Class AB*Product3	7.86*	1.35	5.57	3.01
Degree*Product3	0.72	1.22	-1.03	1.56
High educ.*Product3	-0.71	1.30	-4.73*	1.95
Income*Product4	0.03	0.02	0.02	0.05
Hh size*Product4	-2.14*	0.37	0.04	0.78
Child*Product4	-1.61*	0.66	4.44	3.28
Class AB*Product4	4.15*	0.99	6.91	3.65
Degree*Product4	1.40	1.00	-0.77	1.93
High educ.*Product4	-2.60*	1.13	-5.17*	1.67
Income*Product5	0.08*	0.02	0.07	0.04
Hh size*Product5	-1.39*	0.30	1.43	1.19
Child*Product5	-1.01*	0.90	-4.02	3.02
Class AB*Product5	5.54*	1.34	2.37	2.85
Degree*Product5	2.06*	1.09	-2.59	1.39
High educ.*Product5	0.95	0.87	0.11	1.40
Price	-5.14*	0.38	-3.37*	0.55
Loss	-0.91*	0.37	-1.53*	0.77
Gain	-1.59*	0.28	-0.25	0.46
Sugar	-0.87*	0.13	-0.54*	0.10
Residual price	7.49*	0.54	6.70*	0.96
Private label	-11.72*	1.13	-13.82*	2.19
Size	1.38	2.12	7.55	4.23
SD				
Price	2.56*	0.34	-0.94	0.48
Loss	0.53	0.44	0.64	0.82
Gain	1.07*	0.38	-1.09	0.70
Sugar	1.34*	0.13	2.41*	0.26
Residual price	2.59*	0.50	2.46*	0.98
Private label	14.87*	1.16	20.96*	2.50
Size	26.38*	1.77	47.70*	4.36

Notes: Asterisk indicates the parameter is significant at the α=0.05 level

Table 6. Models diagnostics.

	k	N	LL	LR test	AIC	BIC	RMSE
Traditional stores							
Reduced model	35	38,760	-4128.23	53.96***	8326.46	8626.24	3.22
Complete model	37	38,760	-4101.25		8276.50	8593.41	3.74
Online store							
Reduced model	35	16,134	-1422.19	17.98***	2914.38	3183.48	3.69
Complete model	37	16,134	-1413.20		2900.40	3184.88	3.43

Notes: k=number of model parameters; N=number of observations; *** indicate the test is significant at α=0.01 level, with two degrees of freedom; the Reduced model does not include loss and gain parameters.

would be willing to pay £4 for the brand, ceteris paribus; consumers in traditional stores are willing to pay £2.3 for a cola with brand. Furthermore, results indicate that on average consumers are willing to pay around 15 pence for each 1 gram per 100ml reduction of sugar, both online and offline.

Socio demographics significantly affect probabilities of choosing different colas. For instance, in traditional stores the higher the income, the higher the probability of buying branded colas when shopping; on the other hand, the higher the household size the higher the probability of buying product 6 (private label with sugar), the opposite happens with being in class A or B, which decreases the probability of buying product 6. However, this effect is lower when considering online choices. It is particularly interesting to note that household size and number of children do not affect online choices. This may be due to the limited influence that children have on online shopping (Ayadi and Muratore, 2020; Haselhoff et al., 2014).

Gain and loss parameters significantly affect choices in traditional stores, while only losses seem to significantly affect online choices. Gain and loss are always joined with price, therefore their coefficients have to be interpreted together with the price coefficient, since the higher the loss (gain), the higher (lower) the price.

An example could help to understand how loss and gain really affect choice probabilities: suppose that in traditional stores there is an increase of price of £1 for option j, while prices of other options remain constant. If we do not consider the loss effect, price decreases total utility of alternative j of a certain amount, because of the estimated coefficient -5.14. However, following our assumption, we predict that the choice probability would decrease to a greater extent as effect of loss aversion. In fact, if we jointly consider the effect of price and loss, this is higher than price effect alone, $-5.14+(-0.91) = -6.05$. On the opposite, if price decreases by £1, according to our model we must consider the joint effect of price and gain, and thus $-1*(-5.14)+(-1.59) = 3.55$. Therefore, the model rightly captures the loss aversion behaviour: when a loss occurs, the negative contribution to utility of price and loss is higher than the positive contribution given by price and gain. If we consider the online choice, the gain effect is not significant, meaning that the contribution to utility in case of a gain is prompt only by the price decrease effect and is not modified for effect of a perceived gain. Still, the response in case of a loss is higher due to the loss coefficient.

The change in choice probabilities when loss and gain occur can be simulated by models prediction with different price, loss or gain. We derive elasticities for price increase (and relative loss) and price decrease (and relative gain). Table 7 shows own-price elasticities[7] in traditional and online sample and displays results of the reduced model (without loss and gain parameters) and of the complete model acounting for loss and gain effects. We can see in the table that elasticities of the complete model account for loss aversion. This means

[7]Cross-price elasticities available upon request.

that including loss and gain coefficients affects the change in choice probabilities prompt by an increase or decrease in price. More specifically, compared to the reduced model, the complete model predicts relatively higher decreases in choice probabilities when price increases (i.e. loss aversion); on the other hand, it predicts relatively lower increases in choice probability when price decreases. This is true in general, except for a price increase in the online store, where loss avesion effect seems to be mitigated.

Table 7. Price elasticities in choice probabilities.

	Traditional sample				Online sample			
	Reduced model		Complete model		Reduced model		Complete model	
Percentage change in choice probabilities when price increases by £1	Mean	St. Err.	Mean	St. Err.	Mean	St. Err.	Mean	St. Err.
Product1	-13.65	0.05	-16.57	0.07	-11.66	0.03	-12.37	0.03
Product2	-10.34	0.04	-11.86	0.04	-7.29	0.02	-6.96	0.02
Product3	-11.00	0.04	-12.62	0.05	-8.12	0.04	-8.80	0.04
Product4	-6.79	0.02	-8.14	0.02	-4.22	0.02	-3.83	0.02
Product5	-8.10	0.04	-9.24	0.04	-5.63	0.03	-6.08	0.03
Product6	-6.02	0.03	-6.78	0.04	-2.71	0.02	-2.55	0.02
Percentage change in choice probabilities when price decreases by £1								
	Traditional sample				Online sample			
	Reduced model		Complete model		Reduced model		Complete model	
	Mean	St. Err.	Mean	St. Err.	Mean	St. Err.	Mean	St. Err.
Product1	17.78	0.03	13.64	0.03	12.89	0.02	8.34	0.02
Product2	14.70	0.03	11.10	0.02	9.15	0.02	5.37	0.02
Product3	17.27	0.04	13.20	0.03	11.25	0.04	7.30	0.03
Product4	9.29	0.02	7.72	0.02	5.55	0.02	3.18	0.01
Product5	14.32	0.04	10.93	0.03	8.59	0.04	5.32	0.02
Product6	8.35	0.03	6.61	0.02	3.86	0.02	2.18	0.01

Notes: the Reduced model does not include loss and gain parameters.

5 Conclusion

We analyzed consumer choice behaviour in online versus traditional stores. In particular, we explored whether and how reference price affects the choice of cola on a large sample of observed purchases, based on results of mixed logit models with added coefficients for losses and gains relative to product price in the previous purchase occasion.

Results confirm the existence of a reference price effect and loss aversion both on online and offline choices. Furthermore, the asymmetry in consumer responses due to price variations seems to be larger on choices made in traditional stores; this is in line with previous results on lower price sensitivity of online consumers (Andrews and Currim, 2004; Degeratu et al., 2000).

The existence of loss aversion has important practical implications for marketing activities such as price promotion and store layout, not only in traditional stores, but also in online stores. Results of the present study confirm that consumers behave differently when shopping online, meaning that in the online store particular promotions can be proposed, which must be designed based on the peculiar online consumer behaviour. Moreover, the online setting allows to target specific consumers with ad-hoc promotions that can be based on each individual purchase history.

Further research is needed on reference price effect on online consumer behaviour. In future studies, other definitions of reference effect can be explored (Briesch et al., 1997); other products can be considered, to explore whether loss aversion differes when purchasing e.g. more expensive products or perishable products for which stockpiling under promotions is more difficult. Moreover, studying heterogeneity among consumers is highly relevant.

References

Andrews, R.L., Currim, I.S., 2004. Behavioural differences between consumers attracted to shopping online versus traditional supermarkets: implications for enterprise design and marketing strategy. International Journal of Internet Marketing and Advertising 1, 38–61. doi:10.1504/IJIMA.2004.003689.

Anesbury, Z., Nenycz-thiel, M., Dawes, J., Kennedy, R., 2016. How do shoppers behave online? An observational study of online grocery shopping. Journal of Consumer Behaviour 15, 261–270. URL: http://oro.open.ac.uk/26883/, doi:10.1002/cb.

Ayadi, K., Muratore, I., 2020. Digimums' online grocery shopping: the end of children's influence? International Journal of Retail and Distribution Management 48, 348–362. doi:10.1108/IJRDM-09-2019-0291.

Bell, D.R., Lattin, J.M., 2000. Looking for Loss Aversion in Scanner Panel Data: The Confounding Effect of Price Response Heterogeneity. Marketing Science 19, 185–200. doi:10.1287/mksc.19.2.185.11802.

Briesch, R.A., Krishnamurthi, L., Mazumdar, T., Raj, S., 1997. A Comparative Analysis of Reference Price Models. Journal of Consumer Research 24, 202–214. doi:10.1007/978-3-319-08395-7_3.

Caputo, V., Lusk, J.L., Nayga, R.M., 2020. Am i getting a good deal? reference-dependent decision making when the reference price is uncertain. American Journal of Agricultural Economics 102, 132–153.

Caputo, V., Lusk, J.L., Nayga Jr, R.M., 2018. Choice experiments are not conducted in a vacuum: The effects of external price information on choice behavior. Journal of Economic Behavior & Organization 145, 335–351.

Carson, R.T., Louviere, J.J., 2014. Statistical properties of consideration sets. Journal of Choice Modelling 13, 37–48. URL: http://dx.doi.org/10.1016/j.jocm.2015.01.001, doi:10.1016/j.jocm.2015.01.001.

Chu, J., Arce-Urriza, M., Cebollada-Calvo, J.J., Chintagunta, P.K., 2010. An Empirical Analysis of Shopping Behavior Across Online and Offline Channels for Grocery Products: The Moderating Effects of Household and Product Characteristics. Journal of Interactive Marketing 24, 251–268. URL: http://dx.doi.org/10.1016/j.intmar.2010.07.004, doi:10.1016/j.intmar.2010.07.004.

Chu, J., Chintagunta, P., Cebollada, J., 2008. A Comparison of Within-Household Price Sensitivity across Online and Offline Channels. Marketing Science 27, 283–299. doi:10.1287/mksc.1070.0288.

Cornelsen, L., Mazzocchi, M., Smith, R., 2018. Between preferences and references: Evidence from Great Britain on asymmetric price elasticities. Quaderni di Dipartimento, Department of Statistics, University of Bologna URL: https://ideas.repec.org/p/bot/quadip/wpaper139.html, doi:10.6092/unibo/amsacta/5821.

Danaher, P.J., Wilson, I.W., Davis, R.A., 2003. A comparison of online and offline consumer brand loyalty. Marketing Science 22, 461–476.

Degeratu, A.M., Rangaswamy, A., Wu, J., 2000. Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. International Journal of Research in Marketing 17, 55–78. doi:10.1016/s0167-8116(00)00005-7.

Guadagni, P.M., Little, J.D.C., 1983. A Logit Model of Brand Choice Calibrated on Scanner Data. Marketing Science 2, 203–238.

Haselhoff, V., Faupel, U., Holzmüller, H.H., 2014. Strategies of children and parents during shopping for groceries. Young Consumers: Insight and Ideas for Responsible Marketers 15, 17–36.

Hausman, J.A., 1996. Valuation of new goods under perfect and imperfect competition, in: The economics of new goods. University of Chicago Press, pp. 207–248.

Hole, A.R., 2007. Fitting mixed logit models by using maximum simulated likelihood. The Stata Journal 7, 388–401.

Hu, W., 2007. Bargains or Rip-offs? Reference Price Effects in Conjoint Stated Demand. Journal of Agricultural and Resource Economic 31, 386–408. doi:10.1097/EDE.0b013e3181.

Kahneman, D., Tversky, A., 1979. Prospect Theory : An Analysis of Decision under Risk. Econometrica 47, 263–292.

Kalwani, M.U., Yim, C.K., Rinne, H.J., Sugita, Y., 1990. A Price Expectations Model of Customer Brand Choice. Journal of Marketing Research 27, 251. doi:10.2307/3172584.

Kalyanaram, G., Winer, R.S., 1995. Empirical Generalizations from Reference Price Research. Marketing Science 14, G161–G169.

Krishnamurthi, L., Mazumdar, T., Raj, S.P., 1992. Asymmetric Response to Price in Consumer Brand Choice and Purchase Quantity Decisions. Journal of Consumer Research 19, 387–400.

Lancaster, K.J., 1966. A new approach to consumer theory. Journal of political economy 74, 132–157.

Marschak, J., 1959. Binary Choice Constraints on Random Utility Indicators. Cowles Foundation Discussion Papers 74. Cowles Foundation for Research in Economics, Yale University. URL: https://EconPapers.repec.org/RePEc:cwl:cwldpp:74.

Mazumdar, T., Raj, S., Sinha, I., 2005. Reference Price Research: Review and Propositions. Journal of Marketing 69, 84–102. URL: http://journals.ama.org/doi/abs/10.1509/jmkg.2005.69.4.84, doi:10.1509/jmkg.2005.69.4.84.

McFadden, D., Train, K., 2000. Mixed mnl models for discrete response. Journal of applied Econometrics 15, 447–470.

McFadden, D., et al., 1973. Conditional logit analysis of qualitative choice behavior .

Petrin, A., Train, K., 2010. A Control Function Approach to Endogeneity in Consumer Choice Models. Journal of Marketing Research 47, 3–13. doi:10.1509/jmkr.47.1.3.

Putler, D.S., 1992. Incorporating Reference Price Effects into a Theory of Consumer Choice. Marketing Science 11, 287–309.

Train, K.E., 2009. Discrete choice methods with simulation. Cambridge university press.

Tversky, A., Kahneman, D., 1991. Loss Aversion in Riskless Choice : A Reference-Dependent Model. The Quarterly Journal of Economics 106, 1039–1061.

Winer, R.S., 1986. A Reference Price Model of Brand Choice for Frequently Purchased Products. Journal of Consumer Research 13, 250. doi:10.1086/209064.

A Appendix

Table A1. Reduced Model results.

	Traditional store		Online store	
	Coefficients	S.E.	Coefficients	S.E.
Income*Product1	0.12*	0.02	0.10*	0.05
Hh size*Product1	-4.29*	0.60	2.15*	0.72
Child*Product1	1.71*	0.76	-8.32*	2.18
Class AB*Product1	9.87*	1.29	7.19*	2.23
Degree*Product1	3.51*	1.05	0.49	1.40
High educ.*Product1	1.51	0.85	-2.52	1.66
Income*Product2	0.04	0.02	0.10*	0.04
Hh size*Product2	-0.48	0.26	2.25*	0.73
Child*Product2	-3.12*	0.81	-9.11*	1.98
Class AB*Product2	10.25*	1.25	3.49	2.02
Degree*Product2	0.98	0.99	-2.95	1.55
High educ.*Product2	4.23*	0.83	-2.45	1.59
Income*Product3	0.11*	0.02	0.19*	0.05
Hh size*Product3	-4.69*	0.52	-0.06	0.81
Child*Product3	1.45	0.75	-3.93	2.30
Class AB*Product3	10.10*	1.32	3.29	2.86
Degree*Product3	1.95	1.09	-0.47	1.83
High educ.*Product3	1.30	0.87	-2.67	1.90
Income*Product4	0.07*	0.02	0.07	0.04
Hh size*Product4	-3.89*	0.47	0.13	0.54
Child*Product4	1.98*	0.65	1.56	2.23
Class AB*Product4	6.57*	1.01	5.54*	2.26
Degree*Product4	1.09	0.89	-4.84*	1.73
High educ.*Product4	-1.32	0.75	-3.94*	1.77
Income*Product5	0.05*	0.02	0.10*	0.04
Hh size*Product5	-0.63*	0.30	1.90*	0.69
Child*Product5	-1.81*	0.76	-10.29*	2.10
Class AB*Product5	7.67*	1.25	3.59	2.11
Degree*Product5	-0.56	0.97	-2.45	1.60
High educ.*Product5	3.23*	0.83	-8.30*	1.52
Price	-4.81*	0.32	-4.32*	0.55
Sugar	-0.99*	0.15	-0.40*	0.09
Residual price	7.51*	0.56	7.77*	1.11
Private label	-10.23*	1.22	-15.45*	2.10
Size	0.65	2.42	5.79	4.45
SD				
Price	-1.59*	0.26	-1.57*	0.49
Sugar	1.30*	0.11	1.67*	0.17
Residual price	-2.93*	0.61	3.40*	0.94
Private label	17.74*	1.15	20.15*	1.93
Size	27.56*	1.67	45.22*	3.84

Notes: the Reduced model does not include loss and gain parameters.
Asterisk indicates the parameter is significant at the α=0.05 level