[Submitted November 23, 2020, Revised April 23, 2021 and September 2, 2021, Accepted September 11, 2021]

Using supermarket loyalty card data to measure the differential impact of the UK soft drink sugar tax on buyer behaviour

Andrew Fearne[1], Natalia Borzino[2], Beatrix De La Iglesia[3], Peter Moffatt[4] and Margaret Robbins[5]

Abstract

This paper explores the impact of the soft drinks sugar tax introduced in the UK in 2018 on the purchasing behaviours of different geo-demographic consumer segments. We analyse data for a composite good comprising the most popular Sugar-Sweetened Drinks (SSDs) using loyalty card data from one of the UK's largest supermarkets. We use pre-levy data to predict the effect of the tax and corroborate our predictions by analysing actual consumption of the composite good in the first 5 months post-levy. The results show that the impact of the sugar tax is likely to have the desired effect of reducing the purchase of SSDs. Moreover, whilst the impact of the tax is likely to vary across different geo-demographic segments, the evidence suggests that its impact is likely to be greatest on the most vulnerable market segments – families on low incomes – who are amongst the highest consumers of SSDs in the UK.

***Keywords**:* Sugar tax; supermarket loyalty card data; geo-demographic segmentation; UK

JEL Classifications: D12, D22, I18, L66, M38, Q18, Q28

1. Introduction

Loyalty schemes have become established weapons in the armoury of retailers seeking to gain competitive advantage through more effective product ranging, in-store merchandising and promotional activity tailored to the (heterogeneous) behaviour of their customers. However,

[1] Corresponding Author, Norwich Business School, University of East Anglia (a.fearne@uea.ac.uk)
[2] Institute for Environmental Decisions, ETH Zürich (natalia.borzino@frs.ethz.ch)
[3] School of Computing Sciences, University of East Anglia (b.Iglesia@uea.ac.uk)
[4] School of Economics, University of East Anglia (p.moffatt@uea.ac.uk)
[5] School of Computing Sciences, University of East Anglia (margaret.robins@uea.ac.uk

there is little evidence to date of this rich source of behavioural data being used by policymakers to inform or evaluate the development of policies, legislation or (public) interventions designed to foster more sustainable purchasing decisions and consumption behaviour. One area in which there is a compelling need for behavioural change is diet and health. This paper illustrates the potential value of behavioural insights derived from supermarket loyalty card data for policy-makers and other stakeholder (food retailers, food manufacturers and NGOs) who wish to foster healthier choices in the mainstream food purchasing environment of supermarkets.

Previous studies (Felgate et al, 2012, Yamoah et al, 2014) have demonstrated an important advantage that loyalty card data has over scanner data, which is the ability to analyse purchasing behaviour by distinct consumer segments and the differential impacts of interventions designed to change their behaviour, including changes in the retail price. Such analysis is not possible with store-level scanner data, as no association is made between the sale of an item and the person who purchased it. Loyalty card information is routinely and systematically used by retailers and food manufacturers to inform decisions about marketing (pricing, distribution, ranging, and merchandising) all of which have an impact on purchasing behaviour. Moreover, the information associated with loyalty cards enables inferences to be drawn regarding consumption, as loyalty cards are typically associated with single households with known (geo-demographic) characteristics.

We have a particular interest in consumer behaviour in relation to prices because we want to understand the differential impact of the sugar tax, introduced in April 2018 in the UK. The tax was announced by the UK government in 2016 and is being applied to Sugar-Sweetened Drinks (SSD). Excessive consumption of SSD presents a real problem for public health as they provide little nutritional benefit while contributing to weight gain and probably to the risk of diabetes, cardiovascular heart disease and dental caries (Vartanian et al, 2007, Malik et al., 2006, 2010, Ng et al., 2012).

Through the lens of normative economics, there are two key arguments in favour of the tax. First, given that some people may disregard the effects of over-consuming SSDs on their current and future health, being misinformed or prone to cognitive biases (e.g., time inconsistency), the tax is likely to guide such individuals towards more rational and healthy behaviour. Such effects are termed "internalities" (Gruber, 2002; Griffith et al., 2018). Second, the tax is likely to result in wider societal benefits others through the reduction of the burden of diet-related diseases on the National Health Service (NHS).

The UK sugar tax consists of a tiered levy with two bands, one of 18 pence per litre for soft drinks with more than 5g of sugar per 100ml (low tier) and a higher one of 24 pence per litre for drinks with more than 8g per 100ml (high tier). Other drinks with lower than 5g of sugar are not taxed. It is the first time that such a tiered industry levy has been used, as other countries have opted for a sales tax instead. It is hoped that the levy will help to tackle the nation's obesity problem by reducing consumption of sugar in SSDs, particularly among younger adults. However, there are concerns that it could become a regressive tax if it disproportionally affects poorer members of society.

We analyse consumer behaviour pre-tax as well as their behavioural response as a result of the introduction of the tax. To do so, we use two panel databases containing SSD purchase data from over 2 million loyalty card holders across the UK, divided into ten geo-demographic segments, derived from a major UK supermarket chain.

The objectives of our study are twofold. First, we analyse consumer behaviour using 100 weeks of pre-tax data and predict the effect of the tax on purchasing patterns of SSDs for the different consumer segments. This allows us to establish a methodology for predicting the effect of similar policies in the future. Second, we compare our predictions with the purchases observed for the first few months (23 weeks) after the implementation of the tax (April 2018), as this provides corroboration of our predictive approach using measurements on the initial effect of the policy.

Our approach is based on a composite SSD product using the prices of the principal products relevant to the tax to obtain price elasticities per consumer segment. Then we use these to predict the effect of a price change equivalent to the imposition of the tax on each consumer segment. This is compared with the actual purchasing changes before and immediately after the imposition of the tax.

Our results are of interest to public health experts, to the UK and other governments considering similar policies, and to the public. They provide clear and important evidence about the UK sugar tax on the purchase (and hence supply) of SSDs, particularly amongst the most vulnerable consumer segments. Furthermore, they provide an exemplar for using supermarket loyalty card data for other policy interventions designed to influence food and drink purchasing behaviour.

The paper is organised as follows. Section 2 presents a review of relevant literature. Section 3 describes the materials and methods. Section 4 reports the results and finally, Section 5 presents our conclusions and recommendations for further research.

2. Literature review

The UK was the first country to introduce a tiered tax on SSDs (Briggs et al, 2017). Producers and/or retailers are free to pass on the full tax to consumers (i.e., price pass-through) or to absorb it, fully or in part. Worldwide, there is evidence of price pass-through from SSD taxes ranging from 40% to more than 100% (Aguilar et al., 2018; Alsukait et al., 2020; Arteaga et al., 2017; Berardi et al., 2016; Bollinger and Sexton, 2018; Capacci et al., 2019; Cawley and Frisvold, 2017; Cawley et al., 2018; Etile t al., 2018; Zhong et al., 2018; Falbe et al., 2015; Grogger, 2017; Rojas and Wang, 2017; Seiler et al., 2019).

There are concerns that a tax such as this could become regressive if it is passed through to consumers and disproportionally affects the less affluent social groups (Dubois et al., 2017). In this context, a recent study proposed how to calculate the optimal soda tax range(s) (Allcott et al, 2019) to strike the right balance between corrective and redistributive motives to avoid the possible regressive nature of such taxes.

Another feature of the UK sugar tax design is to incentivise producers to reformulate products – reducing the sugar content to avoid the tax. However, this represents a risk for the manufacturer given that the new recipes may not be well received by consumers (Geykens et al., 2018; Goncalvez and Pereira dos Santos, 2019). Other countries (e.g.France from 2013 and Portugal from 2017) have also structured their SSD taxes to encourage reformulation (Goiana-da Silva et al., 2018).

More than 40 countries have implemented SSD taxes and in some the implementation has been delegated to regional authorities, such as Catalonia in Spain and California, Berkeley, Boulder, Colorado, Philadelphia, and Pennsylvania in the USA (Global Food Research Program, 2020). Several studies have analysed the effect of the taxes on SSD consumption. The evidence suggests that taxes decrease the SSD consumption from 6% in Mexico to more than 20% in Berkeley and Philadelphia and 33% in Saudi Arabia (Aguilar et al., 2018; Alsukait et al., 2020; Arteaga et al., 2017; Castello and Lopez-Casasnovas, 2018; Cawley et al., 2019; Colchero et al., 2016; Colchero et al., 2017; Falbe et al., 2016; Mora et al., 2018; Zhong et al., 2018; Nakamura et al., 2018; Seiler et al., 2019; Silver et al., 2017; Taylor et al., 2019). The differences in the reported impacts could be due to a number of factors, including the quality of the data, the methodologies employed but also the design and scale of the taxes used.

With respect to the type of data used to analyse the effect of the levy, the literature so far has relied primarily on surveys, scanner data (e.g. Kantar World Panel) or hand-collected data on a small number of products or in a small number of stores (Berardi et al., 2016; Castello and Lopez-Casasnovas, 2018; Seiler et al., 2019), with a small number of studies having access to retail data (Goncalvez and Pereira dos Santos, 2019).

In terms of methodologies applied, studies predicting and simulating the impact of the levy *ex-ante* use theoretical models (e.g. Briggs et al 2013; Briggs et al, 2017). Other studies analysing the impact of the tax *ex-post* using counterfactuals (e.g. Colchero et al. 2016); difference-in-difference analysis (e.g. Goncalvez and Pereira dos Santos, 2019) or synthetic control methods (e.g. Grogger, 2017).

Looking at the effect of the UK tax, Briggs et al. (2013) focused on predicting changes in the number and percentage of overweight and obese adults post-tax. They used survey data from various sources, which is problematic as applying results from one dataset to another requires the assumption that the samples are drawn from the same population, which is not always the case. The study predicted that the tax would reduce obesity by 1.3% with the greatest effect occurring in young people. No significant differences were predicted between different income groups.

In another study, Briggs et al. (2017) simulated how the different industry responses to the tax could impact on health. They included product reformulation to reduce sugar content and avoid the tax; an increase in price when the levy is passed on to consumers; and a change in market share as consumer switch to the lower sugar alternatives. Their findings suggested that the greatest health benefit would come with the reformulation of the products. As a result, individuals aged younger than 18 years and those aged older than 65 years would benefit the most from the predicted reduction in obesity, diabetes, and dental decay.

A recent study has assessed the effect of the UK tax on British households one year after the implementation of the policy (Pell et al, 2021). To our knowledge, this is the first paper that estimates the effect of the intervention post-tax. The authors first used pre-tax data to analyse consumer behaviour and predict the effect of the tax using a counterfactual. Then, they compared the predictions with the observed changes in volume and sugar intake one-year post-tax. They analyse a composite product to represent the high/low/no levy drinks. Their findings suggest that in the consumption of high tax tier (the one we study) SSDs decreased by 44.3% and the associated sugar consumption decreased by 45.9%. They also identified decreases in the consumption of low tier drinks but no changes in volumes consumed for drinks that did not attract the tax. Overall, considering all soft drinks, they found that the total volume consumed did not change but associated sugar consumption fell by 9.8%. The authors conclude that reduction in sugar was most likely the result of reformulation. The study used household scanner data from a panel of households reporting their purchasing on a weekly basis (Kantar Worldpanel). However, their data do not allow for analysis of differential effects among different consumer segments.

Our study is the first to analyse the differential impact of the levy across a diverse range of (geo-demographic) consumer segments, using supermarket loyalty card data.

3. Materials and Methods

3.1. Data

The data used for this study are obtained from one of the UK's largest supermarkets. Their loyalty scheme generates a panel dataset of over two million households, which is a 10% sample of the population of loyalty card holders. A considerable advantage of loyalty card data over data from a typical household survey is the level of product disaggregation: it provides access to information on both prices and purchase quantities for thousands of individual products by specific consumers (loyalty card holders) on a weekly basis. Hence, for each geo-demographic consumer segment we have weekly sales by volume, weekly prices and weekly number of customers for the entire soft drinks category, comprising hundreds of products and dozens of brands. Since for SSDs we have unit size in millilitres, we can calculate the total volume of sales in litres, litres per customer and price per litre for each product.

The sales data is segmented by ten different geo-demographic segments, using the classification provided by Cameo, which groups households in neighbourhoods that share certain characteristics, such as lifestage, lifestyle, affluence, ethnicity and employment. (Cameo, 2020). The resulting segments are presented in Table 1, in order of affluence, and comprise: affluent singles and couples in exclusive urban neighbourhoods (YAS); wealthy neighbourhoods nearing & enjoying retirement (WRN); affluent home owning couples & families in large houses (AHO); suburban homeowners in smaller private family homes (SPFH); comfortable mixed tenure neighbours (CMN); less affluent family neighbourhoods (LAF); less affluent singles and students in urban areas (LASS); poorer white and blue collar workers

(PWBCW); poorer family and single parent households (PFSPH); and poorer council tenants including many single parents (PCT).

Geo-demographic segmentation dates back to the mid-1970s and is defined by Birkin & Clarke (1998) as "the study of the population types and their dynamics as they vary by geographical area". What distinguishes geo-demographic segmentation from other segmentation approaches is that the unit of analysis is the neighbourhood rather than the individual. The fundamental rationale is that "the social context in which people live (has) a significant effect on their consumption patterns as well as their attitudes, values" (Webber, 2004). Cameo is one of a number of commercial market research agencies that use a variety of data sources, including Census data, Household Council Tax Band and Property Valuation Data, Consumer Credit data, and residency data from the Electoral Roll, to classify every UK household into distinct marketing types.

Table 1: Summary of Cameo segmentation

Cameo Group	Description	% of UK Households	% of Customers (Period-1)	% of SSD customers (Period-1)	% of Customers (Period-2)
YAS	Young and affluent singles	3.5%	2.2%	2.3%	3%
WRN	Wealthy retired neighbourhoods	3.6%	3.3%	2.7%	3.8%
AHO	Affluent homeowners	11.4%	12.7%	10.8%	12.5%
SPFH	Smaller private family homes	13.7%	15.8%	14.0%	14.3%
CMN	Comfortable mixed neighbourhoods	9.5%	9.4%	8.9%	10.5%
LAF	Less affluent families	13.9%	15.8%	15.6%	13.4%
LASS	Less affluent singles and students	6.1%	5.8%	6.4%	9.3%
PWBCW	Poorer white and blue collar workers	15.7%	14.6%	15.7%	13.3%
PFSPH	Poorer families and single parent households	10.9%	11.2%	13.0%	11.1%
PCT	Poorer council tenants	11.9%	9.2%	10.6%	8.7%

Table 1 shows the population stratification in percentages for the UK per Cameo segment as well as the comparative size of the representative samples in our two SSD datasets, Period-1 and Period-2 as described below. We observe that our segmentation data is a good representation of the UK population. Groups such as AHO, SPFH, and PFSPH (representative of larger household sizes, e.g. families) are slightly over-represented in our sample data compared with the UK population. There is also some small variation in the proportion belonging to some groups (e.g. SPFH or LASS) from Period-1 to Period-2.

We used two separate but related datasets; they both include the weekly sales (by volume), weekly number of customers and weekly prices for the top selling products in the soft drinks category segmented by Cameo. The first dataset (Period-1) contains data over a period of 100 weeks, from June 2014 to May 2016 (i.e., period before the announcement of the tax). The second dataset (Period-2) contains data over a period of 104 weeks from September 2016 to September 2018 (i.e. after announcement and 5 months post implementation of the tax). The data from Period-1 was used to estimate the model parameters and predict sales if the tax had been imposed over that period. We then used the data from Period-2 to measure the actual impact of the tax. For the latter, we compared sales during the 23 weeks from 3rd April to the 4th September 2017 (before the tax) with sales for the corresponding period 12 months later, from the 2nd April to the 3rd September 2018 (immediately after the tax).

3.1.1. Period-1 dataset

Our Period-1 dataset contained a sample with over 700 SSDs. Many of the products in this sample would be non-taxable given their sugar content in the initial period (i.e. < 5mg/litre), or at the time of implementation, as many companies reduced the sugar content to avoid the tax. However, the two top selling brands, Coca Cola and Pepsi, did not change their formulations, so provide us with a suitable sub-group for our analysis (high levy SSDs). The Coke products contained sugar at a level of 10.6g/100ml and the Pepsi products 11.0g/ml. Of the total (64) Coca cola and Pepsi products listed, the top 7, all of which contained sugar levels that would attract the higher tax rate, accounted for 74% of the total volume of Coke and Pepsi sold in the period, 55% of total customers for all SSDs and 63% of the total volume of SSDs sold over the period. The balance of customers/volume included SSDs which would not have attracted the tax (lower tier) represented a very small market share. We therefore used the concept of a composite product for our analysis, focusing on those top 7 high tax SSDs which make the modelling manageable. Table 2 shows summary statistics for the products that make the composite in time Period-1. Table 2 indicates that the products showing high variation in price generally showed high variability in weekly volume and customers.

Table 2: Summary statistics for the composite product set in Period-1

Variable	Variable	Mean (CV) Price (£/Ltr)	Mean (CV) weekly sales volume (Ltrs)	Mean (CV) weekly customers	Mean (CV) weekly purchase per customer (Ltrs)
p_1	Coca Cola 1.75L	£0.68 (36%)	597,530 (40%)	147,323 (37%)	4.1 (11%)
p_2	Coca Cola 500 ml	£2.03 (12%)	56,075 (14%)	81,805 (13%)	0.7 (1%)
p_3	Pepsi 2L	£0.62 (39%)	216,393 62%)	57,821 (63%	3.7 (17%)
p_4	Coca Cola 8 × 330 ml	£1.03 (23%)	161,564 (37%)	42,892 (28%)	3.8 (9%)
p_5	Coca Cola 24 × 330 ml	£0.84 (29%)	351,109 (120%)	36,782 (114%)	9.5 (7%)
p_6	Coca Cola 330 ml	£1.97 (12%)	15,945 (12%)	34,127 (12%)	0.5 (1%)
p_7	Coca Cola 1.25 L	£0.89 (27%)	47,156 (52%)	28,130 (49%)	1.7 (8%)
P	All Coke & Pepsi	£0.76	1,445,772	428,880	3.4

3.1.2. Period-2 dataset

The original dataset for the second period contained slightly over 800 products, of which 47 were Coca Cola and 10 were Pepsi products, all of which were taxable at the high rate. The data was for the 104 weeks from 12th September 2016 to 3rd September 2018. Prices of individual products over this period were far more stable than in Period-1, but many products were removed and new products introduced, generally in smaller pack sizes. For example, 10 Coca Cola products that were selling in 2017 were unavailable during the 2018 post-levy period and 17 new products were introduced after the introduction of the levy. Of the ten highest-volume products in the post-tax period, only one (Pepsi 2l) had significant sales prior to the implementation of the tax and eight were not available at all. For this reason, the composite product in Period-2 cannot be the same as in Period-1. We use Period-2 to analyse the actual changes pre and immediately post tax, hence we can consider all 57 Coke and Pepsi products to be our Period-2 composite dataset. The summary statistics for the composite product in Period-2 are presented in Table 3 at the aggregate level as there are too many products changing over time to present them individually. We present pre and post levy figures separately for comparison.

Table3: Summary statistics for the composite product set in Period-2

Time Period	Mean weekly sales volume (Ltrs)	Mean Price (£/Ltr)	Mean weekly customers	Mean weekly purchase per customer (Ltrs)	Mean weekly sugar purchase per customer (g)
April-Sept 2017 (Pre-levy)	1,658,102	£0.93	545,708	3.038	323
April-Sept 2018 (Post-levy)	1,101,629	£1.29	434,835	2.533	269
%Diff	-33.6%	+39%	-20.3%	-16.6%	-16.8%

Table 3 shows that from the pre-tax period (April-September 2017,) to the post-tax period in 2018 there was a reduction in volume of the composite Coke+Pepsi product purchased of 33.6%, a reduction of 20% in customers purchasing and a reduction of 16.6% in litres per customer purchased. This translates to a reduction of 16.8% in sugar purchased/week as the balance of Coke and Pepsi changed slightly between the two periods. The large reduction in total volume may be associated with a reduction in package volume seen as a result of the tax. The mean price of the composite product set increased by 39% after the introduction of the tax. The expected increase was between 11% and 39% per product, depending on the original price, with the most expensive products expected to see the smallest percentage increases. Hence we observed a high price pass-through for most products.

3.2. Modelling SSDs as a composite good

The key feature of our modelling strategy is that the dependent variable in the empirical analysis is not demand for individual SSDs, but rather the total demand for the composite good representing all SSDs. There are a number of justifications for this choice of approach. First, the focus of the study is the impact of the tax on total demand for SSDs, not the demand for individual SSDs. Second, the form of the econometric model developed here makes it very easy to predict the effect of a proportionate change in the price of every individual good, which is exactly what is required when considering the effect of the sugar tax. Third, using the composite good is a way of side-stepping the problem, raised in Section 3.1.2, of individual products changing over time. Theoretical underpinnings for the composite product approach have been provided by Lewbel (1996) and the approach has been used recently in a context very similar to ours by Pell et al. (2021).

The principal explanatory variable is the weighted average of the prices of the individual products, in the form of an index. Total demand for SSDs is measured using two variables: the total number of customers purchasing one or more SSD; and the total volume (in litres) of SSDs purchased. When we model the former, we are focusing on the "extensive margin", that is, the impact of price changes and taxes on the size of the market; when we model the latter, the focus is on the "intensive margin", that is, the impact on the behaviour of households who are consumers both before and after the change. [6]

The model was estimated for each of the geo-demographic (Cameo) segments separately. However, to minimise notational complexity, we do not include Cameo subscripts in the following specification of the model.

Let N_t be the total number of households who purchase SSDs in week t. N_t is one of our chosen measures of demand for the composite good. The other measure is the total weekly volume of sugary drinks purchased over all the loyalty card holders. To obtain this, we simply sum the weekly volumes (in litres) of individual SSDs. Hence, if there are J sugary drinks, the total volume in period t is:

[6] A problem with the "number-of-customers" variable is that, although we know the number-of-customers purchasing each individual product, we do not know the number-of-customers purchasing the composite product, which is the focus of our analysis. Our chosen measure of number-of-customers is the sum of number-of-customers over products, but we acknowledge that this represents an upper bound for the actual number-of-customers (as a consequence of double-counting of households). In contrast, for total volume of the composite good, we have an accurate measure. We could obtain volume-per-customer by dividing total volume by our estimate of the number-of-customers. However, this would introduce measurement error, and this why we use total volume in the analysis of the intensive margin.

$$V_t = \sum_{j=1}^{J} v_{jt} \quad t = 1,\cdots,T \qquad (1)$$

where v_{jt} is volume purchased of drink j in period t.

To obtain a price index for the composite good, we specify a weighted geometric mean of the J individual product prices. That is, the price of the composite good in period t is assumed to be:

$$P_t = \prod_{j=1}^{J} p_{jt}^{-\alpha_j} \quad t = 1,\cdots,T \qquad (2)$$

where p_{jt} is price per litre of drink j in period t, and α_j, $j = 1,\cdots,J$ are parameters. Note that the α_j parameters capture the importance of each SSD in the budget, and also the responsiveness of consumers to changes in the prices of each SSD.[7] The reason for the minus sign applied to each of the α_j parameters is explained below.

Let Q_t be the measure of demand under consideration; this will be one of the two measures N_t and V_t, defined above. We assume that the demand function for the composite good has the following reciprocal form:

$$Q_t = \frac{\exp(\alpha_0)}{P_t} = \exp(\alpha_0) \prod_{j=1}^{J} p_{jt}^{\alpha_j} \quad t = 1,\cdots,T \qquad (3)$$

where the second equality is obtained using (2). Note that the α_j parameters now have positive signs. Taking logs of both sides of equation (3), we obtain the log-linear equation:

$$\ln Q_t = \alpha_0 + \sum_{j=1}^{J} \alpha_j \ln p_{jt} \quad t = 1,\cdots,T \qquad (4)$$

Equation (4) makes it clear that the that the α_j parameters represent the importance of each individual price in determining demand for the composite good.

[7] The assumption that the α_j parameters are fixed over time guarantees the exogeneity of the price index defined in (2), in a model of composite consumption.

Another important measure is the sum of all the α_j parameters:

$$\eta = \sum_{j=1}^{J} \alpha_j \quad (5)$$

The quantity η defined in equation (5) has the interpretation of a price elasticity: if the prices of *all J* of the component goods rise by 1%, the quantity demanded of the composite good will change by approximately η%. We expect η to be negative.

We also include a time trend variable to allow for changes in tastes over the sample period. There is an upward spike in demand during the Christmas periods for most goods including all carbonated drinks, so to allow for abnormal purchasing behaviour a set of three "Christmas dummies" (C_1-C_3) are also included, representing the first, second and third week of the Christmas period. Finally, we add an error term. The resulting linear regression equation is:

$$\ln Q_t = \alpha_0 + \sum_{j=1}^{J} \alpha_j \ln p_{jt} + \beta t + \sum_{k=1}^{3} \gamma_k C_{kt} + \varepsilon_t \quad t = 1, \cdots, T \qquad (6)$$

Under the tax, companies pay 18p per litre if the product contains more than 5g of sugar per 100 millilitres, and 24p per litre if it contains 8g of sugar per 100 millilitres. Assuming that the tax is passed fully to consumers, the post-tax prices (per litre) can be easily computed as:

$$p_{jt}^{TAX} = p_{jt} + \tau_j \quad j = 1, \cdots, J \quad t = 1, \cdots, T \qquad (7)$$

where τ_j is either 0, 0.18, or 0.24, depending on the rate at which drink *j* is being taxed. For all our 7 products in Period-1 (used to build the model) the amount of sugar is higher than 8g per 100 ml hence the higher levy applies (i.e., £0.24 per litre). Assuming that all of the tax is passed to consumers, the new prices are presented in Table 4 together with the percentage increase over the price per litre they represent. Note that given the nature of the tax as a fixed amount per litre, percentage increases are much more noticeable for the products with lower prices per litre.

Table 4: Prices in Period-1 before and after the tax, assuming the tax is fully passed on

Product	Avg Price before tax (£/Ltr)	Avg Price after tax (£/Ltr)	% Increase
Coca Cola 1.75 L	0.68	0.92	35.5%
Coca Cola 500 ml	2.03	2.27	11.8%
Pepsi 2 L	0.62	0.86	38.6%
Coca Cola 8 X 330 ml	1.03	1.27	23.4%
Coca Cola 24 X 330 ml	0.84	1.08	28.5%
Coca Cola 330 ml	1.97	2.21	12.2%
Coca Cola 1.25 L	0.89	1.13	26.9%

Finally, we predict the impact of the tax by combining the estimates from equation (6) with the assumed after-tax prices from equation (7) to obtain the predicted consumption under a scenario in which the tax is applied over the sample period. This prediction is obtained using:

$$\ln \hat{Q}_t = \hat{\alpha}_0 + \sum_{j=1}^{J} \hat{\alpha}_j \ln p_{jt}^{TAX} + \hat{\beta} t + \sum_{k=1}^{3} \hat{\gamma}_k C_{kt} \quad t = 1, \cdots, T \tag{8}$$

where hats indicate estimates of the parameters in equation (6), and p_{jt}^{TAX} $j = 1,\ldots,J$ are defined in equation (7). Equation (8) gives an unbiased prediction of $\ln Q_t$. To convert this to an unbiased prediction of Q_t, we apply Duan's smearing method (Duan, 1983). This is a non-parametric method that provides consistent predictions whatever the distribution of the error term in equation (6).

We compare this prediction of the tax effect from Period-1 data to the actual consumer response observed after the tax using Period-2 dataset. These comparisons are made in both absolute and relative terms, separately, for each Cameo segment.

4. Results

In accordance with equations (5) and (6), the full results of the model, showing the coefficients/elasticities for each term, are presented in the on-line appendix. In order to test the accuracy of the model we also partitioned the Period-1 dataset into two equal halves. The first 50 data points we used to generate another model, which we tested against the second 50 data points to obtain the in-sample and out-of-sample R^2 respectively. The out-of-sample R^2 values are significant in each case. These values are shown in the on-line appendix.

A summary of results with the overall price elasticities for each of the Cameo segments are presented in Table 5. Results include elasticities obtained from equation (5); volume and customers pre-tax (Pre-tax P-1); predictions obtained using equation (8) after applying Duan's smearing (Post-tax P-1), assuming the full pass-through of the tax; % change predicted (% change P-1) ; and actual percentage change observed in Period 2 (% change P-2). Those are presented for volume (intensive margin) and number of customers (extensive margin).

Table 5: Predicted impact in Period-1 (P-1) of the sugar tax and actual impact in Period-2 (P-2) on mean weekly total volume and mean weekly number of customers by geo-demographic (Cameo) segment

Cameo Segment	Mean weekly Volume (l)					Mean weekly Customers				
	Elasticity	Pre-tax (P-1)	Post-tax (P-1)	% change (P-1)	% change (P-2)	Elasticity	Pre-tax (P-1)	Post-tax (P-1)	% change (P-1)	% change (P-2)
YAS	-1.6	32,941	20,195	-38.7%	-24.9%	-0.9	9,678	7,231	-25.3%	-12.2%
WRN	-1.4	39,817	24,551	-38.3%	-31.2%	-0.6	11,576	8,719	-24.7%	-17.8%
AHO	-0.2	147,166	113,213	-23.1%	-31.3%	-0.02	46,145	37,755	-18.2%	-19.2%
SPFH	-0.9	189,615	128,641	-32.2%	-32.2%	-0.3	60,078	46,733	-22.2%	-20.2%
CMN	-1.1	126,803	86,451	-31.8%	-32.1%	-0.6	38,318	28,668	-25.2%	-18.9%
LAF	-0.7	221,983	156,619	-29.4%	-35.3%	-0.3	66,724	51,837	-22.3%	-23.4%
LASS	-1.4	95,763	59,019	-38.4%	-33.2%	-0.8	27,465	20,618	-24.9%	-18.7%
PWBCW	-1.2	229,521	151,181	-34.1%	-35.0%	-0.6	67,445	51,497	-23.6%	-21.0%
PFSPH	-1.2	196,077	129,748	-33.8%	-35.0%	-0.6	55,839	42,356	-24.1%	-20.9%
PCT	-1.9	166,087	100,127	-39.7%	-35.7%	-1.4	45,613	31,515	-30.9%	-21.9%
Total		1,445,772	969,744	-32.9%	-33.6%		428,880	326,930	-23.8%	-20.3%

Looking at the predicted % change (P-1) and actual % change (P-2), we note that the predictions are good for most groups but over-predict for YAS and WRN and to a lesser extent for LASS. Recall from Table 1 that those Cameo groups contain the smallest proportions of households and are under-represented in our dataset, which may make them harder to model. Focusing on other groups, we see that price elasticities in Table 5 tend to rise over the Cameo groups;

that is, less affluent groups tend to be more sensitive to price changes. The least affluent group (PCT) has the largest elasticities: -1.9 at the intensive margin; and -1.4 at the extensive margin.

In Period-1, we observe that the tax is predicted to decrease both volume and number of customers for all Cameos, with the rate of decrease generally higher for less affluent groups. The predicted percentage reduction in volume is higher than the predicted reduction in number of customers. This may be due to the additional reduction in volume that was caused by manufacturers reducing the volume sold as a result of the tax (e.g. 1.75 L bottle becomes 1.50 L). Table 5 shows that Cameos YAS, WRN, LASS, and PCT are predicted to be highly affected by the tax, as they show greater reductions both in volume and number of customers. The composite product had a mean sugar content of 10.66g/100ml (weighted by the balance between Coke and Pepsi in Period-1), and we therefore expect the changes in sugar purchased to be the same as the predicted volume change i.e. 32.9%. For the actual changes (P-2) portrayed in Table 5, we see an inverse relationship between the affluence of the Cameo segments and % change in volume purchased. This means that less affluent groups show a stronger response.

A further test of this trend is obtained by performing a weighted linear regression with % change in volume purchased as the dependent variable and Cameo group as the explanatory variable, and with initial mean of total volume or number of customers as weights.[8] The results of these regressions are presented in Table 6.

[8] The regressions whose results are shown in Table 6 implicitly assume the ranking of Cameo groups that is routinely assumed by others working with Cameo data (see e.g. Revoredo-Giha et al., 2009), and also that there is an equal distance between ranks. The latter assumption could be avoided if we used a non-parametric measure of association, such as Spearman correlation. However, the great advantage of using a regression in this case is that we are able to incorporate the weights of each Cameo group (using weighted regression) and this weighting is important. We therefore prefer the regression approach for the current purpose, even though we accept that the assumptions underlying this regression may not be fully met.

Table 6: Results of weighted linear regression. Dependent variable: % impact of tax; independent variable: cameo group identifier (1-10)†

		Volume	Customers
Predicted	Constant	-26.43***(3.73)	-18.56*** (2.29)
	Cameo	-1.02*(0.55)	-0.83** (0.34)
	N	10	10
	*R*2	0.30	0.43
Actual	Constant	-28.98***(1.11)	-17.37*** (1.62)
	Cameo	-0.72 ** (0.16)	-0.47*(0.24)
	N	10	10
	*R*2	0.71	0.32

† Standard errors in brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The large negative intercept in these regressions indicates that even the most affluent groups show a large negative response to the tax. The (significantly) negative slope confirms that the response to the tax is even higher for less affluent groups.

5. Discussion and Conclusions

In this paper we use a rich data source from a major UK retailer to estimate a demand equation over the 100 weeks before May 2016 for the most popular SSDs treated as a composite good. We have utilized these estimates to predict the behavioural response after the tax, not only at aggregate level, but also across ten geo-demographic (Cameo) consumer segments. We compare our predicted results with actual consumer response during the five months after the implementation of the tax and calculated both the actual effect of the tax at aggregate level and by geo-demographic group. This allowed us not only to assess the effectiveness of the policy, but also to corroborate our predicted results.

We show that, both in our predictions and the actual post-tax analysis: 1) mean weekly volume of SDDs purchased (i.e. intensive margin) reduced by 32.9% and 33.6%, respectively; and 2) mean weekly number of consumers that purchased SSDs (i.e. extensive margin) reduced by 23.8.% and 20.3%, respectively. A differential analysis, enabled by the loyalty card data, shows that the tax had a significant impact on all consumer segments. However, the greatest impact was felt by the less affluent consumer segments.

We find, first, that the pattern of purchasing of soft drinks varies significantly between consumer groups, with less affluent groups tending to be more likely to purchase SSD. We then analyse the demand for the composite good, made up of the 7 most popular taxable SSD in Period-1 and find that it appeared to be more sensitive to some prices than to others, according to the estimates obtained.

Based on demand in Period-1, we generate predictions of total mean weekly volume purchased and total number of customers purchasing after the introduction of the tax. Our predictions show a clear reduction in purchasing post-tax, both in terms of total volume and number of customers. We also see that this negative impact of the tax is not homogeneous across Cameo segments, being more marked for less affluent segments. Of all the Cameo groups, it was the least affluent (Poorer Council Tenants) that showed the highest percentage decrease in purchasing (at both intensive and extensive margins).

Focusing on the percentage change in purchases, we predicted an overall decrease of 32.9% in volume (intensive margin) and a decrease of 23.8% in the number of customers (extensive margin), of the composite good across all Cameo groups. Overall, these results suggest that the policy intervention would have a significant impact on SSD purchasing as expected by the policy makers.

This predicted overall decrease is consistent with the observed change in purchasing behaviour found in the 5 months post-levy, which was a 33.6% reduction in volume and a 20.3% reduction in customers. Our predictions for the individual Cameo groups showed a very good match to actual reductions for SPFH, CMN, PWBCW, PFSPH, and to an extent for PCT. We over-predicted for YAS and WRN, which are small groups and for LASS. We under-predicted for AHO and LAF.

The two groups for which the prediction error is highest appear to be the first two (YAS and WRN); as previously remarked, this may be associated partly with the higher sampling error arising from these being the smallest groups but may also be a result of unobservable changes in the composition and/or the behaviour of these groups between the period of estimation and the time at which the tax was introduced. Overall, we can conclude that the model appears to be a good tool to predict buyer behaviour after the introduction of the levy.

We have made an important distinction between the extensive and intensive margins. The impact of the levy at the extensive margin is measured by the decrease in the number of customers resulting from the introduction of the levy, and as previously noted this is 20.3%. The impact on total volume is clearly higher than this, at 33.6%. The fall in volume per customer which in real terms was about 16.6% is what is meant by the intensive margin.

Our results show the policy intervention had more impact on the less affluent (higher rank) Cameos compared with the more affluent ones (lower rank). However, because the group sizes are quite different, we used a weighted regression to test this result. The results confirmed the apparent pattern. When observing the effect actually seen in Period-2, which included post-tax data this differential impact was indeed significant.

It is worth noting that the levy being passed in full was expected to have an effect of an increase in the average price per litre of between 11% and 39% depending on the original price, with the more expensive products expected to see the smallest percentage increases. In fact, in the observed period after the implementation, the average price of the composite good increased by about 39%. This is consistent with the tax being passed in full and some over-shifting of the tax, resulting in prices per litre higher than expected after the introduction of the tax.

Our results are broadly consistent with the results reported by Pell et al. (2021) who also analysed the effect of the tax in the UK one year after the implementation of the levy. They found the high-tier SSDs (the comparable composite) purchased volumes decreased by 44.3%

and sugar purchased decreased by 45.9%. This is in line with our total observed 33.5% decrease in volume. The differences might be due to the diverse time period considered in the analysis (five months vs. one-year post-tax) and their population being biased towards lower income groups for which we observed the largest reductions.

Finally, we conclude that loyalty card data affords us the possibility of a unique insight into consumer purchasing behaviour and may enable us to study the impact of other policies (involving taxation or other measures) aimed at influencing consumption behaviour on the food and drinks market. This is one of the first evaluations of the impact of the sugar tax policy on the UK and provides positive lessons for other ’sin taxes’.

5.1. Policy implications

Overall, from our predictions and from the observed period post-tax implementation, we conclude that the soft drinks tax is likely to have an overall sizeable negative impact on SSD purchases (initially −32.9% in volume and −23.8% in customer numbers). The most important aspect of our results is that the impact of the policy intervention is likely to be greatest on those consumer segments with the highest propensity to purchase SSDs, who are also segments characterised as less affluent – including single person households and less affluent families. This indicates that the policy is not regressive, as the least affluent consumers would not pay a higher penalty by not changing their behaviour; instead, they appear to be more price sensitive and are reducing their purchases of SSDs to a greater extent than more affluent shopper segments. This is an important point for governments and for society in general as further “sin taxes” are considered.

Our predicted negative impact on SSD purchasing behaviour is clearly a desirable result from the point of view of policy-makers, who introduced the soft drinks levy with the intention of reducing the consumption of SSD and thereby the sugar intake of the UK population.

5.2 Limitations

An important limitation of our data is that we are able to observe Cameo segment purchasing behaviour but not the behaviour of individual customers, nor their consumption behaviour. We assume they consume what they purchase but this could be over varying periods of time. Higher purchasing may be accounted for by a large family structure resulting in lower individual consumption. Indeed, as households may vary in size and composition and we do not have precise details for those, it would be difficult to translate our analysis to an individual prediction per consumer or to see how it affects population across specific demographic characteristics (e.g. old versus young people).

References

Aguilar, A., Gutierrez, E., & Seira, E. (2019). The effectiveness of sin food taxes: evidence from Mexico. Instituto Tecnológico Autónomo de México (ITAM) Working Paper.

Allcott, H., Lockwood, B. & Taubinsky, D. 2019. Regressive Sin Taxes, with an Application to the Optimal Soda Tax, The Quarterly Journal of Economics, 134 (3), 1557–1626.

Alsukait, R., Wilde, P., Bleich, S. N., Singh, G., & Folta, S. C. (2020). Evaluating Saudi Arabia's 50% carbonated drink excise tax: Changes in prices and volume sales. Econ Hum Biol, 38(100868), 10-1016.

Arteaga, J. C., Flores, D., & Luna, E. (2017). The effect of a soft-drink tax in Mexico: a time series approach.

Berardi, N., Sevestre, P., T´epaut, M. & Vigneron, A. 2016. The impact of a "soda tax" on prices: evidence from french micro data, Applied Economics, 48 (41), 3976–3994.

Birkin, M. & Clarke, G. 1998. GIS, geodemographics, and spatial modeling in the U.K. financial service industry, Journal of Housing Research 9 (1), 87–111.

Bollinger, B., & Sexton, S. (2018). Local excise taxes, sticky prices, and spillovers: evidence from Berkeley's soda tax. Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax (January 12, 2018).

Briggs, A., Mytton, O., Kehlbacher, A., Tiffin, R., Elhussein, A., Rayner, M, Jebb, S, Blakely, T, Scarborough, P. 2017. Health impact assessment of the UK soft drinks industry levy: a comparative risk assessment modelling study, The Lancet Public Health 2 (1), 15–22.

Briggs, A., Mytton, O., Kehlbacher, A., Tiffin, R., Rayner, A. & Scarborough, P. 2013 Overall and income specific effect on prevalence of over-weight and obesity of 20% sugar sweetened drink tax in uk: econometric and comparative risk assessment modelling study, BMJ, 2013;347:f6189

Cameo (2020). https://www.transunion.co.uk/product/cameo

Capacci, S., Allais, O., Bonnet, C., & Mazzocchi, M. (2019). The impact of the French soda tax on prices and purchases. An ex post evaluation. PloS one, 14(10), e0223196.

Castelló, J. V. López-Casasnovas (2018). Impact of SBB taxes on consumption. CRES-UPF Working Paper, 201804, 110.

Cawley, J., & Frisvold, D. E. (2017). The pass-through of taxes on sugar-sweetened beverages to retail prices: the case of Berkeley, California. Journal of Policy Analysis and Management, 36(2), 303-326.

Cawley, J., Willage, B., & Frisvold, D. (2018). Pass-through of a tax on sugar-sweetened beverages at the Philadelphia International Airport. Jama, 319(3), 305-306.

Cawley, J., Frisvold, D., Hill, A., & Jones, D. (2019). The impact of the Philadelphia beverage tax on purchases and consumption by adults and children. Journal of Health Economics, 67, 102225.

Colchero, M., Popkin, B., Rivera, J. & Ng, S. 2016. Beverage purchases from stores in mexico under the excise tax on sugar sweetened beverages: observational study, BMJ 2016;352:h6704

Duan, N. 1983. Smearing estimate: A nonparametric retransformation method, Journal of the American Statistical Association 78 (383), 605–610.

Dubois, P., Griffith, R. & O'Connell, M. 2017. How well targeted are soda taxes?, Toulouse School of Economics (TSE) Working Paper, December 2017, 17-868.

Etilé, F., Lecocq, S., & Boizot-Szantai, C. (2018). The Incidence of Soft-Drink Taxes on Consumer Prices and Welfare: Evidence from the French" Soda Tax".

Falbe, J., Thompson, H. R., Becker, C. M., Rojas, N., McCulloch, C. E., & Madsen, K. A. (2016). Impact of the Berkeley excise tax on sugar-sweetened beverage consumption. American journal of public health, 106(10), 1865-1871.

Falbe, J., Rojas, N., Grummon, A. H., & Madsen, K. A. (2015). Higher retail prices of sugar-sweetened beverages 3 months after implementation of an excise tax in Berkeley, California. American journal of public health, 105(11), 2194-2201.

Felgate, M., Garcia, M, & Fearne, A. 2012. Using supermarket loyalty card data to explore the moderating impact of shopper segments on promotional response, International Journal of Market Research, 54 (2), 221 – 240.

Geyskens, K., Grigoriev, A., Holtrop, N., & Nedelko, A. (2018). Optimal policy design for the sugar tax. arXiv preprint arXiv:1810.07243.

Global Food Research Program (2020). Sugary drink taxes around the world.

Goiana-da-Silva, F., Nunes, A. M., Miraldo, M., Bento, A., Breda, J., & Araújo, F. F. (2018). Using pricing policies to promote public health: the sugar sweetened beverages taxation experience in Portugal. Acta medica portuguesa, 31(4), 191-195.

Gonçalves, J., & Dos Santos, J. P. (2020). Brown sugar, how come you taste so good? The impact of a soda tax on prices and consumption. Social Science & Medicine, 264, 113332.

Griffith, R., O'Connell, M., & Smith, K. (2018). Corrective taxation and internalities from food consumption. CESifo Economic Studies, 64(1), 1-14.

Grogger, J. 2017. Soda taxes and the prices of sodas and other drinks: Evidence from mexico, American Journal of Agricultural Economics, 99 (2), 481–498.

Gruber, Jonathan. "Smoking's internalities." Regulation 25 (2002): 52.

Kantar World Panel (2020). https://www.kantarworldpanel.com/global

Lewbel, A. (1996). Aggregation without separability: a generalized composite commodity theorem. The American Economic Review, 524-543.

Malik, V. S., Popkin, B. M., Bray, G. A., Després, J. P., Willett, W. C., & Hu, F. B. (2010). Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. Diabetes care, 33(11), 2477-2483.

Malik, V., Schulze, M. & Hu, F. 2006. Intake of sugar-sweetened beverages and weight gain: a systematic review, The American Journal of Clinical Nutrition, 84 (2), 274–288.

Mora, T. (2019). Do Consumers Respond to "Sin Taxes" Heterogeneously? New Evidence from the Tax on Sugary Drinks Using Retailer Scanner Data. In 2019 World Congress on Health Economics. iHEA.

Nakamura, R., Mirelman, A. J., Cuadrado, C., Silva-Illanes, N., Dunstan, J., & Suhrcke, M. (2018). Evaluating the 2014 sugar-sweetened beverage tax in Chile: an observational study in urban areas. PLoS medicine, 15(7), e1002596.

Ng, S., Mhurchu, C., Jebb, S. & Popkin, B. 2012. Patterns and trends of beverage consumption among children and adults in great britain, 1986– 2009, The British Journal of Nutrition, 108 (3), 536–551.

Pell, D., Mytton, O., Penney, T. L., Briggs, A., Cummins, S., Penn-Jones, C., ... & Adams, J. (2021). Changes in soft drinks purchased by British households associated with the UK soft drinks industry levy: controlled interrupted time series analysis. BMJ, 372.

Revoredo-Giha, C., Wrieden, W., Kupiec-Teahan, B., Leat, P., Fearne, A., & Cacciolatti, L. (2009). Analysis of red and processed meat purchases in Scotland using representative supermarket panel data. *Food Standards Agency in Scotland, Aberdeen*.

Rojas, C., & Wang, E. Y. (2017). Do taxes for soda and sugary drinks work? Scanner data evidence from Berkeley and Washington. Scanner Data Evidence from Berkeley and Washington (September 23, 2017).

Seiler, S., Tuchman, A. & Yao, S. 2018. The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects, SSRN Electronic Journal, doi:10.2139/ ssrn.3302335.

Silver, L., Ng, S., Ryan-Ibarra, S., Taillie, L., Induni, M., Miles, D., Poti, J. & Popkin, B. 2017. Changes in prices, sales, consumer spending, and beverage consumption one year after a tax on sugar-sweetened beverages in berkeley, california, US: A before-and-after study, PLOS Medicine 14 (4), 1–19.

Taylor, R. L., Kaplan, S., Villas-Boas, S. B., & Jung, K. (2019). Soda wars: The effect of a soda tax election on university beverage sales. Economic Inquiry, 57(3), 1480-1496.

Webber, R. 2004. Papers: Designing geodemographic classifications to meet contemporary business needs, Interactive Marketing 5 (3), 219–237.

Yamoah, F. Fearne, A. & Duffy, R. 2014. Exploring Supermarket Loyalty Card Analysis To Identify Who Buys Fairtrade, The International Review of Retail, Distribution and Consumer Research, 24(3), 328-346.

Zhong, Y., Auchincloss, A., Lee, B. & Kanter, G. 2018. The short-term impacts of the philadelphia beverage tax on beverage consumption, American Journal of Preventive Medicine 55, doi:10.1016/j.amepre.2018.02.017.