Merger Simulation and Policy Evaluation in Markets for Sinful Goods: Insights from Experimental and Empirical Demand Estimation

A thesis submitted to the School of Economics at the University of East Anglia in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Janithe Siriwardana

September 2024

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

The three interconnected studies presented here explore various economic aspects of ‘sin goods’ - notably alcohol and sugar-sweetened beverages (SSBs). The first of these introduces a novel methodology for merger analysis, focusing on the beer industry, by combining data from online experiments to augment real-world data in a structural demand estimation model. Our work provides a cost-effective and timely alternative for conducting merger simulations by addressing the challenges posed by traditional data-intensive approaches while estimating market variables in line with previous studies. We then shift focus into public health policy analysis, evaluating the effectiveness of a recent high in fat, sugar and salt (HFSS) location restriction policy in England through a difference-in-difference model. By analysing aggregate sales data from the UKs largest supermarket chain, we show that although the policy did reduce sales of some sugary drinks, ineffective targeting in terms of storetypes and point-of-consumption means that overall the policy had no significant impact on the sales of SSBs in England. Finally, we combine the structural demand techniques of the first paper, with the data and policy from the second, to investigate the welfare effects of the HFSS location restrictions and simulate the effects of alternate versions of the policy on both sales and welfare. Although we find some small heterogeneous effects between demographic groups, with Families suffering a small decrease in consumer welfare from the policy and an extension, at the aggregate level we find there are almost no changes to welfare, whatever the policy environment.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents

List of Figures

List of Tables

Acknowledgements

This thesis and the entire graduate experience would not have been possible if not for the help and support of a great number of individuals. First and foremost, I would like to thank my supervisors Farasat Bokhari and Stefan Penczynski. Effective supervision is the foundation of any successful PhD and your support and guidance throughout the project has been invaluable. We had some bumps along the way but your confidence in me inspired the belief and determination in myself to complete the work presented here.

I would like to acknowledge the UEA School of Economics and Stefan's research grant for funding and financial support, without which I could not have undertaken this endeavour.

Throughout my 4 years, I have had the pleasure of being a member of two wonderful multi-disciplinary research centres at UEA; the Centre for Behavioural and Experiment Social Sciences (CBESS) and the Centre for Competition Policy (CCP). Having presented my work several times over to both of these audiences and attending the CCP conference for multiple years, I am grateful for the spirited feedback, comments and discussion of my work from all participants of these groups.

I am grateful to everyone, internally and externally, that provided valuable feedback for any of the papers at seminars, conferences and teach-ins. Your insightful observations and constructive critiques have greatly contributed to the development and refinement of this thesis. I deeply appreciate the time and effort taken to engage with the material and share your expertise. Your contributions have helped shaped this work, and for that, I am thankful.

I would like to thank my friends and family for their support throughout this process. To my friends, you have kept me sane throughout this whole endeavour and that is an achievement in itself!

To the Home Hemo Team at the Freeman Hospital - this experience would have been immeasurably more difficult without all your hard work and diligence. Thank you.

Finally, to my family who have always believed in me. Thank you for standing by me every step of the way.

1

Preface

This thesis collects three standalone papers that overlap in the techniques or data sets being used. As a result there is some repetition between Chapters. Chapter 2 and Chapter 4 both rely on methods in demand estimation and build on Berry et al. (1995) style techniques. Accordingly, both have some discussion of the demand and supply side equations used in these models. Nevertheless, there are differences in these repeated sections. For example, in Chapter 2 we use individual-level data and develop the model accordingly. Chapter 4 uses the more traditional aggregate level data.

Similarly, Chapter 3 and Chapter 4 analyse the same policy and use the same data sets, albeit from different angles, with one a reduced-form difference-in-difference and the other a more refined structural demand estimation with welfare calculations and counterfactual exercises on new policies. While there is overlap in the raw data, in answering different questions we focus on different elements of the data set. To improve readability as presented here, we make small changes to the versions to be submitted for publication consideration. We introduce each chapter in more detail below.

Chapter 2 examines the complexities of merger simulation, in the context of beer markets, through a novel approach that combines data from an online experiment with real-world aggregate data to estimate demand parameters via modified random coefficient models. We show our methodology offers a cost-effective and efficient alternative for conducting merger assessments using structural modelling, where traditional data collection methods and analysis are either time-consuming, expensive or impractical. The Chapter also delves into the nuances of incentivization and the role of branding in experimental designs, ultimately providing insights into the practical application of these findings in real-world scenarios. The study examines how brand presence influences consumer decisions in a controlled environment, offering insights into the potential biases that can arise when consumers are not making real-world purchases. By comparing intra-brand and inter-brand scenarios, the Chapter sheds light on how different experimental designs can impact the accuracy of demand estimates, which is critical for understanding the potential effects of mergers in concentrated markets.

Chapter 3 transitions to the public health policy arena, specifically evaluating the effectiveness of England's high in fat, sugar, and salt (HFSS) location restriction policy, which targets the placement of unhealthy products within retail environments in a bid to combat obesity, which Public Health England describes as the single biggest public health issue facing the country. The Chapter's empirical analysis draws from extensive data provided by one of the UK's largest supermarket chains, to assess how these restrictions influence consumer purchasing behavior and the overall sales of HFSS products, through analysing sugar-sweetened beverages (SSBs).

This Chapter shares methodological similarities with Chapter 2, in its focus on demand estimation. Both Chapters analyse consumer responses to changes in market conditions—whether due to a merger or a public health policy—using empirical data to draw conclusions about the effectiveness of these interventions. The analysis in Chapter 3 reveals that while the HFSS policy has led to some reductions in the sales of SSBs, its effectiveness is limited by the exemptions for certain store types. This finding underscores the complexity of influencing consumer behavior through policy measures and the potential for unintended consequences as a result.

Finally, Chapter 4 delves deeper into the welfare implications of the HFSS location restriction policy. Using a more traditional aggregate structural demand model in Chapter 2 alongside the findings from Chapter 3 we assess how welfare was changed before and after the policy at the aggregate level and by demographic subcategories using the same supermarket data from Chapter 3. We then simulate several counterfactual scenarios by altering the parameters of the policy's application to observe what would happen to welfare and sales under these alternatives. Although there are some heterogeneous effects on welfare by demographic groups with Families the most effected, overall we find only a small change in welfare and sales. Balanced against the cost of policy alterations we feel the gains for the policy are insufficient and do not recommend extending the policy further.

A significant linkage between Chapter 4 and the earlier Chapters is the focus on sin goods and the implications of their consumption on public welfare. Both alcohol and SSBs are subject to significant public scrutiny and regulation due to their health impacts. The thesis explores how economic tools and models can be used to evaluate policies aimed at reducing the consumption of these goods, whether through market regulation (as in the case of mergers) or direct public health interventions (as in the HFSS policy).

By comparing and contrasting the welfare effects across different consumer demographics, Chapter 4 enhances our understanding of the distributional impacts of policies and market changes. It highlights how demand for sin goods is not uniform across the population and how policy measures must account for these differences to be effective. This Chapter also reinforces the importance of combining empirical data with robust economic modeling to accurately assess the broader social implications of market interventions, a theme that is central to the entire thesis.

Together, these Chapters form a cohesive body of work that not only advances theoretical knowledge but also offers practical tools for policymakers and industry practitioners. The methodologies developed and applied in this thesis have broad implications, offering new ways to tackle complex economic issues in merger analysis, public health policy, and welfare economics. By bridging experimental and empirical approaches, this thesis contributes to the ongoing discourse on how best to evaluate and implement policies that promote both economic efficiency and public well-being.

2

Merger review using online experiments

Abstract

Merger simulation is a complex exercise and is difficult to implement during merger assessment due to data and time requirements. We use data from an online experiment to estimate demand parameters for the beer market and combine this with aggregate national level data on prices, shares and attributes of beers. These allow us to calculate elasticities, markups, and marginal costs for a set of real products which compare well to reported estimates in the literature. Our proposed method offers a fast and cost-effective way of implementing modern IO methods for evaluating cases in real time.

2.1 Introduction

Merger evaluations often use own and cross-price elasticities and/or diversion ratios. Additionally, a merger simulation uses both demand and cost side parameters. Yet the challenges associated with estimating models derived from the modern workhorse random coefficients mixed logit (RCML) of Berry et al. (1995, henceforth BLP) and Nevo (2000) are numerous.

To circumvent these difficulties with BLP type models, we borrow from the experimental literature on discrete choice experiments, to construct a stated preference (SP) experiment in which subjects are required to make repeated choices on sets of beers. We then combine these estimates derived from the lab with real world data on prices and product characteristics taken from a single market (in our case US national figures from 2019). This step is crucial in our ability to conduct counterfactual analysis such as merger simulations because it incorporates market clearing equilibrium conditions to the experimental demand parameters such that the merger simulations have real world interpretations. As far as we know, we are the first to apply this combination of experimental and real world data in this setting. Although our methods are not immune to the challenges of data collection, they do not require the highly detailed, multi-time-period, multi-market data sets typically required of empirical demand estimation.

We show that an experiment of this type, under the right circumstances, is quick and cheap to implement; data can be collected and analysed in weeks, if not days, rather than months. Our second contribution involves understanding the issue of incentivisation and the use of brands in lab experiments. Incentivisation in this context has a specific meaning; consumers do not face any consequences for their choices in the form of altered payoffs (see section 2.2 for more detail). The literature on SP experiments often discusses labelled versus non-labelled (or branded versus non-branded) products in the choice sets consumers see as an option within the experimental design (Louviere et al., 2000). It may be natural to think that since brands play a part in real purchase decisions, brand effects should be included in the experiment. Alternatively, it could be argued that omitting brands from the experiment would lead to unrealistic demand estimates and elasticities. We show that in lab experiments, where incentivisation is not possible, there is no incentive for subjects to *not* engage in cheap talk when brands are present, such that the non-branded experiment is better suited to elasticity calculations. This is an important finding and as far as we are aware none of the other papers in this direct domain make this distinction.

Beyond speed, the methodology addresses other challenges associated with empirical models which require at a minimum, aggregate level data of purchases obtained from a single market. Data from several markets is advantageous because it results in greater variation in relative prices of the products and/or products offered. However, this can be time consuming and costly to obtain. In many industries there is simply a dearth of information on sales volumes and prices; for example Moshary et al. (2022) use an SP experiment to estimate demand for handguns because there is no centralized database that contains information about either individual-level or aggregate gun purchases matched with prices. Aggregate proxies for purchases that have been used in previous research are neither detailed to the gun model nor matched with prices and so are not suitable for demand estimation either. The models also require data on demographic variables which at best can only be approximated by good census data. Finally, prices are often correlated with unobserved variables resulting in endogeneity; this requires a set of relevant and exogenous instrument variables to solve.

These issues present challenges for any researcher attempting to estimate demand, but particularly for an antitrust agency evaluating a merger in real time, they represent significant hurdles to a timely analysis. Imthorn et al. (2016), from the Netherlands Authority for Consumers and Markets (ACM), is the only paper that we have found to have put a similar method in practice during several merger cases including agricultural fertiliser, hospitals, and bakery products, preferring them over hypothetical surveys (Imthorn et al., 2016). They do however, consistently apply brands in their methodology which as we show can be problematic. They also specified the conditions for merger simulation only after the experiments were conducted which caused problems including lack of variation in price. We show that by considering the purpose of the demand estimation, we can avoid most of these types of issues through careful design of the experiment. Additionally, given technological improvements since the original Dutch paper, we provide ideas for further adaptation in section 2.5.

From the repeated choices obtained in the experiment, we approximate mixed logit choice probabilities and estimate demand parameters to quickly and cheaply enable initial merger simulations. Prices in the experiment are randomly assigned, eliminating the need for instruments. Similarly, we generate variation in product characteristics and repeatedly randomize the assignment of choice sets to consumers to identify the model's demand parameters. To evaluate the effect of branding in the experiment, we conduct two treatments. In the first, choice sets all feature products from a single brand. We call this treatment *intra-brand*. In the second treatment, choices sets are constructed with products from different brands. We call this treatment *inter-brand*. More detail regarding the treatments can be found in section 2.2.3. In the intra-brand experiment, we define a matrix of attributes and levels that yields a set of pseudo-products used to estimate demand parameters. For the inter-brand treatment, all the product characteristics except price are taken from real products. In each treatment, each individual was presented with four alternatives from the set of 18 possible products, with each alternative randomly priced at one of the three values. Individuals were instructed to select their preferred option in each choice set.

We combine this micro-data with aggregate level data on real products to obtain a price elasticity of demand matrix for the product set as well as associated price-cost margins, creating an alternative tool for competition economists to use. As in BLP we augment our model predicted shares with a vector of unobserved heterogeneity parameters so that the model predicted share are equal to observed market shares. This is done using a version of BLP's contraction mapping, described in more detail in section 2.3 and 2.4. For industry/regulatory practitioners, a further advantage of an SP experiment is that once an appropriate experimental design has been conceived it can be retooled for many different products/situations and implemented quickly.[1] An

[1]Regulators often have tight deadlines when conducting merger reviews. The UK

online survey using existing platforms such as Prolific or Amazon Mechanical Turk can produce thousands of observations from very specific groups of consumers in a matter of days. However, our methodology can also be used in situations where empirical data does not exist or would be very difficult to collect (see the handgun example above) or as a complementary method to other merger analysis tools such as upward pricing pressures, diversion ratios or qualitative measures. Further still, some competition authorities around the world including the CMA already use surveys and questionnaires in other forms, often qualitative, during merger assessment such that introducing this methodology will not be technically burdensome.

2.1.1 Literature

Following BLP's seminal work on RCMLs, a range of papers have sought to improve the performance of these models. Nevo (2000, 2001) attempts to guide practitioners through the model using the ready to eat cereal market as an example. Petrin (2002) uses micro moments obtained using consumer level data to augment market-level data and estimate a demand model for mini-vans. There is also a related literature on discrete choice models (e.g. Train (2009)) from which we borrow heavily. Elsewhere, Reynaert and Verboven (2014) and Rossi (2014) focus on instrument variables and their role within RCML type models. Others such as Bajari et al. (2007), Fosgerau and Bierlaire (2007), Train (2008), Bastin et al. (2010), and Fosgerau and Mabit (2013) introduce more flexible distributions to the models to prevent the misspecification that can occur when inappropriate mixing distributions are used. We place our paper in a small but growing strand of literature that uses novel, often experimental, methods to either conduct demand estimation or more generally

Competition Markets Authority (CMA) has 40 working days to complete Phase 1 and a further 24 weeks during Phase 2 to conduct their investigation and submit a final report. In the US, where the Federal Trade Commission (FTC) and Department of Justice (DoJ) are jointly responsible for merger analysis, pre-merger reviews must be completed within 30 days and if necessary the agencies are granted another 30 days to investigate further and take action if required.

assess unilateral price effects arising from some change in the market.

Conlon and Mortimer (2013) conducted some of the earliest experimental work in merger analysis in response to changes in the DoJ/FTC Horizontal Merger Guidelines that set new standards based on upward pricing pressures (UPP) which in turn rely on diversion ratios. They estimate diversion by exogenously removing products from vending machines and analysing changes in demand, firm profits, diversion ratios and UPP. However, this type of field experiment is both costly and time consuming; it does not solve the problems of RCMLs in contrast to our experiment which offers solution to these issues. Conlon and Mortimer (2021) follow up their previous work by establishing a local average treatment effect (LATE) interpretation of diversion ratios and show how diversion ratios (although not demand parameters - hence our experiment is more flexible in its use) can be estimated using different interventions. Although they mention the potential to use a lab (or online) experiment, the paper does not implement any experiments.

Imthorn et al. (2016) advocate the use of conjoint-analysis to overcome biases such as framing effect and those caused by interviewees strategic interests that occur during typical survey methods used by competition authorities.[2] However, Imthorn et al. (2016) themselves state the usage of such methods is limited and we have found no similar implementation by any other competition agency before or since. The authors speculate this may be 'due to a perception that these techniques are complex and time-consuming'. We show in this paper that neither of those limitations hold true. In 2010, the ACM used a choice-based conjoint-analysis (CBC), similar to our inter-brand treatment as part of wider empirical research including interviews and questionnaires to approve the merger between Agrifirm and Cehave, two producers of agricultural products. The resultant merger simulations were used as evidence that the merged entity would not be able to profitably raise prices significantly. Other

[2] E.g. 'what percentage price increase in product A would it take for you to switch to product B?'

attempts to use CBCs by the ACM were less successful in part because of market structure and sometimes unrealistic substitution patterns post estimation. We deal with this issue specifically by comparing our inter- and intra-brand treatments.

Moshary et al. (2022) conduct a similar experiment to ours in that they present subjects with choice sets in an experimental setting in order to elicit demand preferences in the market for firearms. Having obtained substitution patterns for various types of guns, they simulate changes in gun regulations and use the estimated demand model to assess changes in demand and consumer surplus. As mentioned before, Moshary et al. (2022) illustrate an important use case for experiments where empirical data is simply not available. While the foundation of the experiment is similar, crucially their experiment always shows the gun brand as a product characteristic. They do not conduct an intra-brand equivalent in their study and thus face the same challenges we did when using the brands without an alternative procedure as in this paper. While not experimental, as such, Qiu et al. (2021) use win/loss data to identify diversion ratios for merger analysis, recognising the need for simple and efficient methodologies to use in real-time. Incidentally, one could generate this data using survey methods; while this elicitation has been criticised by some U.S. courts, we believe that certain adaptations can be made to improve their external validity.[3] See section 2.5 for more detail.

Magnolfi et al. (2022) take a different approach to experimental demand estimation by using a triplet experiment where subjects are presented with a reference product and are asked to select the two products that are most similar to the reference from a given choice set. They then use a machine learning algorithm to estimate an embedding – a low-dimensional representation of the latent product space. Substitution patterns can be inferred from the distances between product pairs in the embedding. Two other papers also use embeddings in demand estimations. Bajari et al. (2023) use deep neural nets to

[3] See *U.S. v. H&R Block Inc., et al., D.D.C.* (2011).

generate an embedding from products image and text descriptions, useful in cases where the demand relevant information may not easily be defined by a set of measurable characteristics, even though humans are able to process and synthesise the relevant information. However, a key difference between our work is that the embedding serves to augment price and quantity data in a traditional demand estimation model. While we require some data on price and quantity, the requirements are less strenuous (we use readily available national level data) and serve to augment our experimental data. We consider our methodology to be complimentary to other tools used in merger evaluation, both qualitative and quantitative. Armona et al. (2021) use search data to estimate consumer preferences for hotels by using a Bayesian Personalised Ranking to learn products' latent characteristics from consumers web-browsing history. We see these latent attribute methods as complementary to our work using observable product characteristics. As Armona et al. (2021) themselves state 'if the observables are rich, the value add from latent characteristics may be smaller'. Ultimately, the choice of which techniques to apply depends in part on the product(s) of interest.

The results of our experiments are promising. Following the estimation of the demand parameters, we use these to estimate substitution effects and markups so it is these that we ultimately compare to previous studies. We calculate elasticities for a set of real products that consists of the 18 most popular beers in the US by market share in 2019 using parameters estimated from both treatments. This attempts to place our demand parameters in context by comparing them to results observed by Miller and Weinberg (2017) in work analysing the effects of the Miller-Coors joint venture in 2008. It should be noted that the data set they use is not contemporaneous to ours; the product set is different and the structure of the industry has changed so direct comparisons between our results and those of Miller-Weinberg are not possible. We simply use their results to show that our method can produce what appear to be realistic values for individual product elasticities as well as median own-price elasticities. The median own-price elasticity for our real product set, calculated using intra-brand parameters of -6.55, is greater in magnitude but still falls close to the range reported by Miller-Weinberg (–4.73 to –4.33) for their various random coefficient nested logit specifications. Considering the market changes over the 11 years between the datasets, this suggests that the methodology can produce realistic substitution patterns. Since 2008, the largely light-lager culture of the US beer market has been disrupted by the emergence of craft beer and local micro-breweries. The combined market share of the top 5 brands fell by 1% per year in the eight years to 2016, with growth in the beer industry focused in taproom volume, up 15%, and microbrewery volume, up 27% (Watson, 2018). Our predicted markups in the range of 20.2-22.4% are lower than Miller and Weinberg's estimated 34%. In section 2.5, we discuss some of the challenges we faced and lessons we learned during the design and implementation of this methodology, including ways in which to improve the accuracy of estimates.

The rest of the paper is organised as follows. In section 2.2, we detail the experimental design guided by Hensher (1994), including testing of experimental features through Monte Carlo simulations, further detailed in Appendix A. Section 2.3 describes our model that encompasses elements from various strands of the existing literature. We define indirect utility, choice probabilities, price elasticities and price-cost margins. Section 2.4 provides an example of the types of results the estimation procedure can produce and attempts to place them in the context of existing work. In section 2.5 we discuss some of the issues we faced and provide thoughts on how the version of the experiment we conducted can be adapted for real-world use. Finally, we conclude in section 2.6.

2.2 Experiment

We chose beer as our primary product because the industry is an oligopolistic differentiated product market that has been studied in the past. It is also an industry that has seen a significant amount of merger activity over the years. A key requirement of the mixed logit model is to obtain data in long form.[4] We find that it is easier to create the experiment with this consideration in mind rather than attempt to switch later. The design process we use is adapted from Hensher (1994). Firstly, we define our set of product characteristics. These must be relevant to the purchase decision as well as observable and measurable. Price is included because marginal utility of income is a key component of the price elasticity of demand function. Based on previous studies including Miller and Weinberg (2017) and Lerro et al. (2020), we chose ABV (alcohol by volume) to represent alcohol content, volume per unit to represent packaging size and can/bottle to represent packaging material as our remaining product characteristics. These are identified in the 'attributes' column of Table 2.1.

Table 2.1: Attributes and levels of survey products

Attributes	Levels of features
Beers	
Price/6-pack	\$6.49, \$7.99, \$10.99
ABV	3.6%, 4.6%, 5.5%
Can/Bottle	0 = can, 1 = bottle
Volume/unit	8.4-oz, 12-oz, 16-oz

2.2.1 Intra- versus inter-brand treatments

The issue of branding is a key consideration for our experiment. Firms spend heavily on marketing and advertising to increase visibility and recognition of their products and differentiate their brands from competitors in order to reduce the brands' own price elasticity of demand. When choices are made in the real-world, consumers consider brand names in their purchase decision because they confer information to the consumer as a result of advertising, particularly in our case, where there is only a limited amount of information

[4]Each row represents one alternative in a choice set, with either a zero or one to indicate whether that alternative was chosen.

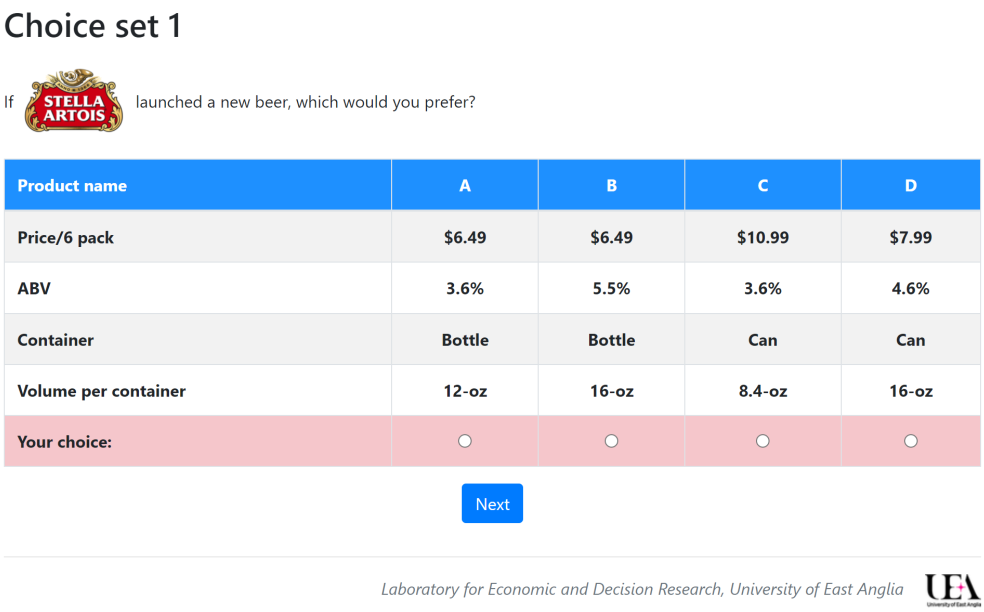
conveyed by our product characteristics. Therefore the inclusion of brands would serve to improve external validity. Branded choices are also more tangible in the minds of subjects and may increase internal validity of the experiment. De Bekker-Grob et al. (2010) find that including brand labels in the choice of colorectal screening programs changes individual choices and reduces the attention that respondents paid to the specified attributes. They suggest unlabelled alternatives are more suitable when investigating attribute tastes and associated trade-offs and labelled alternatives may be more appropriate when the goal is to predict real-life choices. However, as the brand name itself conveys information to the subjects beyond the attributes specified in the experiment, these characteristics are unobserved by the researcher. The key issue is that we, as researchers, have no way of controlling for these unobserved characteristics. Therefore, to avoid problems of endogeneity or omitted variable bias that may arise if unobserved characteristics are correlated with price or the random error term - which in turn can have significant consequences for the magnitude of parameter estimates especially on price where positive associations can lead to underestimating coefficients, while negative associations can lead to overestimating coefficients - it may be prudent to use unbranded alternatives (Louviere et al., 2000). Here, products have no specific names, and are identified only as option 'A', 'B', 'C' etc. for example.

To balance these issues we settle on two treatments which we call intra-brand and inter-brand. Brands are present in both treatments but are utilised differently.

Treatment 1: Intra-brands

In this treatment products are hypothetical and each individual product is constructed as a combination of attributes from the values in Table 2.1. The levels in column 2 were chosen to balance realism with econometric considerations. The range of values should be believable and large enough to ensure sufficient variability to identify model parameters but not so large that

Figure 2.1: Intra-brand treatment example screen



Choice set 1

If [STELLA ARTOIS] launched a new beer, which would you prefer?

Product name	A	B	C	D
Price/6 pack	$6.49	$6.49	$10.99	$7.99
ABV	3.6%	5.5%	3.6%	4.6%
Container	Bottle	Bottle	Can	Can
Volume per container	12-oz	16-oz	8.4-oz	16-oz
Your choice:	○	○	○	○

Next

Laboratory for Economic and Decision Research, University of East Anglia

there are a high number of dominated alternatives. The more levels for each attribute, the more choice tasks are required. In order to increase the salience of the choice in the minds of a subject, we included a randomly chosen brand logo to appear at the top of each choice set. The subject was then asked *"If* [brand name] *launched a new beer, which would you prefer?"* As a result any brand effects are fixed across all four options. An example screen is shown in Figure 2.1.

Treatment 2: Inter-brands

In this treatment, we use real branded products with real product characteristics except for price which is randomly allocated to a product from the prices in Table 2.1. An example screen is shown in Figure 2.2. Now the first row contains the brand name as well as a picture of the product to simulate the choice a consumer might face on a supermarket shelf. Otherwise, the presentation remains the same as in treatment 1.

As it is more practical (and indeed cheaper) to ask fewer respondents to make

Figure 2.2: Inter-brand treatment example screen

Choice set 2

Product name	A	B	C	D
Brand	 Heineken	 Keystone Light	 Michelob Ultra	 Corona Extra
ABV	5.0%	4.1%	4.2%	4.5%
Container	Bottle	Bottle	Bottle	Bottle
Volume/unit	12-oz	15-oz	16-oz	12-oz
Price/6-pack	$10.99	$6.49	$7.99	$7.99
Your choice:	○	○	○	○

Next

Laboratory for Economic and Decision Research, University of East Anglia

repeated choices rather than ask more respondents to each make a single choice, we use a panel data set. There is some disagreement in the literature regarding an appropriate number of choice sets in the context of subject fatigue. Bradley and Daly (1997) argue that fatigue caused by a large number of choice sets increases the error term variance. Hess et al. (2012) provide evidence that these concerns are overstated. Ultimately, we follow Chung et al.'s (2011) recommendation that different specifications and functional forms should be pretested in order to identify optimal numbers of products and choice sets. This pretesting is done through a simulation exercise using 'fake' data. The methodology and output of this is described in appendix A. As a result of the simulation, we settle on 4 alternatives in a choice set and 8 choice sets per subject in the intra-brand treatment. Having found that subjects completed the task sooner than we expected, we increased the number of choice sets to 10 in the inter-brand treatment.

2.2.2 Identification

As Holmes et al. (2017) state 'an experimental design must contain sufficient independent variation among attribute levels within and across alternatives so that each preference parameter can be identified. For example, if the levels of an attribute are always identical across alternatives, it will not be possible to identify the effect of that attribute on responses.' To illustrate that our experiment adheres to this principal, Table 4.2 presents summary statistics for each product characteristic across each treatment.

In the intra-brand treatment, 486 subjects each saw 8 choice sets composed of 4 products resulting in 15,552 products that were shown across the entire treatment. For the inter-brand treatment we included 10 choice sets of 4 products to 493 participants resulting in 19,720 products that were shown across the entire treatment. The units are the units of measurement of each particular attribute.[5] The number of observations, M in each treatment is equal to the total number of products shown. We refer to the number of participants in each treatment as m. Finally, $\overline{T}$ is the average number of choice sets an individual sees in each treatment. Since our panel is balanced $\overline{T}$ is a whole number for both treatments. The *overall* row for each product characteristic is the mean, standard deviation and min/max across these M products. The dimensions of the panel are the participant i and the set of choices t. Variation across individuals, i.e., a cross section is called *between (B)* while variation over choice sets for a given individual is called *within (W)*. The *between* output first estimates participant-level averages, $\bar{x}_i$ for each of the 486 participants, then calculates s for these means such that $s_B^2 = \frac{1}{m-1}\sum_i (\overline{x}_i - \overline{x})^2$. The *within* output shows how much a product characteristic varies within the products a person sees, while ignoring all variation between participants. In other words, we calculate the standard deviation of a product characteristic for each participant separately and then average these values to get the reported within

[5] Price is measured in dollars; ABV in percentage; container is a 0/1 dummy and volume is measured in fluid ounces.

s so that $s_W^2 = \frac{1}{M-1}\sum_i \sum_t (x_{it} - \overline{x}_i)^2$. As we expect, for every product characteristic, the within variation is greater than the between. Intuitively, this means that the options different participants see are similar but the variation between these options is greater. Later, in section 2.4, we present further summary statistics of the number of times each product was shown in a treatment and the number of times it was chosen, which further illustrates how subjects responded to changes in price for a given product.

Table 2.2: Summary statistics: Variation in product characteristics

Variable		Mean	Std.Dev	Min	Max	Obs
		Intra-brand Treatment				
Price	Overall	8.502	1.867	6.49	10.99	15552
	Between		0.320	7.568	9.724	486
	Within		1.840	-3.234	3.422	8
ABV	Overall	4.559	0.780	3.6	5.5	15552
	Between		0.134	4.175	5.006	486
	Within		0.768	-1.406	1.325	8
Container	Overall	0.5	0.500	0	1	15552
	Between		0.085	0.250	0.843	486
	Within		0.493	-0.844	0.750	8
Volume	Overall	12.163	3.113	8.4	16	15552
	Between		0.526	10.463	13.913	486
	Within		3.068	-5.513	5.538	8
		Inter-brand Treatment				
Price	Overall	8.500	1.869	6.49	10.99	19720
	Between		0.306	7.653	9.603	493
	Within		1.843	-3.113	3.338	10
ABV	Overall	4.523	0.342	4.1	5.2	19720
	Between		0.050	4.353	4.665	493
	Within		0.339	-0.565	-0.818	10
Container	Overall	0.550	0.498	0	1	19720
	Between		0.070	0.375	0.775	493
	Within		0.493	-0.775	0.625	10
Volume	Overall	12.551	1.399	11.2	16	19720
	Between		0.207	12.02	13.38	493
	Within		1.384	-2.180	3.945	10

Panel dimensions are participant, i by choice set, t. Variation over time for a given individual is called *within (W)*, and variation across individuals (cross-section) is called *between (B)*. The overall variation is $s_O^2 = \frac{1}{M-1}\sum_i \sum_t (x_{it} - \overline{x})^2$ while $s_B^2 = \frac{1}{m-1}\sum_i (\overline{x}_i - \overline{x})^2$ and $s_W^2 = \frac{1}{M-1}\sum_i \sum_t (x_{it} - \overline{x}_i)^2$. Observations are listed as overall (M), over number of participants (m) for between, and average number number of choice sets per partipant $(\overline{T})$ for within.

2.2.3 Experimental Design

Historically, capacity constraints in the lab meant that the number of observations one could obtain was limited. Therefore, alternatives in each choice set had to be selected in such a way that they extracted the maximum amount of information so that the model could be correctly identified. This is particularly true of the intra-brand treatment as it consists entirely of hypothetical products. For laboratory studies, orthogonal arrays in which the attribute levels are independent both within and between alternatives became the preferred experimental design when choosing alternatives for a choice set.

The benefit of online experiments is that they are easily scaleable. Random sampling theory guarantees that if we take large enough samples from the complete factorial, we should closely approximate the statistical properties of the factorial itself (Louviere et al., 2000). Since we require a large number of observations to achieve consistent and efficient parameter estimates anyway and our simulation exercise indicates that beyond a few thousand observations the marginal gains in accuracy decrease significantly, we are able to draw on random sampling from the full factorial set as the selection method for alternatives in a choice set, without the need for deriving several complex orthogonal arrays. In fact, for certain cases, Rose and Bliemer (2009) show that an orthogonal design is not the most efficient design and so-called 'efficient' designs are able to produce more efficient data in the sense that more reliable parameter estimates can be achieved with an equal or lower sample size. Random assignment of alternatives to choice sets across a large number of choice sets also achieves attribute level balance which ensures the parameters can be estimated well on the whole range of levels, instead of just having data points at only one or few of the attribute levels. Identification is then achieved because we have variation in our product characteristics by construction from Table 2.1 across and within subjects, alongside attribute level balance and variation in choice sets between and within subjects. In the top panel of Table 4.2 we can see that there is less variation in the ABV and volume when using real products in the inter-brand treatment. When we randomise these product characteristics in the intra-brand treatment we get greater variation. Price is randomised in both treatments which is why the mean within choice set variation is similar and container only has two options so its variation is similar as well. In part this explains why we find smaller (in magnitude) parameter estimates in the inter-brand treatment; there is simply far less variation in these specific product characteristics in the set of real products that parsing out preferences is difficult. This lack of variation in product characteristics is not uncommon in empirical data. Panel B shows the number of times we held the non-price product characteristics constant and varied price for individuals and the mean number of times this occurred for each individual.

2.2.4 Realism and External Validity

Of primary concern for any SP type experiment are issues of realism and external validity. By construct, the surveys elicit hypothetical responses and so minimising hypothetical bias, or 'the potential error induced by not confronting the individual with an actual situation' (Schulze et al., 1981) is paramount. It is possible to achieve high levels of realism through complex choice tasks yet this must be balanced with the levels of stress and cognitive burden placed on participants which can reduce the quality of responses (Hensher and Cherchi, 2015).

Incentivisation

One of the biggest challenges for any stated choice experiment is to convince external validity and realism exist when consumers are not making consequential choices (Bergman et al., 2020). If consumers are not spending their own money, they may simplify their decision process for example, always choosing option A. As mentioned earlier, lengthy surveys can result in boredom and cognitive fatigue which increases survey noise and correspondingly reduces the quality of responses. We include attention checks at random points within each round to ensure the participant is not just randomly clicking through choices. However, as of the current experiment we have not devised a satisfactory methodology of incentivising choices which would increase external validity. Experimenting with various incentivisation strategies is an area for further research, but beyond the scope of this paper.

The challenge in our experiment is to provide incentives in such a way that it recreates the experience of consumers in an actual supermarket. One possibility is to give subjects an endowment at the beginning of the experiment so that one of their choices could be randomly chosen to 'purchase' the actual goods. However, we know from the mental accounting literature (see Arkes et al. (1994)) that subjects treat this not as part of their regular endowment but as a windfall and what we observe is how they treat this windfall rather than how they behave with their own money. On the goods side, depending on the products in question, an actual provision or delivery may be prohibitively expensive or simply infeasible. Finally, the close recreation of incentives involves an outside option of no purchase in the experiment. But then the actual outside option – outside of the experiment – becomes relevant and is difficult to control or observe. Ultimately, we posit that when faced with a choice in our experiment consumers default to their past shopping experiences in the absence of any other information and thus mimic those choices closely.

2.2.5 Data

We administered the treatments described above in June/July 2021 on the online subject recruitment platform Prolific. The subject pool was restricted to US residents aged between 21-30, which gave us the largest geographical market to operate in. Previous work in the US beer market also enabled us to make some comparisons to existing data. The age restriction included the minimum drinking age in the US and an age range most likely to be found on a student campus.[6]

In total, 1,000 subjects, divided equally between treatments, made a choice for each of the eight/ten choice sets presented to them in each treatment, resulting in 4000 / 5000 observations per treatment.[7] Participants were paid a fixed fee

[6] We did intend to collect some data in-person to compare to our online experiment, but Covid restrictions in 2021 prevented us from using the experimental lab on the UEA campus.

[7] A small number of participants in each treatment were excluded from the final analysis

for their time. As a result of an unexpected surge in sign ups to Prolific of young women aged 18-30 around the time of our experiment, many studies including ours suffered from a severe gender bias; 79% of subjects were female.[8] We felt the data remained suitable for our methodological purposes, but we recognize that any predictive claims could be weakened by the unrepresentative sample. In addition to their product choices, data on the demographics of the subjects including age, gender, income, and location by state was also collected.

2.3 Model Specification

The mixed logit model is in the class of random utility models (RUM) derived from assumptions of utility maximisation. Individual n faces a choice between products $j \in J$ over a set of $t \in T$ choice situations in the experiment. The utility individual n derives from product j in choice situation t is

$$U_{njt} = \beta_n' x_{njt} + \varepsilon_{njt}. \tag{2.3.1}$$

An individual will choose product j if and only if $U_{njt} > U_{nlt}\ \forall j \neq l$. β_n is a vector of coefficients on the product characteristics shown in the experiment that is unobserved for the sample and varies in the population with density $f(\beta|\theta^*)$ where θ^* are the true location and scale parameter of the population distribution. x_{njt} is a vector of observed product characteristics for each beer in each choice set.

Each individual has their own value of β_n that can be estimated and represents their tastes and preferences over the defined product characteristics. The values of these β_n's are distributed over the population with parameters θ^*. It is these population parameters, θ^*, that we seek to estimate through the mixed logit

because they failed one of the random attention checks during the experiment or for missing or inappropriately answering the demographic questions. For example several participants stated their age was outside of the specified range.

[8] The flood of new participants was subsequently attributed to a viral TikTok in which a teenager promoted Prolific as a 'side-hustle'; an easy way to make a few extra dollars. The video garnered 4.1 million views in a month (Letzter, 2021).

model.

Since each individual's β_n is unobserved, the exact unconditional probability of n's sequence of choices made during the experiment is the integral of the conditional probability over all possible values of β as defined by the true parameters of the distribution of β_n, θ^*,

$$P_n(y_n|\theta^*) = \int P_n(y_n|\beta_n) f(\beta|\theta^*) d\beta. \tag{2.3.2}$$

However, since the integral in (2.3.2) does not have a closed form solution, $P_n(y_n|\theta^*)$ must be approximated via simulation by taking R draws of β_n for a given θ, calculating the statistic $P_n(y_n|\beta_n)$ for each draw and averaging.[9]

As discussed earlier, we use the mixed logit because of its flexibility. Utility is composed of a mean component that is common to all members of the population and a stochastic portion for each individual. This stochastic portion is correlated over alternatives and choice situations because it is a common term so that the model can allow for general models of substitution and is not constrained by independence of irrelevant alternatives (IIA). Any RUM model can be approximated by a mixed logit through appropriate selection of product characteristics and distribution for the coefficients (McFadden and Train, 2000); we specify a normal distribution for all non-price characteristics and a log-normal distribution for price such that the coefficient is always negative.

Having estimated coefficients on our model parameters, we obtain a set of model predicted market shares, $\mathbf{s} = (s_0, s_1, ..., s_J)$ that are distinct from actual observed shares in the real world. In order to reconcile these with observed shares from our real world data, $\mathbf{S} = (S_0, S_1, ..., S_J)$ we employ the inversion of Berry (1994), under the principle that we wish to obtain $\mathbf{s}$ as close to possible as $\mathbf{S}$ by finding

[9] We do not cover these derivations as they are covered in detail elsewhere. For an excellent presentation see Train (2009), chapters 3 and 8.

θ such that

$$\min_{\theta} \sum_{j=1}^{J} [\mathbf{S}_j - \mathbf{s}_j(\alpha, \beta, \xi_1, \xi_2, ..., \xi_J)]^2. \quad (2.3.3)$$

Note that α is the coefficient on price and is simply one element of the vector of coefficients β. The key term here is ξ which is the unobserved product characteristics that enter the predicted market shares. ξ enters the mean utility, $\delta_j = \alpha(-p_j) + x_j\beta + \xi_j$, giving us a J-system of equations in the form of $\mathbf{S} = \mathbf{s}(\delta_1, \delta_2, ...\delta_J)$. By inverting this system we can estimate $\hat{\delta}$ to give us estimates for ξ, which we use in combination with our previously estimated values of β (and α) going forward. The presence of random coefficients means these equations are non-linear and must be solved numerically using a contraction mapping

$$\delta^{h+1} = \delta^h + [\ln(\mathbf{S}) - \ln(\mathbf{s}(\delta^h))] \quad (2.3.4)$$

where we take an initial value of δ from the initial coefficient estimates combined with the real world product characteristics and iteratively solve the above until $||\delta^{h+1} - \delta^h||$ is less than some specified tolerance level.

Estimating $\hat{\theta}$ provides a foundation for further analysis. In merger simulations, the demand estimates can be used to calculate price elasticity of demand, which when combined with data on marginal costs and ownership structures can be used to predict the price and welfare effects of a merger. Let $\eta_{jk} = \frac{\partial q_j}{\partial p_k}\frac{p_k}{q_j}$ be the price elasticity of demand where p_j and q_j are the price and quantity of good j in the market. Instead of quantities, in the logit case, we use predicted market shares $s_j = \frac{q_j}{M}$ where M is the total size of the market. Market shares in turn are equivalent to the predicted probabilities such that

$$\eta_{jk} = \begin{cases} -\frac{p_j}{s_j} \int \alpha_n P_{nj}(1 - P_{nj}) f(\beta) d\beta & \text{if } j = k, \\ \frac{p_j}{s_k} \int \alpha_n P_{nj} P_{nk} f(\beta) d\beta & \text{otherwise.} \end{cases} \quad (2.3.5)$$

This results in a $J \times J$ (or a $J \times (J+1)$ with an outside good) matrix in which the main diagonals are the own price elasticities of goods $j \in J$ and the off-diagonals are the cross-price elasticities of goods $j, k \in J$. Therefore, by combining our demand estimates with real world observations on price and product characteristics we should be able to obtain measures of price elasticity for real products.

In a monopolistic market obtaining price elasticity is sufficient to infer marginal cost, c because at the profit-maximising price, the price-cost margin is equal to the negative reciprocal of the price elasticity of demand.

Following our mixed logit specification, in an oligopoly of F firms in which the fth firm produces a subset $\mathcal{F}_f \in J$ products, a firm's joint profit is given by

$$\Pi_f = \sum_{k \in \mathcal{F}_f} (p_k - c_k) s_k(\mathbf{p}; \theta), \tag{2.3.6}$$

where c_k is the constant marginal cost of the kth product and $\mathbf{p}$ is a vector of all relevant prices. Assuming Nash-Bertrand competition, the profit maximisation first order condition can be written as

$$\mathbf{p} = \mathbf{c} + \mathbf{\Omega^{-1}} \mathbf{s}, \qquad \text{where} \qquad \Omega_{jk} = -\phi_{jk} \frac{\partial s_k(\mathbf{p} : \theta)}{\partial p_j}, \tag{2.3.7}$$

$\mathbf{s}$ is a vector of market shares and Φ is a 1/0 $J \times J$ matrix where element ϕ_{jk} is 1 if j, k are produced by the same firm and 0 otherwise. We can therefore simulate a merger by changing the elements of Φ and iteratively re-estimating 2.3.7 to obtain post-merger prices.

2.4 Results

We begin by summarising the experimental choice sets and choices made for each treatment, in Table 2.3 and Table 2.4. Each table is structured identically in the following fashion. The first column is a product identifier; numerical in the intra-brand treatment and the names of real products in the inter-brand treatment. The next three columns indicate non-price product characteristics. In the intra-brand treatment these are created from the options in Table 2.1. In the inter-brand treatment these are real product characteristics. Columns labelled Appearances indicate first the total number of times a product appeared in the experiment across all choice sets for all subjects, then number of times a product appeared at each price point. We can see that aside from some random variation the total number of appearances for each product within each treatment is broadly the same. Likewise, each product is equally likely to be seen at each price point within each treatment. The final four columns in each table indicate the number of times each product was selected from its choice set, in total and then at each price point. In Table 2.3 we can see that the number of times a product is chosen in total increases as we move down table from product 1 to product 18. This represents a general preference in the population for a larger ABV and more volume. Preferences between can and bottle are less visible in this table. Further, for every product the number of times it was chosen decreases as the price increases, illustrating well-behaved demand curves. The same cannot be said for the choices made in Table 2.4. There is no pattern to which beers are preferred in total. In fact, since the beers are ordered by real world market share we would expect the figures to broadly descend from Bud Light to Coors Banquet but this is not the case. Finally, when we observe the choices at each price point, the demand curves are not well-behaved. An illustration of both sets of demand curves is presented in the appendix.

We first estimate a mixed logit model on the data from each treatment using PyBLP (Conlon and Gortmaker, 2020). The results are presented in Table 2.5.

Table 2.3: Summary Statistics: Intra-brand choice table

				Appearances*				Chosen†			
Product	ABV	Type	Size	Total	\$6.49	\$7.99	\$10.99	Total	\$6.49	\$7.99	\$10.99
1	3.6	Can	8.4	880	308	274	298	81	47	23	11
2	3.6	Can	12	876	280	298	298	88	56	20	12
3	3.6	Can	16	872	291	277	304	139	82	34	23
4	3.6	Bottle	8.4	904	288	314	302	79	44	24	11
5	3.6	Bottle	12	863	276	294	293	137	70	53	14
6	3.6	Bottle	16	900	276	298	326	206	108	70	28
7	4.6	Can	8.4	885	280	323	282	125	68	48	9
8	4.6	Can	12	840	275	309	256	195	108	62	25
9	4.6	Can	16	865	298	285	282	238	126	86	26
10	4.6	Bottle	8.4	828	273	285	270	146	87	50	9
11	4.6	Bottle	12	812	276	281	255	246	132	85	29
12	4.6	Bottle	16	852	301	266	285	327	182	107	38
13	5.5	Can	8.4	811	247	288	276	166	84	66	16
14	5.5	Can	12	878	270	292	316	286	157	82	47
15	5.5	Can	16	869	301	286	282	377	193	135	49
16	5.5	Bottle	8.4	843	298	282	263	245	130	84	31
17	5.5	Bottle	12	861	240	295	326	326	140	126	60
18	5.5	Bottle	16	913	320	302	291	481	222	185	74

*Shows number of times each product appeared out of 15,552 products shown in the treatment, in total and at each price point. †Shows number of times each product was chosen in total and at each price point. Size is in fluid ounces (oz).

Table 2.4: Summary Statistics: Inter-brand choice table

				Appearances*				Chosen†			
Product	ABV	Container	Volume	Total	\$6.49	\$7.99	\$10.99	Total	\$6.49	\$7.99	\$10.99
Bud Light	4.2	Can	16-oz	1,083	386	319	378	252	107	75	70
Coors Light	4.2	Can	12-oz	1,129	360	359	410	195	71	68	56
Miller Lite	4.2	Bottle	12-oz	1,105	345	364	396	237	78	83	76
Budweiser	5	Can	12-oz	1,110	355	368	387	233	90	82	61
Michelob Ultra	4.2	Bottle	12-oz	1,080	369	384	327	303	106	122	75
Corona Extra	4.5	Bottle	12-oz	1,091	389	366	336	470	191	167	112
Modelo Especial	4.5	Bottle	12-oz	1,104	377	365	362	411	154	159	98
Natural Light	4.2	C	12-oz	1,163	381	407	375	70	33	22	15
Busch Light	4.5	Can	12-oz	1,113	373	391	349	120	45	49	26
Busch	4.3	Can	12-oz	1,073	362	348	363	77	37	26	14
Heineken	5	Bottle	12-oz	1,092	364	366	362	387	163	129	95
Keystone Light	4.1	Bottle	15-oz	1,033	334	347	352	62	26	24	12
Miller High Life	4.6	Bottle	12-oz	1,075	324	370	381	262	99	82	81
Stella Artois	4.8	Bottle	11.2-oz	1,103	346	387	370	429	160	150	119
PBR	4.7	Can	12-oz	1,099	381	372	346	231	94	80	57
Blue Moon	5.2	Bottle	12-oz	1,086	324	388	374	652	213	246	193
Dos Equis	4.2	Bottle	12-oz	1,070	356	353	361	262	105	90	67
Coors Banquet	5	Can	12-oz	1,111	369	377	365	277	114	103	60

*Shows number of times each product appeared out of 19,720 products shown in the treatment, in total and at each price point.

†Shows number of times each product was chosen in total and at each price point.

The model parameters, $\hat{\theta}$, refer to the mean and standard deviation of each of the elements of the vector β. Each product characteristic is specified to have a random component such that there is heterogeneity in preferences and we do not include any demographic variables. The random parameter on price, which is commonly referred to as α is an element of β and is specified as log-normal for two reasons. Firstly, prior studies have shown that this is typically the shape for the distribution of preferences on price. Secondly, it ensures all parameter estimates have the same sign so that the parameter estimate on price α is negative for all n. All other random parameters are specified to be normally distributed. This is of course, an a priori assumption but it is straightforward to estimate the parameters of any parametric distribution including a uniform or triangular distribution where appropriate. Estimating non-parametric distributions is possible; as McFadden and Train (2000) state, it is possible to estimate any RUM model to any degree of accuracy by a mixed logit with appropriate observed product characteristics and mixing distribution. However, as the number of parameters to estimate per characteristic increases, the estimation becomes computationally complex. Although the likes of Fosgerau and Mabit (2013) and Train (2016) have detailed methods to navigate these estimations, we have no reason to believe preferences on our chosen characteristics are distributed in such fashion.

2.4.1 Treatment 1: Intra-brand

Column 1 of Table 2.5 shows the results of the intra-brand treatment. We can see that consumers prefer a higher ABV, and volume per unit but a negative coefficient on container indicates that subjects prefer cans to bottles. The standard errors on these non-price product characteristics are all small and the estimates are statistically significant. Similarly, the standard deviations are all statistically significant which suggests the presence of unobserved heterogeneity in preferences and that a random specification is appropriate. For the parameter on price, the log-normal coefficients m and s are estimated such that the reported mean is equal to $\exp(m + (s^2/2))$ and the reported standard deviation is equal to $m * \sqrt{\exp(s^2) - 1}$. The sign is negative, indicating utility goes down as price goes up, but of course this is a result of the log-normal specification we used.

Interaction Effects

Where available individual-specific demographic data can be included in the model as a source of observed heterogeneity through an interaction with relevant product characteristics. Some of the unobserved heterogeneity in the population model can then be 'explained' by the observed demographic characteristics of sampled individuals. Although it may be tempting to add pairwise interactions between each demographic variable and each product characteristic, the larger the number of interactions, the greater the number of moment restrictions required. Hence a researcher must decide which demographic and product characteristics interact in reality.

We collected data on income, age, ethnicity and home state for each individual. As Hensher and Greene (2003) state, these demographic effects can be included in the model by interacting the variable with the random parameter and adding it in as a fixed parameter. In this specification, $U_{nj} = \beta_n' x_{nj} + \kappa(z_n x_{nj}) + \varepsilon_{nj}$ where z_n is a vector of demographic characteristics, and κ is a fixed parameter (we drop t for notational simplicity). A common and plausible interaction is between price and income. Of the 486 subjects, six declined to provide information on their income so they were dropped from the sample for this specification. The results are presented in column 2 of Table 2.5.

The results show that there is a small interaction effect and the positive sign suggests that as income rises subjects are slightly less sensitive to price. However, this effect is not statistically significant which means that there is absence of heterogeneity around the mean on the basis of observed income. This is not to say that income has no effect on the distribution of preferences on price, simply that we have failed to discover its presence. It must be noted at this point that there is an issue with our data with regards to income. Participants were asked, 'What is your monthly income in dollars?'. Some subjects clearly stated their annual income but more importantly around 13% of subjects responded with 0. This is likely to be students not in any form of employment. Of course, these subjects still have a monthly budget and it is this including all loans, stipends and allowances that was required. As a result, we question the non significance of the income interaction. To further illustrate the point we include a second specification, in column 3, that includes an interaction between age and ABV, and gender and ABV. The results suggest that younger people and women prefer a stronger beer, although neither estimate is statistically significant. Again, we do not place too much emphasis on the result itself because the gender bias in the sample means that female preferences drive the estimates. Nevertheless, it serves to illustrate the mechanism of the interaction.

Table 2.5: Mixed logit estimates

		Intra-brand			Inter-brand	
		(1)	(2)	(3)	(4)	(5)
Price	Mean (μ_α)	-1.027*	-1.025*	-1.017*	-0.275*	-0.220*
		(0.077)	(0.068)	(0.068)	(0.047)	(0.030)
	SD (σ_α)	1.148*	1.112*	1.133*	0.836*	0.652*
		(0.185)	(0.154)	(0.161)	(0.354)	(0.224)
ABV	Mean (μ_{β_1})	1.444*	1.439*	2.864*	1.605*	0.202*
		(0.089)	(0.090)	(0.752)	(0.104)	(0.027)
	SD (σ_{β_1})	1.430*	1.455*	1.442*	1.741*	
		(0.089)	(0.088)	(0.089)	(0.109)	
Container	Mean (μ_{β_2})	-0.686*	-0.701*	-0.276*	1.215*	0.589*
		(0.099)	(0.101)	(0.869)	(0.081)	(0.037)
	SD (σ_{β_2})	1.675*	1.710*	1.699*	1.411*	
		(0.111)	(0.114)	(0.113)	(0.081)	
Unit volume	Mean (μ_{β_3})	0.256*	0.261*	0.260*	-0.001	-0.147*
		(0.016)	(0.017)	(0.017)	(0.026)	(0.009)
	SD (σ_{β_3})	0.257*	0.249*	0.250*	0.369*	
		(0.019)	(0.018)	(0.018)	(0.032)	
Price × income			0.001			
			(0.001)			
ABV × age				-0.059		
				(0.031)		
ABV × gender				-0.017		
				(0.035)		
Observations		3888	3840	3728	3944	3944

Column (1) shows results from base intra-brand specification; columns (2) and (3) add demographic interaction terms for price and ABV respectively. Column (4) shows estimates from the inter-brand experiments without brand dummies included the model, whereas column (5) includes brand dummies in the specification.
Standard errors in parentheses: $^*p < 0.05$

2.4.2 Treatment 2: Inter-brand

The results of the same mixed logit specification as column 1 but using the data from the inter-brand treatments are shown in column 4; although brands were included in the experiment, no brand fixed effects are included in the model. In column 5 we add brand dummies to account for brand fixed effects. The addition of brand dummies captures a large proportion of unobserved (to the researcher) effects. The difficulty is that the number of parameters to estimate increases in proportion to the number of brands, and characteristics that are fixed across choice situations are difficult to identify. This second problem requires the use of a minimum distance procedure (Chamberlain (1982), Nevo (2000)) to estimate taste coefficients β. We first estimate a $J*1$ vector of brand dummy coefficients, $d=(d_1,...,d_j)'$ using the previously described mixed logit procedure. From the original indirect utility equation 2.3.1 it follows that

$$d = X\beta + \xi, \tag{2.4.1}$$

where X is a $J*K$ matrix of product characteristics that are fixed and ξ' is a vector of $J*1$ unobserved product characteristics. Assuming that $E(\xi|X)=0$ then

$$\hat{\beta} = (X'{V_d}^{-1}X)^{-1}(X'{V_d}^{-1}\hat{d}), \qquad \hat{\xi} = \hat{d} - X\hat{\beta} \tag{2.4.2}$$

The difference between columns 1 and 4/5 are stark. The magnitude of the mean value of price is much smaller at -0.275 and -0.220 compared to -1.027. Without brand dummies the magnitude of the coefficients on ABV and container are much larger than without. The mean coefficients for container and unit volume also reverse signs between the inter and intra-branded experiments. Finally the standard deviations are all smaller in the inter-brand experiment save for unit volume. As per Nevo (2000) we consider the specification with brand dummies going forward for the inter-brand treatment.

2.4.3 Unobserved heterogeneity

The model predicted shares using the demand parameters from 2.5 are some way off from the observed market shares. This is unsurprising given our sample population is not representative of the US population, especially given the gender bias revealed previously. Even if the sample population were representative, unobserved heterogeneity results in discrepancies between model predicted and observed market shares. In the inter-brand treatment this heterogeneity comes from unobserved (to the researcher) product characteristics. In the intra-brand treatment, heterogeneity can arise if subjects have prior assumptions or make their choice based on anything other than the specified product characteristics.[10] To account for this heterogeneity and bring the model predicted shares in line with observed share we calculate δ and iterate the contraction mapping in equation 2.3.4 to solve the $J-$system of non-linear equations that give a vector of mean utilities, δ^* that minimises the difference between predicted and observed shares. Going forward, in the post-estimation calculation for elasticities, markups and during the merger simulation, when we are required to calculate utilities we use this δ^*.

2.4.4 Substitution patterns and markups

We then use the various demand estimates to calculate price elasticity matrices and price cost margins for our set of pseudo-products and a set of real products (a) using the intra-brand estimates from column 1 of Table 2.5 and (b) using the inter-brand with brand dummy estimates and compare them to existing estimates from previous studies. The real product set contained 18 of the most popular beers in the US plus an outside good matching the size of the pseudo-product set. Ownership of the brands was split between five firms; AB InBev, Molson Coors, Constellation Brands, Heineken and Blue Ribbon specified in the ownership matrix Φ. It must be noted that studies on aggregate

[10] As we discuss later, choosing meaningful product characteristics becomes vitally important.

data use observations from the entire population while our sample was restricted to ages 21-30. Table 2.7 presents a sample of the estimated elasticity matrix for the real products using the intra-branded experiment estimates. Tables 2.B.2 and 2.B.3, in the appendix presents the same for the set of pseudo-products and the real products using parameters estimates from the inter-brand specification with brand dummies. Each entry i, j, where i indexes the row and j indexes the column, gives the elasticity of brand i with respect to a change in the price of j. As the full matrix is too large to include here, only columns of brands owned by the two largest manufacturers ABInBev (green) and Molson Coors (orange) are shown in the table as these products were most scrutinised following the joint-venture between Miller and Coors investigated by Miller and Weinberg.[11] We can see evidence of the flexibility of the mixed logit in the heterogeneity in cross-price elasticities that exists within a single column. We also compare our estimates with those achieved by Miller and Weinberg (2017) in a study that uses a random coefficient nested logit model to compare predictions from demand estimation to ex-post merger price effects. Own-price elasticities for a selection of products that appear in both studies as well as summary statistics are presented in Table 2.6.

We do not present this comparison as a benchmarking exercise. As we have mentioned before, the product sets, sample populations, time periods and product characteristics are all too different between our study and that of Miller-Weinberg to make direct comparisons and hypothesis testing is not possible. We include these here to show our estimates are *broadly* in line with previous studies as an illustration that our methodology produces what appears to be realistic estimates of elasticities.

Indeed, when we look at the summary statistics for our intra-brand real set, the median own-price elasticity at -6.55% is somewhat higher in magnitude compared to Miller-Weinberg's range. However, given the changes in the

[11] The brands in red, blue and pink are owned by Constellations Brands, Heineken and Pabst, respectively.

Table 2.6: Comparison of beer elasticity estimates

	(1) Pseudo	(2) Intra-brand	(3) Inter-brand	(4) Miller-Weinberg
Own-price elasticities				
Bud Light		-4.020	-1.116	-4.389
Coors Light		-6.142	-1.487	-4.628
Miller Lite		-6.479	-4.081	-4.517
Budweiser		-6.628	-1.468	-4.272
Michelob Ultra		-4.678	-1.025	-4.970
Corona Extra		-5.595	-1.086	-5.178
Heineken		-6.200	-1.158	-5.147
Miller High Life		-8.673	-1.148	-3.495
Coors Banquet		-7.907	-1.084	-4.371
Summary Statistics				
Median Own-PED	-4.71	-6.55	-1.39	-4.73 – -4.33
Mean PCM		22.39%	81.5%	34%
Median PCM		20.22%	91.8%	

Abbreviations: PED is price-elasticity of demand; PCM is price cost margin

industry in the intervening decade in which the demand for light lagers has fallen, to be replaced by growth in craft beers and microbreweries (Watson, 2018), the higher price elasticities make sense. The pseudo-product set has a very similar median-own price elasticity. This suggests that if the aim to get a general understanding of a market rather than make predictions about specific products the kind of experiment we conducted in treatment 1 can be useful. Despite this, our model struggled to accurately predict market shares of beers because some beers with similar observed characteristics had markedly different actual market shares suggesting factors other than our observed characteristics were driving choices. Unobserved characteristics such as taste are most likely to be the cause as well as an unrepresentative sample. When we use the estimates

Table 2.7: Intra-brand parameters applied to real product set elasticity matrix

	Brand	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Bud Light	-4.02	0.327	0.497	0.082	0.235	0.101	0.063	0.575	0.276	0.061	0.046	0.045	0.047
2	Budweiser	1.124	-6.628	0.167	0.524	0.599	0.355	0.121	1.029	0.561	0.197	0.208	0.137	0.215
3	Michelob Ultra	1.116	0.109	-4.678	0.019	0.083	0.03	0.107	0.21	0.376	0.014	0.054	0.07	0.016
4	Natural Light	0.671	1.248	0.07	-7.398	0.601	0.725	0.045	1.135	0.909	1.211	0.428	0.056	0.18
5	Busch Light	1.353	1.004	0.213	0.423	-7.021	0.318	0.103	1.081	0.582	0.182	0.167	0.09	0.144
6	Busch	1.145	1.17	0.151	1.004	0.626	-8.088	0.077	1.252	0.813	0.479	0.285	0.076	0.168
7	Stella Artois	0.647	0.363	0.492	0.057	0.184	0.07	-6.159	0.361	0.645	0.019	0.152	0.203	0.052
8	Coors Light	1.488	0.776	0.243	0.359	0.487	0.286	0.091	-6.142	0.571	0.169	0.14	0.069	0.112
9	Miller Lite	0.811	0.479	0.492	0.326	0.297	0.21	0.184	0.647	-6.479	0.157	0.509	0.157	0.069
10	Keystone Light	1.011	0.949	0.104	2.452	0.524	0.7	0.031	1.079	0.887	-8.314	0.374	0.04	0.137
11	Miller High Life	0.602	0.799	0.317	0.69	0.383	0.331	0.195	0.715	2.289	0.298	-8.673	0.231	0.115
12	Blue Moon	0.657	0.586	0.457	0.1	0.231	0.099	0.291	0.392	0.786	0.036	0.258	-6.865	0.084
13	Coors Banquet	1.124	1.495	0.167	0.524	0.599	0.355	0.121	1.029	0.561	0.197	0.208	0.137	-7.907
14	Corono Extra	0.78	0.28	0.615	0.055	0.169	0.068	0.205	0.368	0.677	0.022	0.136	0.147	0.04
15	Modelo Especial	0.78	0.28	0.615	0.055	0.169	0.068	0.205	0.368	0.677	0.022	0.136	0.147	0.04
16	Heineken	0.689	0.419	0.518	0.061	0.189	0.072	0.258	0.348	0.635	0.022	0.168	0.246	0.06
17	Dos Equis	0.853	0.259	0.655	0.077	0.18	0.084	0.183	0.424	0.851	0.034	0.162	0.119	0.037
18	Pabst Blue Ribbon	0.923	1.597	0.115	1.3	0.686	0.585	0.084	1.215	0.807	0.545	0.35	0.101	0.23
19	Outside	0.09	0.047	0.02	0.139	0.045	0.05	0.007	0.128	0.125	0.068	0.027	0.003	0.007
	Median X-PeD	0.832	0.533	0.280	0.233	0.266	0.156	0.114	0.611	0.661	0.113	0.168	0.110	0.077
	Mean X-PeD	0.881	0.677	0.328	0.458	0.349	0.250	0.132	0.686	0.724	0.207	0.212	0.115	0.097

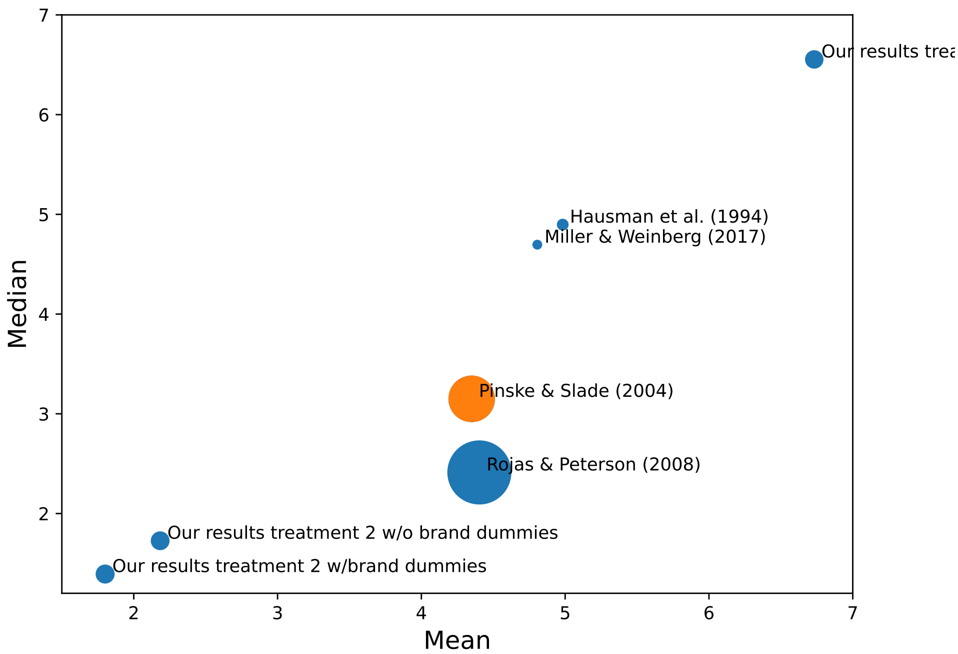
from the inter-brand experiment with brand dummies to calculate elasticities we can see the impact that these differences have. The median own-price elasticity is now -1.39 compared to -5.32. We attribute these changes to the non-incentivisation of our experiment. At the inter-brand level, when there is no consequence to a subjects wealth, it appears they pick their favourite brand regardless of price, and for reasons not captured by our observed product characteristics. This is supported by the fact that in general subjects are less sensitive to changes in the observed characteristics in the inter-brand experiment as seen by the smaller absolute values of the taste coefficients. Therefore, for our purposes we prefer the intra-brand experiment as it focuses subjects on the observed product characteristics, especially price, which is crucial for downstream estimation of elasticities and the merger simulation.

While there exist many papers that estimate elasticities on the market for beer as a whole,[12] there are fewer papers that estimate elasticities for a set of differentiated products. Figure 2.3 shows mean and median own-price elasticity estimates from studies that have previously estimated a differentiated demand system in the beer industry in the US or UK. Markers in blue are from the US whereas markers in orange are from the UK. The size of the marker represents the standard deviation of own-price elasticities i.e. studies with smaller standard deviations are represented by smaller markers. Although our intra-brand results are higher than other comparisons, again for reasons mentioned before including changes in market structure and the make-up of our sample we find they are less of an outlier than the inter-brand results with brand dummies which supports our preference to use the intra-brand estimates for the proceeding merger simulation.

Finally, we use the elasticity matrix to calculate marginal costs using equation 2.3.7 for the real product sets. (It is not possible to do this for the pseudo-set as there is no ownership matrix). We obtain median and mean price cost

[12]see Fogarty (2010) and Nelson (2014) for two meta-studies that compare market elasticites across countries and time-periods.

Figure 2.3: Comparison of differentiated demand estimates with previous studies



Markers in blue are from the US; markers in orange are from the UK. The size of the marker represents the standard deviation of own-price elasticities. Tighter standard deviations are shown by smaller markers.

margins of around 22-23% in the intra-brand version compared to 34% in Miller and Weinberg. This is equivalent to a markup of around \$3.50 compared to an average markup of \$3.60 in Miller and Weinberg. In the inter-brand experiment with brand dummies, this increases to 60% further reinforcing our idea that for our purposes the non-branded experiment is preferable.

2.4.5 Merger Simulation

With all the ingredients in place we are able to simulate the effects of a potential merger between firms in the industry using the unbranded elasticities. As an illustrative exercise, we choose to observe the effects of a merger between Molson-Coors and Constellation Brands.

Table 2.8: Simulated merger between Molson-Coors and Constellation Brands

	Pre-merger values				Post-merger values				
	Price	Share	MC	PCM	New Firm	Price	Share	PCM	%Δ Price
Bud Light	15.99	13.24	10.48	34.5	-	15.09	23.74	30.6	-5.62
Budweiser	11.99	5.14	8.58	28.4	-	12.54	6.51	31.6	4.59
Michelob Ultra	18.99	4.97	12.98	31.6	-	18.88	7.1	31.3	-0.6
Natural Light	7.99	3.24	5.83	27	-	8.61	3.83	32.3	7.75
Busch Light	11.99	3.07	8.49	29.2	-	12.53	3.72	32.2	4.47
Busch	9.99	1.87	7.21	27.8	-	10.61	2.19	32	6.17
Stella Artois	15.99	1.29	11.91	25.5	-	17.23	1.49	30.9	7.77
Coors Light	11.99	6.82	9.64	19.6	CC	15.15	2.05	36.4	26.37
Miller Lite	11.99	6.02	9.59	20	CC	15.55	1.72	38.3	29.67
Keystone Light	7.99	1.6	6.55	18	CC	10.11	0.32	35.2	26.55
Miller High Life	10.99	1.46	8.81	19.8	CC	13.57	0.58	35.1	23.48
Blue Moon	14.99	0.96	12.14	19	CC	18.84	0.38	35.6	25.67
Coors Banquet	11.99	0.74	9.84	17.9	CC	14.76	0.25	33.3	23.11
Corono Extra	15.99	4.95	12.73	20.4	CC	20.21	1.97	37	26.41
Modelo Especial	15.99	4.56	12.73	20.4	CC	20.21	1.81	37	26.41
Heineken	15.99	1.78	13.36	16.4	-	17.71	1.68	24.6	10.75
Dos Equis	14.99	0.9	12.56	16.2	-	16.98	0.7	26	13.29
Pabst Blue Ribbon	9.99	0.97	8.87	11.2	-	11.42	0.58	22.3	14.28

*Firms are colour-coded as follows: ABInBev; Molson-Coors; Constellation Brands; Heineken; Pabst. CC is the new firm arising through the merger of ABInBev and Molson-Coors. PCM is price-cost margin $= (p - c)/p$ expressed as a percentage.

To simulate a merger, we change the ownership matrix, $\mathbf{\Phi}$ to reflect the brands that would be under common ownership, and solve equation 2.3.7 to predict the new prices and quantities. Table 2.8 shows our predictions where the merged entity has the same marginal costs as pre-merger. The new entity, referred to as CC in the table, now owns 8 of the top 18 brands in the market. Nearly, all prices rise, with an average price increase of 15%, and the total market share of the top brands falls from 63.6% to around 61%. Again, we reiterate that these results are illustrative only and demonstrate that using the demand estimates from our experiment, augmented with unobserved heterogeneity can produce convergent merger simulations. Given the sample population and the specific product characteristics used, these results should not be used to make any conclusions or prediction about the real-world beer market.

2.5 Discussion

As we have discovered, brand effects and the issue of incentivising the experiment go hand in hand. In our inter-brand treatment we saw that subjects appear to undervalue the importance of price when brands are present, a result that is also mentioned in Moshary et al. (2022). The reason for this is because the experiment was entirely hypothetical and subjects were not required to make any purchases based on the choices they made. This, of course is in contrast to what they would experience purchasing beer in the supermarket where there is a trade-off between preferences for a particular brand and the associated price. For example, an individual's all else equal 'favourite' brand may also be the most expensive, while their close second favourite brand is significantly cheaper such that they usually purchase the second favourite. This type of behaviour is not captured by a non-incentivised branded experiment because there is no consequential trade-off between price and brand; the hypothetical individual in the previous example would choose his favourite brand because the price is not relevant to him - the brand effect supersedes all other product characteristics. Branding provides signals to consumers about product quality and enables firms to charge higher prices. In the real-world we are able to calculate the willingness to pay for these brands through revealed preference data but this is not transferable to the current non-incentivised experiment. Typical structural empirical work includes prices on the right hand side, but does not usually have advertising information, so that advertising information then sits in the error term. These studies almost always use instrument variables which makes price orthogonal to the error term, such that they are able to capture the price effect consistently. In contrast, to understand the coefficient on the price more accurately, we must focus on the intra-brand treatment because in the inter-brand treatment people are focused on the wrong aspect of the purchase decision. A question then for further investigation is whether there is a design that allows for brands to be included alongside some form of incentivisation? Regardless, the success of these models depends on correctly identifying relevant observable product characteristics and the density of preferences for these characteristics in the population (Train, 2009). We were limited to measurable characteristics but in future the use of machine learning techniques and deep neural nets could allow the inclusion of qualitative characteristics to synthesise substitution patterns.

The use of similar techniques by agencies around the world is varied. Survey methods, questionnaires and experiments have been used more often in the UK by the Competition Markets Authority, (CMA, 2018) than in the US.[13] As Imthorn et al. (2016) argue, the specific use of discrete choice experiments avoids many of the biases inherent in more general survey methods. In their guidelines, the CMA discuss best practices for recruitment of subjects and questionnaire-type survey design including identifying choice attributes by either asking consumers to identify the most important reason(s) for their purchase or how important each attribute is to the customer, using a categorical scale.[14] However, the guidelines suggest discrete choice experiments of the type we use are not extensively used because of time constraints. We show that once an effective framework is designed, an experiment can be quickly deployed for goods or markets that exhibit similar characteristics.

The guidelines also express concerns with finding representative samples using online surveys. There are however, several ways in which our current methodology can be easily adapted. For certain demographics/products current online platforms including Prolific and Amazon MTurk will allow for representative samples - particularly as the numbers on these platforms are growing. But competition authorities that already use questionnaires in their analysis can utilise their existing recruitment methods. As the experiment only needs a mobile device and internet connection to administer and takes between 10-15 minutes to complete, customers can even be intercepted outside of stores

[13] In *U.S. v. H&R Block Inc., et al., D.D.C.* (2011), Judge Howell criticises several elements of a survey used by defendant in the merger case including the leading nature of the questions and potential biases in recruitment.

[14] For example 'essential', 'very important', 'fairly important' and so on.

if an agency so wishes. Else, participants can be drawn from third-party marketing firms such as YouGov in the UK or Harris Poll in the US that have large subject lists with considerable geographic and demographic coverage. The proliferation of video calling and online meetings even allows for the adaptation of interviewer led telephone surveys to our methodology which enables researchers to sample potentially more representative populations than are available through online recruitment platforms. However, this is beyond the scope of this paper. Agencies can even recruit participants to a physical lab(s) across the country - the only requirement is that there are sufficient devices on which subjects can make choices, although this is considerably more expensive and time-consuming. The limited success of the ACM and NMa shows that experiments can be useful, certainly in various types of merger cases, but also in answering other policy questions. Beyond understanding incentives and the role of brands in experimental preferences many of the underlying capabilities required for implementation already exist within mature competition agencies, particularly subject recruitment infrastructure and coding abilities. Nevertheless, there has not been a broader uptake elsewhere in the European Commision, or further afield suggesting that perhaps the biggest challenge for experiments is to convince institutions to begin using them. We hope that this paper, alongside others advocating similar methods, can begin to do just that.

2.6 Conclusions

So far we have presented a background and methodology that can be used to estimate demand parameters and utilise these estimates in further analysis relevant to merger evaluation. Among our primary goals was to simplify the process so that it could be easily adapted and replicated.

Although we were able to obtain estimates of taste parameters, elasticities and PCMs, the observed characteristics we used for beer did not always accurately predict market shares, even when we used the set of real products. While the characteristics we used were guided by previous studies and a survey published by the Craft Brewing Business, it was apparent that many products in the real world were very similar in these characteristics. Despite these apparent similarities however, the products enjoyed different market shares. These differences must be a result of unobserved factors such as taste and branding. Although it is difficult to measure taste, information about this may be conferred through the brand for well known brands. Since we used unlabelled alternatives in our initial choice sets, we have no information about specific brand fixed effects. This may be sufficient if the aim is to simply obtain demand estimates which could be confounded by brand effects.

The addition of brands saw our calculated elasticities move significantly away from both our non-branded results and those from Miller-Weinberg. It appears some factor in the presentation of the choice sets leads to consumers only considering the brand such that the other product characteristics, including price are less salient in the choice. An area for further research is to explore alternative experimental designs to solve this issue. It is infeasible and inadvisable to provide subjects with a choice set of all brands in the market. If however, we consider the purpose of the demand estimation to be evaluating a merger and that a merger will only come to the attention of regulators when there are competition implications, then a possible solution might be that only the largest brands in a market need be considered. Therefore it should be possible to present subjects with a choice of, for example, the top 10 brands in a single choice set, while the other product characteristics are allocated randomly as before. This would allow brand fixed effects for each of these 10 products to be estimated and lead to more accurate predicted market shares.

One problem with using brand dummies that we mentioned earlier was that it confounds identification of demand parameters. However, Nevo (2000) provides an elegant solution to this using a two stage projection method. First, the brand dummy coefficients and their variance-covariance matrix is estimated. Then a GLS regression is used to retrieve the taste parameters where the brand dummies are the independent variable and the number of observations is the number of brands used. Even in an empirical model, however, this restricts the number of observations. Where we suggest using only the top J brands the ability of this method to identify taste parameters must be examined further. The requirement of brands may vary by industry such that where observable product characteristics are more salient in consumer decisions, correct specification of these characteristics may result in a sufficiently identified model.

From an estimation perspective, there are several alternative methods we could explore. Rather than maximum likelihood, hierarchical Bayes estimation can be used and should achieve the same results if the model is correctly specified and identified. Even with an SLL estimation there are a number of different algorithms and methods for drawing from sample in simulation that can be tested. However, in our experience the marginal gains can often be small if there are sufficient observations.

In general we have presented a method that uses experimental data to estimate demand parameters and useful measures in merger evaluation quickly, with some degree of success. We managed to obtain estimates of elasticities and markups that appear to be realistic when compared to previous studies. However, there are several areas in the very simple experiment we conducted that could be improved to enhance the accuracy of estimates further. The precise experimental requirements are likely to be industry dependant, and indeed the model only suited to consumer goods, but once a satisfactory experiment has been designed it can be easily reworked to the specific products in question to provide guidance in initial merger evaluations.

References

Arkes, H. R., C. A. Joyner, M. V. Pezzo, J. G. Nash, K. Siegel-Jacobs, and E. Stone (1994). "The psychology of windfall gains". *Organizational behavior and human decision processes* 59(3), pp. 331–347.

Armona, L., G. Lewis, and G. Zervas (2021). "Learning Product Characteristics and Consumer Preferences from Search Data". *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 98–99.

Bajari, P., Z. Cen, V. Chernozhukov, M. Manukonda, S. Vijaykunar, J. Wang, R. Huerta, J. Li, L. Leng, G. Monokroussos, et al. (2023). "Hedonic prices and quality adjusted price indices powered by AI". *arXiv preprint arXiv:2305.00044*.

Bajari, P., J. T. Fox, and S. P. Ryan (2007). "Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients". *American Economic Review* 97(2), pp. 459–463.

Bastin, F., C. Cirillo, and P. L. Toint (2010). "Estimating nonparametric random utility models with an application to the value of time in heterogeneous populations". *Transportation science* 44(4), pp. 537–549.

Bergman, A., A. Chinco, S. M. Hartzmark, and A. B. Sussman (2020). *Survey Curious? Start-Up Guide and Best Practices For Running Surveys and Experiments Online*. Available at SSRN: https://ssrn.com/abstract=3701330. [Accessed 15-April-2021].

Berry, S., J. Levinsohn, and A. Pakes (1995). "Automobile prices in market equilibrium". *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Berry, S. T. (1994). "Estimating Discrete-Choice Models of Product Differentiation". *The RAND Journal of Economics* 25(2), pp. 242–262. ISSN: 07416261. URL: http://www.jstor.org/stable/2555829.

Bradley, M. A. and A. J. Daly (1997). "Estimation of logit choice models using mixed stated preference and revealed preference information". *Understanding* *travel behaviour in an era of change*. Ed. by P. Stopher and M. Lee-Gosselin. Oxford: Pergamon, pp. 209–232.

Chamberlain, G. (1982). “Multivariate regression models for panel data”. *Journal of econometrics* 18(1), pp. 5–46.

Chung, C., T. Boyer, and S. Han (2011). “How many choice sets and alternatives are optimal? Consistency in choice experiments”. *Agribusiness* 27(1), pp. 114–125.

CMA (May 2018). *Good practice in the design and presentation of customer survey evidence in merger cases*. URL: https://www.gov.uk/government/publications/mergers-consumer-survey-evidence-design-and-presentation/good-practice-in-the-design-and-presentation-of-customer-survey-evidence-in-merger-cases.

Conlon, C. and J. Gortmaker (2020). “Best practices for differentiated products demand estimation with PyBLP”. *The RAND Journal of Economics* 51(4), pp. 1108–1161. DOI: https://doi.org/10.1111/1756-2171.12352. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12352. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12352.

Conlon, C. and J. H. Mortimer (2021). “Empirical properties of diversion ratios”. *The RAND Journal of Economics* 52(4), pp. 693–726.

Conlon, C. T. and J. H. Mortimer (2013). “Demand estimation under incomplete product availability”. *American Economic Journal: Microeconomics* 5(4), pp. 1–30.

De Bekker-Grob, E. W., L. Hol, B. Donkers, L. Van Dam, J. D. F. Habbema, M. E. Van Leerdam, E. J. Kuipers, M.-L. Essink-Bot, and E. W. Steyerberg (2010). “Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening”. *Value in Health* 13(2), pp. 315–323.

Fogarty, J. (2010). “The demand for beer, wine and spirits: a survey of the literature”. *Journal of Economic Surveys* 24(3), pp. 428–478.

Fosgerau, M. and M. Bierlaire (2007). "A practical test for the choice of mixing distribution in discrete choice models". *Transportation Research Part B: Methodological* 41(7), pp. 784–794.

Fosgerau, M. and S. L. Mabit (2013). "Easy and flexible mixture distributions". *Economics Letters* 120(2), pp. 206–210.

Hensher, D. A. (1994). "Stated preference analysis of travel choices: the state of practice". *Transportation* 21(2), pp. 107–133.

Hensher, D. A. and W. H. Greene (2003). "The mixed logit model: the state of practice". *Transportation* 30(2), pp. 133–176.

Hensher, D. A. and E. Cherchi (2015). "Stated preference surveys and experimental design: an audit of the journey so far and future research perspectives".

Hess, S., D. A. Hensher, and A. Daly (2012). "Not bored yet – revisiting respondent fatigue in stated choice experiments". *Transportation Research Part A: Policy and Practice* 46(3), pp. 626–644.

Holmes, T. P., W. L. Adamowicz, and F. Carlsson (2017). "Choice experiments". *A primer on nonmarket valuation*, pp. 133–186.

Imthorn, M., R. Kemp, and I. Nobel (Apr. 2016). *Using Conjoint Analysis in Merger Control*. URL: https://www.acm.nl/sites/default/files/old_publication/publicaties/15747_acm-working-paper-conjoint-analysis-2016-04-25.pdf.

Lerro, M., G. Marotta, and C. Nazzaro (2020). "Measuring consumers' preferences for craft beer attributes through Best-Worst Scaling". *Agricultural and Food Economics* 8(1), pp. 1–13.

Letzter, R. (Sept. 2021). *A teenager on TikTok disrupted thousands of scientific studies with a single video*. URL: https://www.theverge.com/2021/9/24/22688278/tiktok-science-study-survey-prolific.

Louviere, J. J., D. A. Hensher, and J. D. Swait (2000). *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press.

Magnolfi, L., J. McClure, and A. T. Sorensen (2022). "Embeddings and Distance-based Demand for Differentiated Products". *Available at SSRN 4113399*.

McFadden, D. and K. Train (2000). “Mixed MNL models for discrete response”. *Journal of applied Econometrics* 15(5), pp. 447–470.

Miller, N. H. and M. C. Weinberg (2017). “Understanding the price effects of the MillerCoors joint venture”. *Econometrica* 85(6), pp. 1763–1791.

Moshary, S., B. Shapiro, and S. Drango (2022). “Preferences for Firearms and Their Implications for Regulation”. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (115).

Nelson, J. P. (2014). “Estimating the price elasticity of beer: Meta-analysis of data with heterogeneity, dependence, and publication bias”. *Journal of Health Economics* 33(C), pp. 180–187. DOI: 10.1016/j.jhealeco.2013.1.

Nevo, A. (2000). “A practitioner’s guide to estimating random-coefficient logit models of demand”. *Journal of Economics and Management Strategy* 9(4), pp. 513–548.

— (2001). “Measuring market power in the ready-to-eat cereal industry”. *Econometrica* 69(2), pp. 307–342.

Petrin, A. (2002). “Quantifying the benefits of new products: The case of the minivan”. *Journal of political Economy* 110(4), pp. 705–729.

Qiu, J., M. Sawada, and G. Sheu (2021). “Win/Loss Data and Consumer Switching Costs: Measuring Diversion Ratios and the Impact of Mergers”. *Masayuki and Sheu, Gloria, Win/Loss Data and Consumer Switching Costs: Measuring Diversion Ratios and the Impact of Mergers (November 6, 2021).*

Reynaert, M. and F. Verboven (2014). “Improving the performance of random coefficients demand models: the role of optimal instruments”. *Journal of Econometrics* 179(1), pp. 83–98.

Rose, J. M. and M. C. Bliemer (2009). “Constructing efficient stated choice experimental designs”. *Transport Reviews* 29(5), pp. 587–617.

Rossi, P. E. (2014). “Even the rich can make themselves poor: A critical examination of IV methods in marketing applications”. *Marketing Science* 33(5), pp. 655–672.

Schulze, W. D., R. C. d'Arge, and D. S. Brookshire (1981). "Valuing environmental commodities: some recent experiments". *Land Economics* 57(2), pp. 151–172.

Train, K. (2016). "Mixed logit with a flexible mixing distribution". *Journal of choice modelling* 19, pp. 40–53.

Train, K. E. (2008). "EM algorithms for nonparametric estimation of mixing distributions". *Journal of Choice Modelling* 1(1), pp. 40–69.

— (2009). *Discrete Choice Methods with Simulation*. 2nd Edition. New York: Cambridge University Press.

U.S. v. H&R Block Inc., et al., D.D.C. (2011). URL: https://www.justice.gov/atr/case-document/memorandum-opinion-2.

Watson, J. (Jan. 2018). *The Great Fragmentation of Beer*. URL: https://research.rabobank.com/far/en/sectors/beverages/great-fragmentation-of-beer.html.

Appendix

2.A Simulation

While this will not help in selecting appropriate product characteristics it will help choosing the number and combination of attributes and choice sets as well as an idea of the required number of observations. 'Participants' are generated and assigned values of tastes and preferences for the observed characteristics, drawn from distributions specified by the researcher. Additional noise, drawn from a standard Type 1 extreme value (Gumbel) distribution is assigned per participant per alternative per choice set. This 'birthing' of respondents can be repeated as many times as required for the sample size. Each participant is presented with repeated choice situations as in the real experiment. Each choice situation contains four alternatives and the program chooses the alternative that has the highest utility based on the preferences of each individual in the sample.[15] This process is repeated over n participants and t repetitions per person to obtain observations $= n * t$. Once the data is obtained, analysis is via the process outlined in section 3; we use SLL of mixed logit probabilities to estimate mean and standard deviation of the distribution in the population with the aim of estimating parameters from the previously specified distribution as consistently and efficiently as possible

In order to test for consistency and efficiency of parameter estimates, for a given

[15] Choice behaviour need not be utility maximisation - the model simply describes the relation of explanatory variables to the outcome of a choice, without reference to how the choice is made.

seed, we presented increasing numbers of computer generated participants with a single choice and estimated the value of the mean and standard deviation of $f(\beta)$ in the population and the associated standard errors, presented in figures A.1 and A.2. In each graph, the red line represents the true parameter values (-5 and 1 respectively). We can clearly see that as the number of observations increases, the parameter estimates converge quickly to the true values, for both mean and standard deviation. We can also see that as the number of observations increase, the standard errors of the estimate, denoted by the gold bars, reduce significantly. Together, these results indicate that we can achieve consistency and efficiency using this model and data collected in a similar fashion.

The next step was to ensure that these results were not as a result of peculiar phenomenon occurring within the particular seed we had randomly chosen. In order to test this, we repeated the experiment 50 times each for specified combinations of participants, n and choice sets, t and then reported mean values for parameter estimates and standard errors. The results of this exercise are presented in table A.1. The pre-specified, 'true' values are given in parentheses next to the name of the attribute. The first 4 columns show the results for increasing numbers of participants each making a single choice. This is essentially the same as presented in figures A.1 and A.2 except the experiments have been repeated 50 times with different samples. The key thing to note is that as we move from column 1 to 4, the mean value of the mean price coefficient approaches -5, the mean value of the standard deviation of the price coefficient approaches 1, and the standard errors, denoted in parentheses for each parameter, drop significantly. Of course, as we mentioned earlier it is impractical to only ask one choice of each participant, so we conduct our 50 repetitions for different combinations of n and t, shown in columns 5-8. What we can see is that if we can obtain at least 1000 observations then the point estimates for mean and standard deviation are very close to the true values in the population. Increasing the observations to 5000 serves to improve the standard errors. We focus on the price coefficient as the observed characteristic of most interest, however, it can be seen that the mean point estimates for μ and σ all converge to their true values as the number of observations increases for all observed attributes. Similarly, the standard errors all decrease significantly as we move from left to right from column 1 to column 8. This suggests the mixed logit of the experimental data is able to accurately derive the true population parameters, θ^*.

Figure 2.A.1: Estimates of population mean of price coefficient for increasing sample sizes

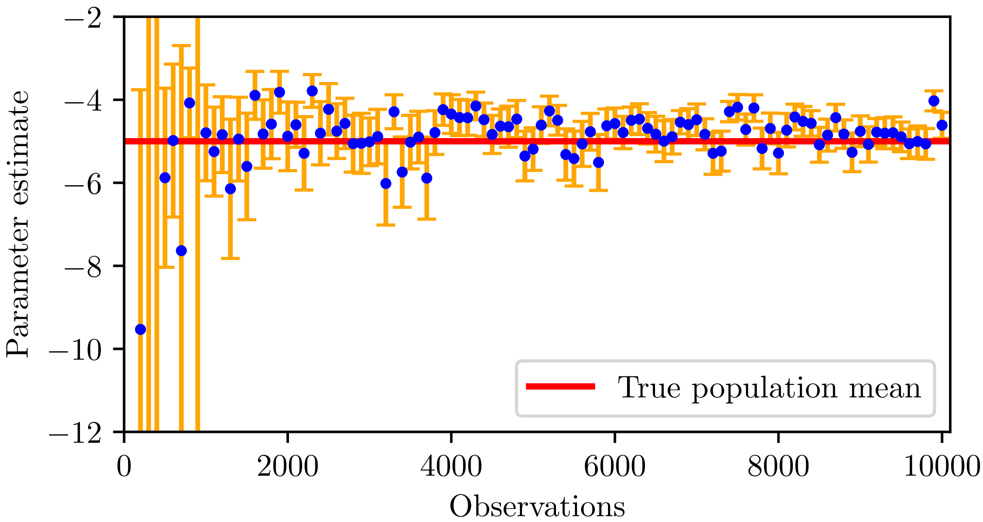
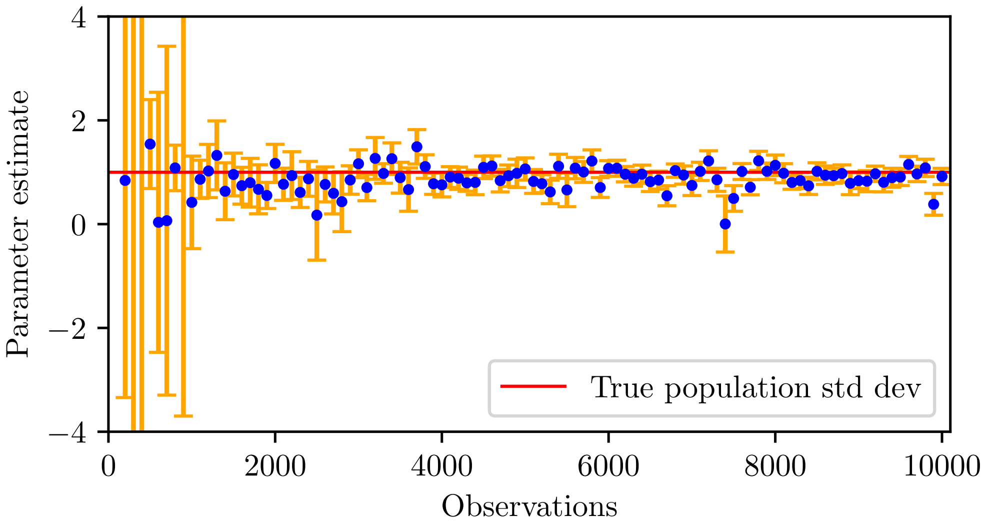


Figure 2.A.2: Estimates of population standard deviation of price coefficient for increasing sample sizes



(a) *Note*: Dots represent point estimates of parameters and gold bars represent standard errors.

Table 2.A.1: Mean values of the estimated parameters on all coefficients over 50 samples

	1	2	3	4	5	6	7	8
n	200	500	1000	5000	200	1000	100	500
t	1	1	1	1	5	5	10	10
observations $= n * t$	200	500	1000	5000	1000	5000	1000	5000
Price (true values: $\mu = -5; \sigma = 1$)								
mean	-20.34	-8.866	-5.501	-4.729	-5.195	-4.928	-5.407	-5.003
	(124.6)	(17.99)	(1.634)	(0.485)	(0.832)	(0.312)	(0.746)	(0.262)
standard deviation	2.865	1.088	0.928	0.863	0.924	1.024	1.164	1.127
	(24.96)	(2.542)	(0.858)	(0.232)	(0.645)	(0.180)	(0.561)	(0.157)
ABV (true values: $\mu = 2; \sigma = 1.5$)								
mean	8.822	3.689	2.296	2.028	2.216	2.054	2.194	2.074
	(53.70)	(7.791)	(0.656)	(0.198)	(0.365)	(0.136)	(0.338)	(0.124)
standard deviation	6.712	2.835	1.711	1.434	1.615	1.518	1.598	1.514
	(42.39)	(6.653)	(0.750)	(0.234)	(0.379)	(0.145)	(0.345)	(0.124)
Can (true values: $\mu = 1.5; \sigma = 0.8$)								
mean	6.979	2.653	1.698	1.458	1.563	1.461	1.632	1.499
	(43.74)	(5.620)	(0.512)	(0.160)	(0.315)	(0.121)	(0.308)	(0.114)
standard deviation	6.760	2.504	1.232	0.766	0.996	0.853	1.064	0.881
	(2459)	(389.8)	(65.14)	(23.75)	(33.45)	(12.46)	(25.60)	(9.243)
Volume(true values: $\mu = 4; \sigma = 2.5$)								
mean	16.71	6.954	4.376	3.814	4.175	3.916	4.247	3.913
	(102.9)	(14.23)	(1.250)	(0.370)	(0.656)	(0.243)	(0.606)	(0.216)
standard deviation	10.99	4.692	2.827	2.444	2.684	2.549	2.653	2.565
	(64.41)	(11.07)	(0.992)	(0.302)	(0.537)	(0.199)	(0.489)	(0.173)

2.B Additional Tables

2.B.1 Conditional Distributions

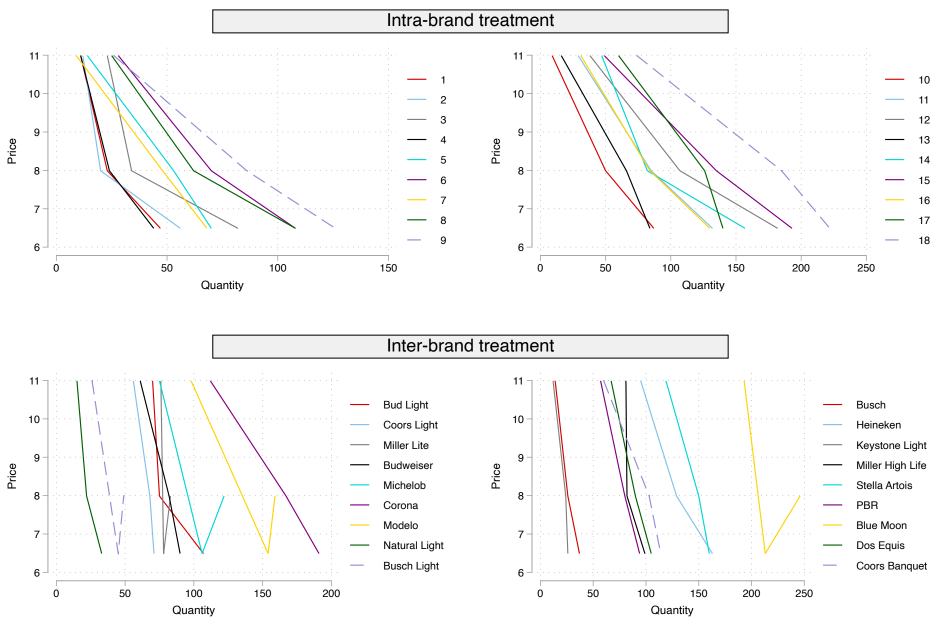
Using the point estimates of $\hat{\theta}$ from the intra-brand treatment we can calculate each subject's tastes conditional on the sequence of choices they made, presented in column 2 of Table 2.B.1, which shows the mean and standard deviation of the 486 individual coefficients, β_n. Column 1 includes the base intra-brand results from table 2.5 for easy comparisons. The means of β_n are very close to the population mean in all cases. This similarity is expected for a correctly specified and consistently estimated model. The standard deviations are considerably greater than zero and are also similar to their population counterparts. For example, the conditional estimate of the standard deviation on ABV is 1.116, and the population estimate of the standard deviation is 1.430. Thus, variation in $\bar{\beta}_n$ captures more than 78% of the total estimated variation in the coefficient.

Figure 3.4 shows a similar pattern for all the other coefficients. The dashed line shows the kernel density of the individual parameters and the standard deviation of this distribution is marginally less than the standard deviation of the equivalent population distribution. This shows that the mean of a subjects conditional distribution captures a large share of the variation in coefficients across subjects and has the potential to be meaningful in distinguishing customers.

Table 2.B.1: Mixed logit estimates

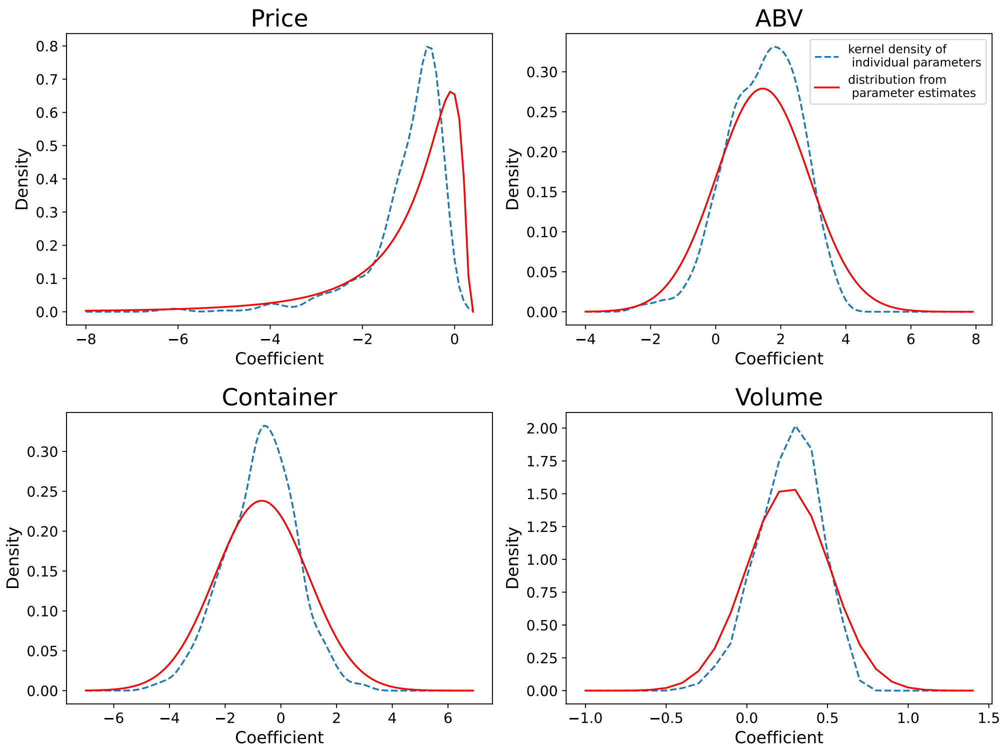
		Population	Individual
		(1)	(2)
Price	Mean (μ_α)	-1.027	-1.041
	SD (σ_α)	1.148	0.876
ABV	Mean (μ_{β_1})	1.444	1.468
	SD (σ_{β_1})	1.430	1.116
Container	Mean (μ_{β_2})	-0.686	-0.720
	SD (σ_{β_2})	1.675	1.266
Unit volume	Mean (μ_{β_3})	0.256	0.260
	SD (σ_{β_3})	0.257	0.190
Observations		3888	3888

Figure 2.B.2: Demand curves



Here we present the demand curves for all the products that appear in our

Figure 2.B.1: Comparison of population and individual parameters



treatments, based on Tables 2.3 and 2.4. Products from the intra-brand treatment appear in the top two panels while those from the inter-brand treatment appear in the bottom two panels. In each treatment, the 18 products are separated to increase legibility of the graphs. We see that in the intra-brand treatment, all the products exhibit well behaved demand functions that are downward sloping. In the inter-brand treatment, however, we see much more inelastic demand curves that in some cases are non-downward sloping. This supports our theory that in the inter-brand exercise, participants engage in cheap-talk, choosing their favourite brand while paying less attention to price.

Table 2.B.2: Pseudo product set elasticity matrix

Name	1	2	3	4	6	7	9	10	11	13	15	16	17	18
1	**-3.868**	0.077	0.166	0.013	0.043	0.097	0.208	0.025	0.030	0.230	0.395	0.065	0.070	0.120
2	0.042	**-4.144**	0.284	0.010	0.078	0.062	0.310	0.017	0.030	0.131	0.535	0.038	0.061	0.167
3	0.025	0.076	**-4.185**	0.006	0.134	0.031	0.435	0.009	0.025	0.059	0.679	0.018	0.045	0.218
4	0.025	0.036	0.083	**-3.752**	0.097	0.048	0.113	0.056	0.069	0.125	0.230	0.155	0.169	0.296
5	0.019	0.039	0.143	0.021	0.174	0.031	0.168	0.037	0.067	0.071	0.307	0.089	0.144	0.403
6	0.011	0.037	0.236	0.013	**-4.175**	0.015	0.236	0.019	0.055	0.031	0.384	0.04	0.104	0.513
7	0.041	0.047	0.087	0.010	0.025	**-4.629**	0.183	0.031	0.033	0.437	0.617	0.135	0.130	0.198
8	0.030	0.051	0.150	0.008	0.044	0.071	0.284	0.021	0.033	0.261	0.887	0.082	0.118	0.290
9	0.017	0.047	0.243	0.005	0.075	0.036	**-5.038**	0.011	0.028	0.122	1.190	0.039	0.090	0.397
10	0.018	0.022	0.043	0.021	0.053	0.055	0.098	**-4.511**	0.073	0.237	0.347	0.309	0.299	0.464
11	0.014	0.025	0.076	0.016	0.096	0.036	0.152	0.045	**-4.808**	0.141	0.496	0.186	0.270	0.670
12	0.008	0.023	0.125	0.010	0.163	0.018	0.222	0.024	0.061	0.065	0.661	0.088	0.203	0.904
13	0.020	0.021	0.035	0.006	0.01	0.091	0.129	0.028	0.027	**-4.961**	0.857	0.228	0.202	0.287
14	0.014	0.022	0.059	0.004	0.018	0.059	0.201	0.019	0.027	0.422	1.265	0.140	0.187	0.428
15	0.008	0.020	0.093	0.002	0.030	0.030	0.291	0.010	0.022	0.199	**-4.537**	0.067	0.144	0.591
16	0.009	0.010	0.017	0.012	0.022	0.047	0.069	0.061	0.059	0.377	0.475	**-4.992**	0.456	0.651
17	0.007	0.011	0.030	0.009	0.039	0.031	0.108	0.040	0.059	0.229	0.701	0.312	**-5.361**	0.966
18	0.004	0.010	0.048	0.005	0.065	0.016	0.157	0.021	0.048	0.108	0.957	0.148	0.321	**-4.803**

Table 2.B.3: Brand dummies elasticity matrix

	Brand	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Bud Light	**-1.116**	0.017	0.015	0.005	0.007	0.005	0.016	0.006	0.032	0.159	0.029	0.048	0.044
2	Budweiser	0.009	**-1.468**	0.014	0.017	0.011	0.011	0.025	0.009	0.095	0.290	0.038	0.079	0.053
3	Michelob Ultra	0.007	0.012	**-1.025**	0.003	0.005	0.003	0.011	0.005	0.021	0.124	0.026	0.035	0.040
4	Natural Light	0.007	0.043	0.009	**-6.698**	0.017	0.039	0.039	0.014	3.008	0.641	0.032	0.122	0.040
5	Busch Light	0.009	0.028	0.014	0.017	**-1.483**	0.011	0.025	0.009	0.095	0.289	0.037	0.078	0.053
6	Busch	0.009	0.037	0.014	0.055	0.015	**-2.140**	0.034	0.012	0.294	0.447	0.040	0.106	0.055
7	Stella Artois	0.009	0.027	0.014	0.017	0.011	0.011	**-1.467**	0.009	0.096	0.289	0.037	0.078	0.053
8	Coors Light	0.008	0.025	0.015	0.016	0.010	0.010	0.023	**-1.487**	0.099	0.300	0.038	0.072	0.055
9	Miller Lite	0.007	0.039	0.010	0.501	0.015	0.034	0.036	0.015	**-4.081**	0.651	0.033	0.112	0.042
10	Keystone Light	0.008	0.030	0.014	0.026	0.012	0.013	0.027	0.011	0.161	**-1.370**	0.041	0.085	0.056
11	Miller High Life	0.007	0.018	0.014	0.006	0.007	0.005	0.016	0.006	0.037	0.189	**-1.148**	0.051	0.047
12	Blue Moon	0.009	0.028	0.014	0.017	0.011	0.011	0.025	0.009	0.095	0.290	0.038	**-1.416**	0.053
13	Coors Banquet	0.007	0.016	0.014	0.005	0.006	0.005	0.015	0.006	0.031	0.167	0.031	0.046	**-1.084**
14	Corona Extra	0.007	0.016	0.014	0.005	0.006	0.005	0.015	0.006	0.031	0.167	0.031	0.046	0.046
15	Modelo Especial	0.007	0.016	0.014	0.005	0.006	0.005	0.014	0.006	0.030	0.167	0.031	0.046	0.045
16	Heineken	0.007	0.018	0.014	0.006	0.007	0.006	0.016	0.007	0.038	0.188	0.032	0.050	0.048
17	Dos Equis	0.009	0.038	0.013	0.055	0.015	0.019	0.034	0.012	0.294	0.448	0.041	0.107	0.055
18	Pabst Blue Ribbon	0.007	0.016	0.013	0.005	0.006	0.005	0.014	0.006	0.030	0.168	0.031	0.046	0.045
19	Outside	0.008	0.027	0.013	0.080	0.011	0.013	0.024	0.010	0.465	0.335	0.036	0.077	0.050

3

Evaluating the effectiveness of the high in fat, sugar and salt location restriction policy in England

Abstract

In a bid to reduce consumption of high in fat, salt and sugar foods (HFSS), the Department of Health and Social Care introduced a policy that restricts the sale of HFSS products in prominent locations within stores, such as near checkouts and store entrances, in October 2022. Using a dataset of weekly drinks sales from one of the largest supermarkets in the UK over a two year period, we analyse the effectiveness of this policy in reducing the consumption of sugar-sweetened beverages (SSBs) in a difference-in-difference framework. We construct several possible control groups, including non-SSBs in England and SSBs in unrestricted stores in both England and combined Scotland-Wales. We find the policy was effective in reducing consumption of immediate consumption drinks by 10%. However, these drinks comprise only 7% of total sales, and so overall we conclude the policy had no significant impact in reducing sales of SSBs in England. While the policy appears ineffective for SSBs, it covers a wide range of products, and so we cannot comment on the full effects of the policy.

3.1 Introduction

The UK Department of Health and Social Care (DHSC) in a recent policy paper described tackling obesity as 'one of the greatest long-term health challenges this country faces' (DHSC, 2020). Obesity reduces life expectancy, increases mortality rates, and reduces quality of life (Abdelaal et al., 2017). It raises the risk of chronic conditions such as heart disease, type 2 diabetes, and various cancers, all of which put a strain on medical services; the cost of treating obesity related conditions to the NHS is £6 billion per annum (DHSC, 2022). Typical interventions to reduce the consumption of goods with a combination of negative externalities (that impose costs to society) and negative internalities (that are harmful to the individual) towards a socially optimum level have come in the form of 'sin' taxes. Common examples include alcohol duties and taxes on cigarettes. Yet there is evidence to suggest that 'sin goods' are more heavily consumed by low-income consumers such that sin taxes are highly regressive - see Allcott et al. (2019) for an application to a sugar-sweetened beverage tax. In part for this reason, but also in order to reduce sugar intake even further, policy makers are looking for alternative tools to affect consumption. One such measure proposed by the UK government in 2020 and brought into force in England in October 2022 is legislation to end the placement of unhealthy food and drink in prominent locations within a store, both physically and online.

This paper examines the effectiveness of the policy by observing sales of sugar-sweetened beverages (SSBs) over a two year period in stores from one of the largest supermarket chains in the United Kingdom. While placement restrictions have been used in other products such as cigarettes[1], this ban is the first of its kind in the breadth of products it covers because it applies to all products considered high in fat, sugar and salt (HFSS). Our primary contribution therefore is to estimate the effects on purchases, prices, and

[1] So called tobacco display bans are in effect in several countries including the UK, Australia and Canada. These bans mandate that tobacco products are kept out of sight from customers, often in special cabinets behind counters, and are sold only on request to customers of legal age.

revenues of SSBs as a result of this novel legislation. This work sits in contrast to the vast majority of empirical studies in the domain of reducing SSB consumption, including influential work such as Griffith et al. (2017, 2019) which focus on tax regimes and their welfare effects. As far as we are aware, the UK is among the first countries to implement this type of policy on non-alcohol/tobacco products. Accordingly, understanding the consequences of placement restrictions is critical from a public health perspective, but also to producers, retailers and consumers from a competition perspective. To evaluate the restriction, we utilise a difference-in-difference framework to compare consumption of SSBs affected by the restriction pre- and post-policy, against a control group that was not affected. In actuality we create three different control groups each of which is appropriate under specific assumptions. These assumptions are described in detail in Section 3.3; we do this because we recognise that limitations in our data prevent us from making concrete conclusions using any single one of our control groups.

We limit our analysis to SSBs for two reasons. Firstly, sugar from drinks is considered worse for health than sugar from food (DiMeglio and Mattes, 2000) and so reducing SSB consumption would contribute most to achieving the government's stated aims. Secondly, exactly what constitutes an HFSS product is clearly defined for drinks. Taking its cue from the Soft Drinks Industry Levy (SDIL), the placement restriction only applies to drinks with >5g of *added* sugar per 100ml. Natural fruit juices and smoothies are exempt. For other HFSS products however, the distinction is not so clear; manufacturers and retailers are required to work together to decide what constitutes 'high' in the HFSS definition which is problematic when separating products into those which are 'treated' by the policy and those which are not. In control group 1, we compare SSBs in stores affected by the policy in England with non-SSBs from the same stores. Since non-SSBs make up over 90% of the market by litres sold in the UK (see Table 3.1), control group 1 is the largest control by number of SKUs included (stock-keeping units, see Table 3.2). Yet there is a possibility that consumers switch into purchasing non-SSBs instead of SSBs if the policy causes the price of SSBs to rise or makes it more difficult to find SSBs in the store - akin to a 'reduction' in quality. If this contamination of the policy into sales of non-SSBs occurs, group 1 no longer serves as a suitable control. Group 2 uses SSBs from a different, non-treated store type but also within England as a control. Control group 3 attempts to exploit the (quasi-) natural experiment arising from devolved healthcare policy in Great Britain. Placement restrictions came into effect only in England in October 2022; while discussions have taken place in Scotland and Wales for similar types of action, their governments have not enacted a comparable policy to date. Yet all three mainland British home nations share monetary and fiscal policy as well as similar supply side costs to the supermarket chain, and exhibit common tastes and demographic characteristics. Therefore, we compare SSBs in England with SSBs in Scotland and Wales. The challenge with groups 2 and 3 is the drastic reduction in observations and SKUs available. We expand the discussion on controls in Section 3.3.

In all three instances, our primary dependent variable of interest is the change in the quantity of SSBs, measured in litres, purchased by consumers following the placement restrictions. This is used as a proxy for consumption given that consuming less liquid will result in a lower intake of sugar from drinks. Additionally, as retailers and manufacturers have the option to change prices in response to the policy (which in turn ought to affect quantity), we estimate changes to the price of SSBs. Finally, we estimate the effects on revenues, as any restriction that limits a seller's freedom to operate in a particular way should lead to suboptimal outcomes that reduce profits.

Unsurprisingly, the three different control groups indicate different outcomes from the policy. When using non-SSBs in England as the control, we find that there is no overall effect on quantity sold of SSBs, suggesting the policy had no effect in reducing sugar intake. Groups 2 and 3, using SSBs in petrol stations in England and Scotland & Wales, respectively, as the control suggest a 17-20% decline in

SSB sales as a result of the policy. Here the policy appears successful in its stated aims. This apparent contradiction occurs because of the difference in the types of products sold in different store types. As a rule of thumb, the smaller the store the greater the proportion of drinks that are purchased for immediate consumption. Therefore, we further break down the comparisons by whether a product is for immediate consumption or take-home consumption.[2] When we do so, we find that using control groups 1 and 2, there are statistically significant reductions to the order of 10% and 20% in the quantity sold of immediate consumption SSBs, but no significant effect on the quantity of take-home SSBs.

This point of consumption heterogeneity in effects, supports our expectations about consumer behaviour. We expect the biggest inconvenience to be for immediate consumption drinks which were previously placed in refrigerators in prominent store locations, because consumers are attempting to satisfy an immediate need that is more susceptible to impulse. Although take-home products should also be removed from special displays at the end-of-aisles as a result of the policy, because these products are larger in size and more expensive, they are less likely to be impulse purchases and so those intending to purchase are not as inconvenienced by the policy. As a result we conclude that there is some evidence to suggest the policy is successful in reducing consumption of immediate consumption SSBs. Yet as we show later, these drinks make up only 6-7% of total SSB consumption. Therefore, when it comes to truly reducing the nation's sugar intake, the policy has only a minimal effect. Ultimately, it is targeting the wrong kinds of drinks (immediate consumption), in the wrong kinds of stores (larger stores above 2000 sq ft, which sell much more take-home products). We do of course acknowledge the policy applies to more than just drinks and we cannot comment on its effectiveness in reducing consumption of other HFSS products. For sugary drinks however, we find it is largely ineffective. In order to truly reduce the volume of SSB consumption, any policy ought to (1) target take-home products (2) in larger stores instead.

[2]We proxy for this using product size.

Our work contributes to the literature around 'sin goods' that focuses on policies to reduce consumption, and their effects. Much of this work has centered around alcohol and cigarettes, while taxation has been the traditional policy tool preferred in the first instance by most governments around the world. The first contribution therefore is to the literature targeting sugar consumption and HFSS in general. The second is to the evaluation of non-tax policies in this domain.

Griffith et al. (2017, 2019) study optimal tax design in the alcohol market by estimating a model of demand for the UK. Recognising that products vary in their alcohol content, they show that if consumer preferences are heterogeneous, and correlated with marginal externalities, then varying tax rates across products by ethanol content can improve welfare in comparison to a flat-rate tax. In particular they find that 'heavy drinkers have systematically different patterns of alcohol demands and welfare gains from optimally varying rates are higher the more concentrated externalities are among heavy drinkers' (Griffith et al., 2019). Despite some differences in the market for alcohol and SSBs, these findings are still relevant to the sugary drinks market; indeed the SDIL is a two-tiered tax dependent on the amount of sugar in a drink.

A number of studies look more specifically at SSBs and sodas, including tax regimes but also more broadly at policies to reduce consumption. A public health study by Ejlerskov et al. (2018) utilises a natural experiment whereby six UK supermarkets banned the sale of confectionery at checkouts for various periods between 2013 and 2017 in what might be considered a precursor to our policy of interest. Using a large-scale panel of household purchases of food brought home, they find there is a 17.3% reduction in four-week purchases in stores that applied the ban compared to stores that did not. This reduction in purchases was maintained a year later remaining almost the same, at a 15.5% reduction. The immediate effect but not the later effect was robust to sensitivity analysis. Judging by the effects of this natural experiment, since our policy goes further in limiting sales, we would expect to see even greater reduction in purchases of SSBs. However, the authors note that the data does not reveal where in the store a purchase was made from or indeed how purchases translate into consumption. This is something faced by all studies that use supermarket data to make inferences about consumption.

Following the introduction of the SDIL and other sugar-taxes, including in certain parts of the US, Dubois et al. (2020) ask 'How well targeted are soda taxes?'. Again they use a panel of micro-level data, in contrast to our aggregate level data, to estimate individual-specific preferences on on-the-go drinks (what we refer to as immediate consumption). They find that 'soda taxes are relatively effective at targeting the sugar intake of the young, are less successful at targeting the intake of those with high total dietary sugar, and are unlikely to be strongly regressive especially if consumers benefit from averted internalities' (Dubois et al., 2020). Considering one of the UK government's stated aims is to tackle childhood obesity by targeting sugar intake, soda taxes appear to be a useful policy. However, this study shows that they are not sufficient to target those with high sugar intake. It is for this exact reason that governments seek alternative tools to complement tax regimes. More generally, Griffith et al. (2020) consider policies to reduce young people's sugar consumption. Beyond taxation, they consider advertising restrictions and potential restrictions on the availability of products or changing the characteristics of products (such as moving from sugar to sweetener). However, they do not conduct any empirical work to test the potential effects of these policies, largely because there have been few policies available to analyse. This underscores the importance of the wider HFSS policy as one of the first of its kind globally, providing all stakeholders with an opportunity to empirically understand the effects of these types of policy on purchases, prices and revenues.

Finally, we highlight two studies that have close links to ours. Fearne et al. (2022) use the same database of supermarket aggregate sales data as us (albeit over a different time period) to estimate the impact of the SDIL on buyer behaviour. Using pre-SDIL data they predict the effect of the tax and then test their predictions using actual post-levy data, finding that mean weekly SSB volume falls by 32.9% in the prediction period and 33.6% in the post-SDIL period. They further find that the tax has the greatest impact on low income families who disproportionately consume SSBs. Unfortunately, the 10 CAMEO groups they used are no longer available in the data, replaced by six of the supermarket's own demographic categorisations. As the data is not at the individual level, they face the same data limitations we do, and like us choose to model demand on the aggregate (or composite) SSB good. As we discuss later, this avoids problems caused by a large number of zero observations at the SKU level and is justified because we are interested in the overall quantity of SSBs purchased. As we must also do, Fearne et al. (2022) caveat their results by noting that purchases are not perfectly correlated with consumption for reasons such as stockpiling.

Bokhari et al. (2023) also look at non-tax policies, this time on alcohol, by examining the effects of a volume discount ban, applicable in Scotland but not the rest of the UK, thereby creating a natural experiment. Importantly, they found that the ban increased rather than reduced sales of alcohol because retailers responded by reducing the price of standard, single items so that high consumption households increased the frequency of their store visits and ended up purchasing more alcohol. In other words, the volume discounts acted as a commitment device, in which households would bulk buy to take advantage of the volume discount but not buy again until the previously purchased volume was exhausted. This highlights there are potential unintended consequences from these types of policies. For example, if the absence of SSBs in prominent locations pushes customers with inelastic demand to travel to the part of the store where SSBs are stocked, in search of better value-for-money multipacks they may end up purchasing a greater volume.

The rest of the paper is organised as follows. In Section 3.2 we provide background detail on the location restriction policy, alongside a description of the supermarket chain and the store types used in our sample. This section also outlines the dataset and contains summary statistics for the classes of products that we analyse. Section 3.3 begins with our empirical specification, followed by descriptions of the three control groups we use in our DD model, alongside analysis of pre-policy trends. Section 3.4 presents the results of the baseline aggregate DD models for each control together with analysis of further sub-groups from the data as well as some analysis by lifestage (demographics). Finally, we provide a discussion and concluding remarks in section 3.5.

3.2 Policy Description and Data

3.2.1 Background

Scientific and public discourse surrounding the individual and social effects of sugar intake has been building over the last 10-15 years. In 2015, the World Health Organization (WHO) published an influential set of guidelines based on meta-analysis of randomised controlled trials that recommended a 'reduced intake of free sugar throughout the lifecourse, reducing the intake of free sugars to less than 10% of total energy intake in both adults and children, and a further reduction of free sugars to below 5% of total energy intake' (World Health Organization, 2015). Following a Public Health England (PHE, 2015) report and a Colchero et al. (2016) analysis of an SSB excise tax in Mexico that showed an average purchase decrease of 6% in taxed beverages, the UK government announced the Soft Drinks Industry Levy in the Budget of 2016. In the five years since its introduction in 2018, more than 45,000 tonnes of sugar have been removed from soft drinks in the UK and tax revenues of £334 million were raised in 2021-22 in the process (O'Mara and Vlad, 2023).

Despite the apparent successes of the SDIL, the UK government has reiterated its aim to target sugar intake as a leading cause of obesity (DHSC, 2020), the treatment of which is estimated to cost the NHS £6 billion per annum, rising to nearly £10 billion per annum in 2050 (DHSC, 2022). Wider costs to society through absenteeism, productivity reduction, and accommodations of heavier individuals are estimated to be around £27 billion (PHE, 2017). As part of this, DHSC (2023) introduced legislation in 2020 'to restrict the promotion of HFSS products by volume price (for example, 'buy one get one free') and location, both online and in store in England'. The first of these is due to come into force on 1 October 2025, following two postponements. The location restriction policy however, came into effect on 1 October 2022.

The policy restricts the placement of HFSS products in certain areas of a store, for stores that have relevant floor area of 185.8m^2 (2,000 sq ft) or greater, including any area within 2m of the checkout facility, any area within 2m of a designated queuing area or queue management system, the ends of aisles, store entrances, and covered external areas. Small businesses with less than 50 employees and specialist retailers are exempt from the location restriction policy (though not from the volume price policy). Although both policies apply to a wide range of HFSS products, we focus on SSBs in this study. The definition of an SSB for the purposes of the policy is derived from SDIL definitions which state that any drink with *added* sugar > 5g per 100ml of liquid is subject to the restrictions. In other words, all beverages that are covered by the SDIL are also impacted by the new policies. Drinks that contain natural sugars, dairy products, and alcoholic beverages are exempt.

3.2.2 Supermarket chain

Our data is taken from one of the UK's largest grocery retailers. As seen in the dataset, this supermarket chain operates four types of physical store. Petrol stations refer to those attached to fuel stop forecourts and are the smallest of the stores included in the dataset. 88% of all UK petrol station forecourt shops are less than 2,000 square feet in size (ACS, 2022). The remaining 12% are all less than 3,000 square feet. Express stores are next in size and are usually urban or inner-city convenience locations. Originally, Express stores had an upper limit in size of 3,000 square feet. However, since the retirement of the Metro brand which operated stores between 7,000 and 15,000 square feet, some smaller Metro stores were renamed Express. As such, there is a wide variety in size of Express stores. Express stores are also the most common type of store, with 1,713 locations in 2016 (Vasquez-Nicholson, 2016). The two remaining store types, Superstore and Extra, are the largest stores that typically contain a full range of products and are usually located in more out-of-town or retail park locations. Of these, Extra stores are usually the largest[3] and are so called because they include extra services such as opticians, dry cleaners and pharmacies. However, the size delineation is not clear as some Superstores are larger than certain Extra stores. We find that the pricing policies between Extra and Superstores are largely the same and so for some of the regression analysis that follows, we combine the two store types into a single group named 'large' stores.

3.2.3 Data

Our primary data is taken from a proprietary database from one of the UK's largest supermarkets, which contains weekly sales of stock-keeping units (SKUs) by home nation, store type, and lifestage - an internal demographic classification - from a random 20% of its UK loyalty program customers. Loyalty program transactions represent approximately 80% of the supermarket's sales (West, 2023). The raw sample contains observations from 113 weeks beginning 3rd May 2021 to 2nd July 2023 on over 2,500 SKUs. For each SKU, we have information on the number of units in a pack, the size of each unit, the number of units/packs sold and the revenue generated. From this information we are able to calculate quantity in litres and price per litre as revenue divided by quantity. We supplement this data with information on the nutritional content of the drinks including energy, protein, sugar per 100ml, and a dummy for whether the drink contains any artificial sweeteners.[4] For

[3]Some have a size of over 100,000 square feet or 9300 m^2 (Vasquez-Nicholson, 2016).

[4]Although in some cases the names of the particular sweeteners are listed on the product, in most instances the amount of sweetener used is not specified.

estimation, we log transform each of our three dependent variables, quantity, price and revenue. This allows us to make comparisons between groups that have sales that differ by orders of magnitude. Moreover, we can directly interpret our coefficient estimates as percentage changes.

The placement restriction was implemented on 1 October 2022, which corresponds to week 75. Many of the SKUs exhibit near zero sales for the entire sample period; due to the construction of the database, products that are discontinued from shelves are not immediately removed from the dataset. Additionally, products in this segment exhibit frequent churn; new products are often added and others discontinued. To avoid churn driven by factors unrelated to the policy and to minimise the problems caused by the data collection methodology, we limit our analysis to a stable set of products that are available to purchase for 22 weeks either side of the policy introduction. We refer to this 44 week period from May 2022 to February 2023 as the policy window. To account for seasonality, we include the equivalent 44 week period from the previous year, May 2021 to February 2022, referred to as the pre-policy year. The final set of SKUs in the analysis appear in every week for both the pre-policy year and the policy window. We also discard SKUs that account for the bottom 5% by cumulative sales. This restriction removes the majority of SKUs with negligible sales; the policy is not targeting these products so it is reasonable to exclude them from our analysis.

Table 3.1 presents the percentage of sales by store type and group across the UK. We can see that the largest stores, for whom the restriction definitely applies, account for over 80% of sales by volume for all drinks and a similar figure for non-SSBs. However, when we look at SSBs sales - row *SSBs/Total* - they constitute a greater proportion of sales the smaller the store. The distinction between immediate consumption and take-home products has an element of fuzziness to it; we define immediate consumption drinks as those sold in single units usually in 330ml cans or 500ml bottles (but up to 750ml). These products are typically found in refrigerated cabinets placed at many of the prominent locations mentioned in the policy. Table 3.2 shows the number of

Table 3.1: Percentage of litres sold by category and store type in Great Britain

	Small Stores may be $< 2,000$ sqft		Large Stores $> 2,000$ sqft		
	Petrol	Express	Extra	Super	Total
England					
Total Sales	0.34	16.60	42.61	40.79	100
SSBs	0.10	1.99	2.52	2.58	7.09
Non-SSBs	0.24	14.61	40.09	38.21	92.91
SSBs/Total	29.41	12.00	5.91	6.33	7.09
Immediate	0.23	3.42	1.28	1.66	6.36
Take home	0.11	13.18	41.33	39.13	93.64
Immediate/Total	67.65	20.60	3.00	4.07	6.36
Scotland/Wales					
Total Sales	0.51	8.34	43.52	47.63	100
SSBs	0.16	1.14	3.34	4.01	8.64
Non-SSBs	0.35	7.20	40.18	43.62	91.36
SSBs/Total	30.84	13.64	7.67	8.41	8.64
Immediate	0.37	2.05	1.88	2.65	6.95
Take home	0.14	6.28	41.64	44.98	93.64
Immediate/Total	71.90	24.64	4.32	5.56	6.95

All values expressed as percentages. SSBs/Total measures the proportion of all drinks in a particular store-type that are SSBs.

SKUs in our final sample. Currently, SSBs constitute only 6-7% of SKUs which correlates with the fact they only account for around 7% of sales from Table 3.1.

Table 3.2: Number of SKUs in sample

	England	Scotland	Britain
SSBs	28	32	33
Non-SSBs	379	393	434
Immediate Consumption	64	69	77
Take home	343	356	390
Total	407	425	467

Bottom 5% sales removed, product set stabilised through analysis periods

3.3 Methodology

3.3.1 Empirical Specification

We use a difference-in-differences (DD) approach to compare quantity, price and revenue associated with SSBs before and after the policy in a panel data framework, observing weekly sales for each SKU for a total of 88 weeks comprising the pre-policy year and the policy window. Let y_{it} represent either log quantity, log price per litre, or log revenue in week t. For the majority of this paper, where we aggregate SKUs into treated and control groups, i refers to one of these two groups. In a SKU level analysis, i would refer to individual SKUs. The DD model is given by

$$y_{it} = x'_{it}\beta + u_{it} = \beta_1 D_i + \beta_2 B_{it} + \beta_3(D_i B_{it}) + x'_{4it}\beta_4 + u_{it} \tag{3.3.1}$$

where D_i is an indicator variable equal to 1 if the group (or SKU) is treated and 0 otherwise. Likewise, B_{it} is a dummy variable set to 1 in the post-policy period, 0 otherwise. The coefficient of interest is β_3 on the interaction term $D_i B_{it}$ which represents which captures the causal effect of the policy on treated SSBs, relative to the counterfactual in which those products had not been treated in the post-policy period.. The vector x'_{4it} includes additional covariates such as seasonal dummies, year dummies, interactions between seasonal dummies, and D_i and indicators for 'immediate consumption' versus 'take home'.

3.3.2 Assumptions

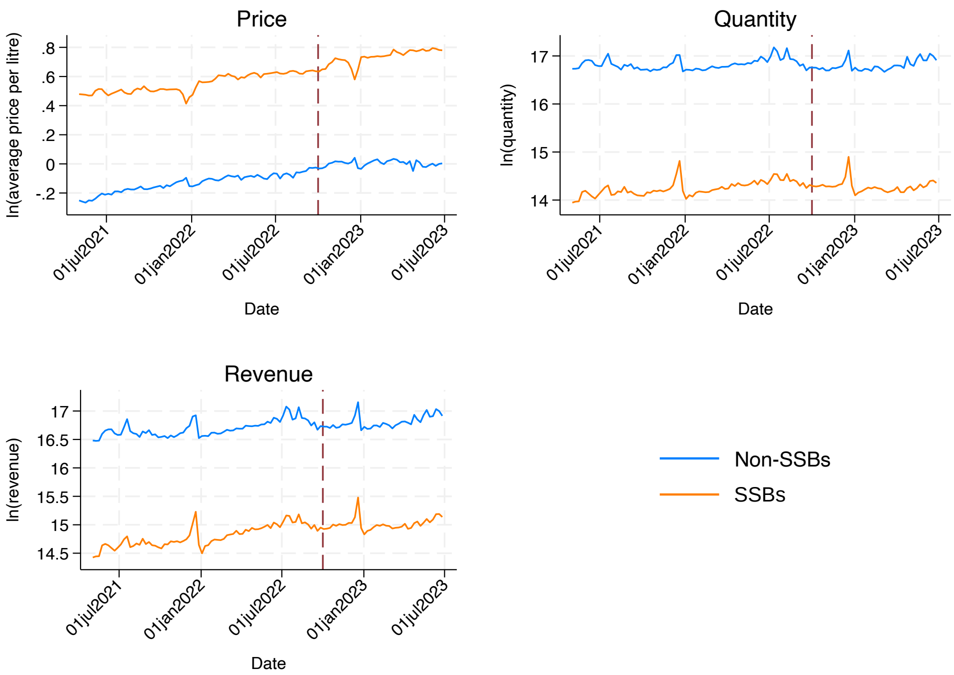
The key assumption of the canonical DD specification is that common trends exist between the treatment group and the control group in the pre-policy period such that in the absence of the treatment, the treated group would have continued to trend parallel to the control. While some studies include several pre-trends tests such as event studies to illustrate the suitability of a control, these come with caveats highlighted by recent literature (see Kahn-Lang and Lang (2020), Bilinski and Hatfield (2018), and Roth (2022)). For example 'pre-trend tests can have low power and conditioning the analysis on the results of pre-trends tests can induce distortion to estimation and inference from pre-testing' (Roth, 2022). Similarly, recent papers have even proposed relaxations of the parallel trends assumption (see, e.g., Rambachan and Roth (2019); Freyaldenhoven et al. (2019)). The literature on innovations in DD methods is beyond the scope of this paper (see Roth et al. (2023) for a review). Like most papers that use a DD model we present graphical evidence to illustrate the presence - or not - of common trends, for each control group in the subsections that follow.

We make two further assumptions in our initial estimations. First, we assume there is no policy anticipation—that is, the policy has no effect prior to implementation. Anticipation confounds the pre/post differential because the pre/post delineation is fuzzy. We justify this by arguing that for revenue generating drinks, there is little incentive to alter characteristics of the product, including formula or price, before the legislation applies. Second, we assume binary and static treatment effects: all treated products are affected simultaneously, and that products are either treated or not; there are no differential effects between different treated products.

3.3.3 Control Group 1

The first control group we use consists of non-SSBs in England as a comparison to SSBs in England. Non-SSBs are not restricted to within-store locations by the policy. To assess the suitability of this control group we look at pre-trends for aggregated quantity, price and revenue for the entire raw sample period, for all stores excluding petrol stations. We discuss the issues surrounding petrol stations in the next subsection.

Figure 3.1: Pre-trends between SSBs and non-SSBs in England in stores exc. petrol (Control 1)



Note: The dashed red line indicates the date the legislation came into effect

From the graphs in Figure 3.1, it appears that the blue line, representing non-SSBs and the orange line, representing SSBs follow a similar pattern in each of our three variables of interest, especially during the period 1st January 2022 up to the date the policy came into effect. Looking at only this period, we can see there is a steady and parallel increase in both these lines in the price graph, a small increase to summer 2022 before a decline to October in both lines in the quantity graph and a similar pattern in the revenue graph. This suggests that using non-SSBs in England could be an appropriate control group. However, there is a possibility of contamination that we must consider. If the policy causes consumers to switch out of SSBs then they must either be reducing their overall consumption of drinks, switching into non-SSBs or switching into other types of drink such as tap water, tea or coffee. In the case of the second scenario, our control group would be affected by the policy, violating the requirements for DD inference. This is a serious problem and is part of the reason that we consider alternative control groups below. We also estimate two subspecifications using this control group as a result of the minimum store size required for the policy to apply. Extra and Super-stores definitely exceed the 2,000 sq ft requirement and so are definitely subject to the policy. Unfortunately, we have no way of verifying the size of each Express store in our sample, so some may well be under the 2,000 sq ft lower limit. Since Express stores account for over a quarter of all SSB sales in England (from Table 3.1) we cannot simply ignore them. Thus we separate our initial analysis using this control group to (1) Large stores only and (2) All stores excluding petrol stations. Later we test whether contamination occurs by estimating an almost-ideal demand system model (AIDS) to obtain cross-price elasticities between SSBs and non-SSBs. Regardless of this result, the following two controls can act as robustness checks.

3.3.4 Control Group 2

In this group we compare SSBs in large stores in England with SSBs in petrol stations in England as the control. Although we do not know the exact breakdown of the supermarket petrol stations that are above and below the floor space cut-off for the placement restrictions to apply, we present two possible scenarios that enable us to utilise petrol stations as a control.

First, given the breakdown of petrol station sizes nationally, we can assume the vast majority of the supermarket's petrol stations fall below the 2,000 square

Figure 3.2: Pre-trends between SSBs in Extra stores and SSBs in petrol stations in England (Control 2)

Note: The dashed red line indicates the date the legislation came into effect

feet minimum size. Therefore, in the supermarket's petrol stations, SSBs are not affected by the policy. Second, there is a possibility that *some* petrol stations are large enough such that the policy is applicable to these stores. In this instance, it may be that instead of individually mandating which petrol stations do and which do not have to abide by the policy, the supermarket issues a blanket notice for all petrol stations to adhere to the policy. But given the size of the stores, even when products are moved away from prominent locations, they remain within easy reach of consumers. The policy aims to inconvenience consumers away from SSB purchases but this effect will be minimal to be almost negligible in the petrol station stores. Since the products themselves are identical to the SSBs of the same SKU sold in large stores and the supermarket is likely to have single contracts with manufacturers for all their store types such that costs are relatively constant between store types, we believe this is an appropriate control group. The problem now is whether there are sufficient common pre-trends between SSBs in petrol stations and SSBs in larger stores to use the former as a control. The graphs in Figure 3.2 are far less convincing than those from Figure 3.1.

Figure 3.3: Pre-trends between immediate consumption SSBs in large stores and in petrol stations in England (Control 2)

Note: The dashed red line indicates the date the legislation came into effect

Part of this is likely caused by the different consumption types that occur in each store type. From Table 3.1 we can see that drinks for immediate consumption comprise around two-thirds of all beverage sales in petrol stations but only 3% of beverage sales in Extra stores. Therefore we compare immediate consumption products in large stores with those in petrol stations in Figure 3.3. The first thing to notice is that price now clearly moves in a similar fashion in both store types which is unsurprising given, as mentioned earlier, that the products are identical and contracts will have been negotiated centrally at the supermarket. Quantity and revenue appear to have a sharper increase in petrol stations in the months immediately following January 2022. However, from around March/April 2022 up until the policy implementation, the blue and orange lines do appear parallel.

3.3.5 Control Group 3

With this control group we attempt to utilise the quasi-natural experiment that arises from the devolution of certain public health policies around the UK. England is the only home nation in which the placement restriction is legislation. It does not apply to Scotland, Wales or Northern Ireland. However, since the countries are very similar in many ways including tax regimes, monetary policy, costs to the supermarket, culture, demographics, and tastes, we can use the other home nations as a control. We focus on Scotland and Wales primarily as they, together with England, constitute mainland Great Britain and so have greater similarities, especially from a supply side perspective, with England than Northern Ireland. As they are much smaller in population and sales, we combine sales in Scotland and Wales. However, it appears that the supermarket has an internal price parity policy such that in a given week the price of a SKU is the same within the same store type across Great Britain. For example, if you were to purchase a 6*330ml pack of Coca-Cola cans from an Express store, the price should be the same in London, Edinburgh and Cardiff. We test for this price parity in Figure 3.4 and Table 3.3. The top row of panels show prices in England in a given week in a given store type against prices in Scotland and Wales in the same week and store type for the same SKU. The bottom row of panels compares prices in England with Northern Ireland. We also split the figures into before (the left hand column) and after (the right hand column) the policy. The black line in each figure represents the 45 degree line - points on the line are exactly the same price in both countries. Two tolerance bands are also shown: +/- 5% in pink and +/- 10% in grey. This is because we might observe differences in average prices if the take-up of promotional offers is different. The shape of the scatter graphs suggest price parity does indeed occur and there is little difference before and

Figure 3.4: Is there price parity across the UK?

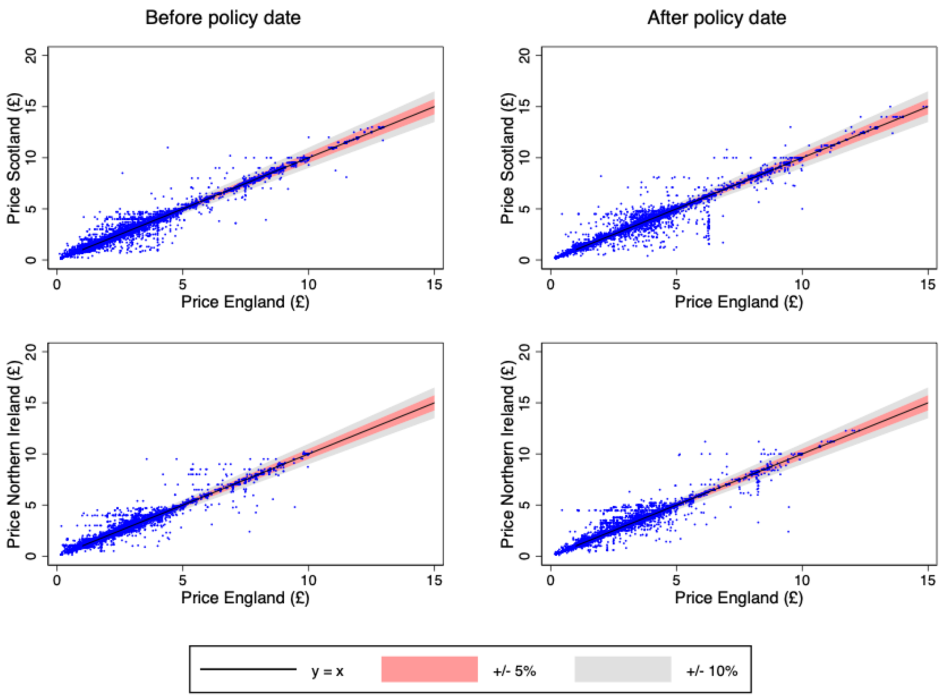
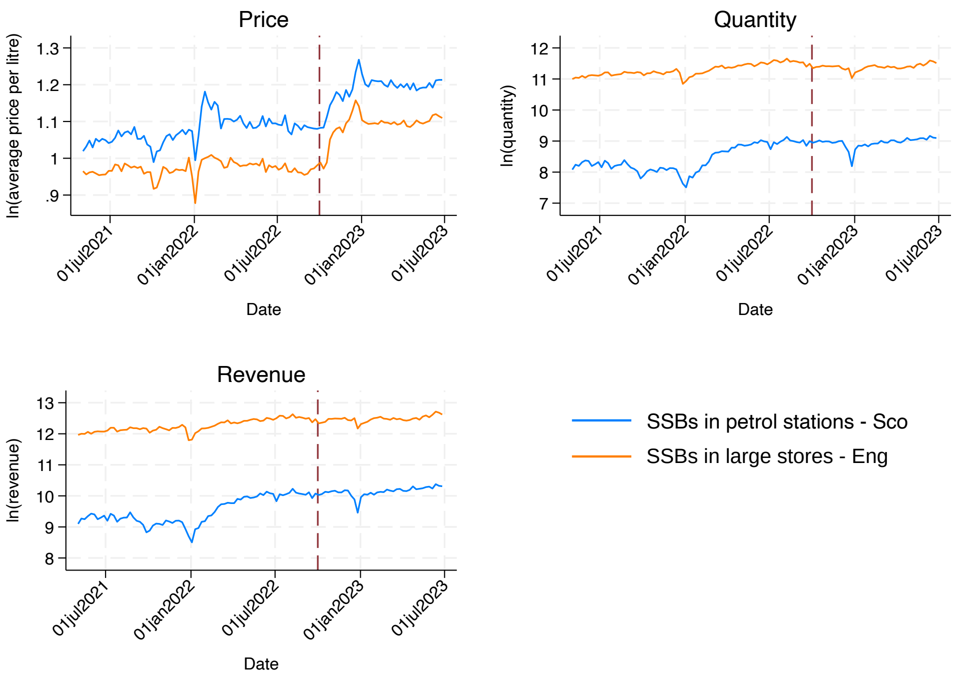


Table 3.3: Percentage of prices within given range of England

	Scotland & Wales	Northern Ireland
Before policy		
Exactly equal	52.6%	51.4%
Within 5%	93.9%	87.0%
Within 10%	97.7%	94.7%
After policy		
Exactly equal	51.2%	45.2%
Within 5%	94.1%	85.9%
Within 10%	97.5%	93.4%

after policy for the majority of SKUs in this regard. Indeed when we examine the numbers in Table 3.3 in Scotland and Wales, around 50% of SKU/week observations are exactly the same price as in England. Including a 5% tolerance level, this figure increases to 94.1%. However, price parity introduces a risk of spillover effects that complicate cross-country comparisons. If firms respond to the policy by for example, reducing prices of SSBs in England then since England represents 85% of the UK market, to achieve price parity in Scotland and Wales, the supermarket would have to reduce prices in these two nations also. We would then expect to see an increase in demand in Scotland and Wales as a result of contamination from a policy which does not apply there.

Figure 3.5: Pre-trends between immediate consumption SSBs in large stores in England and in petrol stations in Scotland & Wales (Control 3)



Note: The dashed red line indicates the date the legislation came into effect

As the number of drinks that are classified as SSBs is already low, it is not possible to observe a control group in large Scottish stores without overlap in products that is necessary to address the price parity spillover effect. However, we can compare changes to SSBs in large stores in England with SSBs in petrol stations in Scotland. The same assumptions regarding petrol store size apply as in Control Group 2. Now, we also assume that there is no price parity across store types across countries. Nevertheless, as SKUs are identical whether they are sold in petrol stations or elsewhere we would expect a high degree of correlation in prices to exist. Figure 3.5 compares large stores in England with petrol stations in Scotland and Wales, for immediate consumption goods only. Figure 3.A.1 in the Appendix shows the same for all goods. We see the price graph is very similar to the graph from 3.3, with quantity and revenue in Scotland/Wales petrol stations following a similar pattern although less pronounced than their English counterparts used in Control Group 2.

3.3.6 Other considerations

The long panel nature of the data in aggregate form is such that $T > N$; we must account for serial correlation in the error u_{it}. Starting with a two-way effects model $y_{it} = \alpha_i + \gamma_t + \mathbf{x}'_{it}\beta + \varepsilon_{it}$ since our panel now has so few individuals, the individual effects α_i can be incorporated into $\mathbf{x}_{it}$ as dummy variable regressors while time effects can be included with a linear time trend. Intuitively, it is easy to imagine that both price and sales in week t are correlated with the corresponding values from $t-1$. Formally, at the greatest level of generality $u_{it} = \rho_i u_{i,t-1} + \varepsilon_{it}$ where ε_{it} are serially uncorrelated but are correlated over i with $\text{Cor}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_{ts}$. The results presented in the following sections utilise a flexible pooled feasible GLS model with errors correlated across each group and a distinct auto-regression order 1 process for each group.

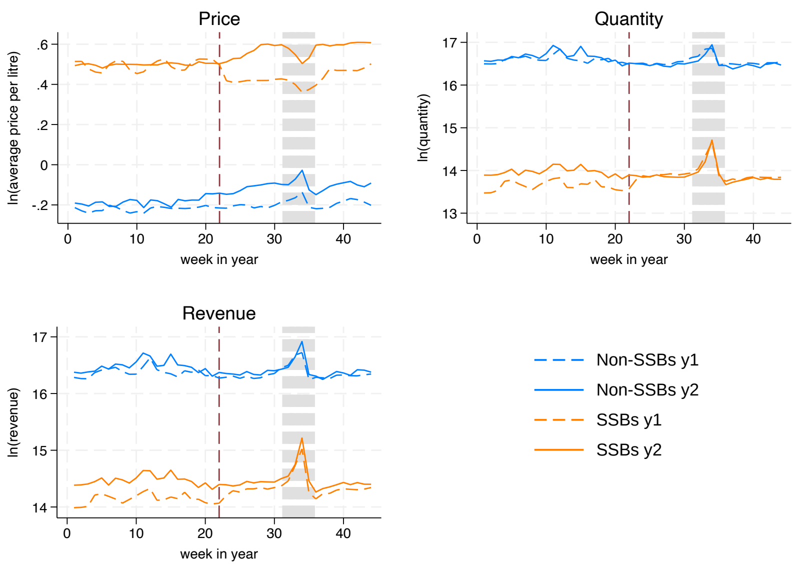
Additionally, grocery shopping has cyclical elements and so we also include some time and seasonality variables in our specifications to account for these effects. Given that we have data from two equivalents periods exactly one year apart, we add 43 week dummies to account for seasonality during the 44-week 'windows'. An additional 'year' term accounts for linear annual trends. Finally, we concentrate our analysis at the aggregate level, separating our sample into two groups for each set of controls; a treatment group that contains SSBs subject to the policy and the respective control that contains either SSBs or non-SSBs that are not subject to the policy. This is similar to Fearne et al. (2022) who use what they call a composite good rather than the demand for individual SKUs as the dependent variable. Our primary justification is that we are interested in the total demand for SSBs as a result of the policy, not the impact on individual drinks and therefore the aggregate measure is sufficient. Secondly, although we attempt to create a stable set of products across the sample period, this is far from an exact process and aggregation to some extent side-steps this issue. Further, there are many weeks in which an individual SKU has 0 sales, especially when we divide sales by country, store type, lifestage or some combination of the three. As such individual level data has a strong right skew with a large number of zero observations, requiring solutions to the log-zero problem such as using a hyperbolic sine function or 2-step model in which demand in the second stage is conditional on a positive probability of purchase in the first stage. However, we are not satisfied that either of these represent reality better than the aggregate log model we employ.

3.4 Results

3.4.1 Descriptive Analysis

We begin with a series of figures illustrating yearly comparisons between the treated and untreated groups for each set of controls having 'cleaned' the data. Figure 3.6 shows the difference in trends between SSBs (treated) and non-SSBs (non-treated) in English large stores only. The solid lines represent observations from the policy year (year 2) while the dashed lines are observations from the pre-policy year (year 1). When it comes to price, certainly there are some year effects indicated by the difference in the solid and dashed lines of the same colours. This justifies our inclusion of a year dummy. Comparing the solid lines in the period before the policy date (denoted by the dashed vertical red line) we can see that in all three dependent variables they follow a similar pattern. The graphs suggest that in the post-policy period in year 2, there was little effect on quantity and revenue but perhaps a slight increase in the price of SSBs compared to the counterfactual where they are not treated. The figure for all stores excluding petrol stations is very similar and so is omitted here in the interests of space. We include it in the appendix for posterity (see Figure 3.A.2).

Figure 3.6: Yearly comparisons between SSBs in English large stores only



Figures 3.7 and 3.8 show the same figures but compare SSBs in large stores in England with Petrol Stations in England and Petrol Stations in Scotland respectively. We include graphs that compare sales of immediate consumption goods only here but include the figures for all goods in the appendix as Figures 3.A.3 and 3.A.4. The reason for this is that the majority of sales (and SKU overlap) in petrol stations are for these immediate consumption products. In both Figure 3.7 and Figure 3.8 the orange lines are identical because the treated group is the same in both. However, the blue lines are also very similar indicating that petrol stations across mainland Great Britain exhibit similar characteristics. We know there is price parity so the shape of the price lines is unsurprising, but quantity and revenue also follow each other very closely. The pre-policy period is flat in quantity and revenue in both figures. Price contains much more movement and it is difficult to say that the treated and untreated products move in parallel. Looking at the post-policy period, price does appear to move parallel in both figures suggesting no effect on price when using this

control group to estimate the counterfactual. Quantity and revenue also both appear parallel in the post-policy period, or perhaps a slight decline in the treated group.

Figure 3.7: Yearly comparisons between SSBs in English Extra stores and Petrol Stations - immediate

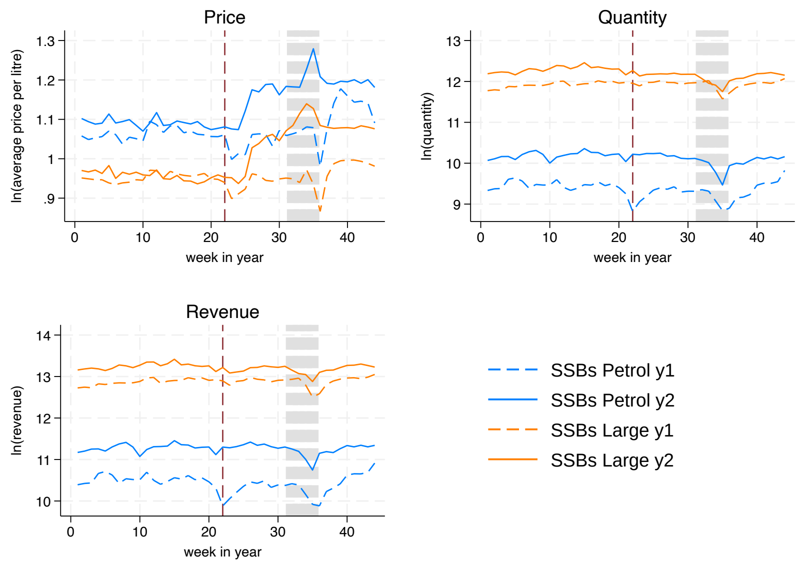
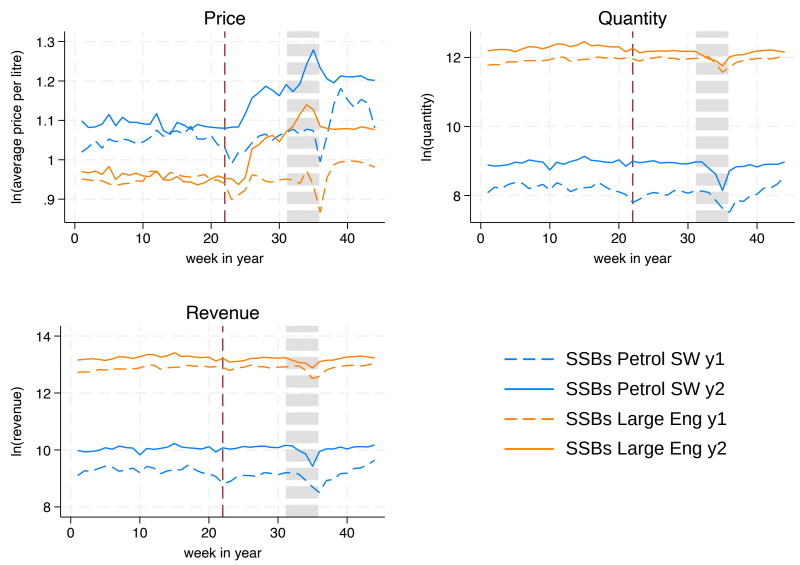


Figure 3.8: Yearly comparisons between SSBs in English Extra stores and Scottish/Welsh Petrol Stations - immediate



3.4.2 Control Group 1

In Table 3.4 we present the results using non-SSBs in England as our control group for large stores only (columns 1-3) and all stores excluding petrol stations (columns 4-6). The row labelled *D*B* is the coefficient on the interaction between the before/after dummy and the treated/non-treated dummy and is our variable of interest. The results are similar for both types of store; there is an estimated increase in prices of 5-6% for SSBs had they not been treated and this effect is highly significant in both cases. The results however, indicate there is no significant difference in quantity or revenue as a result of the policy. Of course, we have previously highlighted the potential difference in effects between beverages for immediate consumption and those consumers take home. Table 3.5 illustrates these differences for beverages in large stores only. The difference in results between this table and the same for all stores excluding petrol stations is negligible so we do not include the latter here, although we include them in Table 3.A.1 in the Appendix. The results show that there is a negative and statistically significant change in both quantity and revenue for immediate consumption SSBs in columns 1 and 3, of 10% and 5.3% respectively. Together with a significant increase in price of 3.3% in column 2, this suggests the policy had the desired impact in reducing consumption for immediate consumption SSBs. For consumers attempting to purchase immediate consumption drinks, the absence of SSBs from prominent refrigerated units appears enough to reduce the number of purchases. The same cannot be said for take-home products, as there are no significant changes in quantity (column 4) or revenue (column 6), although there is a 4% increase in price in column 5. If consumers are traversing the store anyway as part of their weekly shopping, they are still likely to arrive at the relevant aisle and purchase take-home beverages even if these products are absent from end-of-aisle or store-front displays. From Table 3.1 we can see that take-home beverages represent around 95% of all sales in large stores. Thus, although the policy appears to work for immediate consumption drinks, this is only a small fraction of the sugary drinks volume sold and is not targeting the real source of high consumption i.e. from take-home products.

Table 3.4: Control 1: Fully aggregated results

	Large stores only			All stores exc. petrol		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.186***	0.058***	-0.136***	-0.193***	0.049***	-0.159***
	(0.037)	(0.007)	(0.033)	(0.046)	(0.009)	(0.042)
D: Treated	-2.821***	0.695***	-2.130***	-2.675***	0.736***	-1.945***
	(0.053)	(0.011)	(0.048)	(0.063)	(0.013)	(0.059)
*D*B*	-0.006	0.050***	0.050	0.021	0.058***	0.081
	(0.046)	(0.009)	(0.041)	(0.059)	(0.010)	(0.055)
N	176	176	176	176	176	176
*D*wk* χ^2	175.16***	343.21***	102.97***	118.91***	312.11***	61.08**

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes 43 week seasonal dummies, 43 *D*week* interactions and year dummy

3.4.3 Control Groups 2 & 3

In Table 3.6 we present the fully aggregated model for control 2: SSBs in large stores versus SSBs in petrol stations all in England, in columns 1-3 and control 3: SSBs in large stores in England versus SSBs in petrol stations in Scotland & Wales, in columns 4-6. These results are markedly different from the aggregated results in Table 3.4. In both control 2 and 3, we see that quantity and revenue are significantly lower than the counterfactual scenario in which SSBs in large stores in England are not treated by the policy. For control group 2, quantity sold is 20.4% lower and revenue is 23.7% lower while using control group 3, quantity and revenue are 17% lower and 12.6% lower respectively. The magnitude of coefficients is certainly smaller using control group 3, but these results are consistent in their

Table 3.5: Control 1: Immediate consumption versus take-home products in large stores

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.037	0.028***	-0.006	-0.204***	0.055***	-0.159***
	(0.023)	(0.009)	(0.020)	(0.038)	(0.007)	(0.035)
D: Treated	-0.835***	0.136***	-0.703***	-2.997***	0.603***	-2.400***
	(0.042)	(0.014)	(0.040)	(0.054)	(0.011)	(0.051)
*D*B*	-0.100***	0.033***	-0.053**	-0.037	0.041***	0.014
	(0.029)	(0.013)	(0.026)	(0.045)	(0.009)	(0.041)
N	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes 43 week seasonal dummies, 43 D*week interactions and year dummy

presentation across both controls. Price is the outlier here in the sense that it appears flat in control 2 in column 2 but has a small positive and significant effect of 3.9% in control 3 from column 5. Looking at the results in this table, the policy could again be considered a success; both quantities and revenues of SSBs treated by the policy are down.

Shifting focus to immediate consumption and take-home goods, in Table 3.7, for control group 2, the pattern in columns 1-3 is similar to columns 1-3 in Table 3.6 although the magnitude of the effect of the policy is slightly greater at nearly 30%. Considering the product range in petrol stations is more skewed towards immediate consumption drinks this is unsurprising. Again, for these immediate consumption drinks, the policy appears effective. However, looking at take-home products there is an increase in both quantity and revenue as a result of this policy by over 30%. Given that take-home products account for a much larger proportion of sales, if these results are to be believed the net result of the policy would be an

Table 3.6: Control 2 & 3: Fully aggregated results

	Control 2			Control 3		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.027	0.070***	0.040	0.016	0.010	0.035
	(0.074)	(0.012)	(0.074)	(0.040)	(0.014)	(0.042)
D: Treated	3.919***	-0.564***	3.353***	7.042***	-0.342***	6.716***
	(0.109)	(0.016)	(0.113)	(0.092)	(0.034)	(0.085)
*D*B*	-0.204**	-0.010	-0.237**	-0.170***	0.039*	-0.126***
	(0.101)	(0.013)	(0.103)	(0.046)	(0.022)	(0.047)
N	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes 43 week seasonal dummies, 43 D*week interactions and year dummy

increase in the total quantity sold of SSBs. However, we acknowledge that using petrol stations to judge sales of take-home drinks is probably not structurally appropriate so we pay less attention to the results in columns 4-6 than those in 1-3.

Interestingly, when we look at immediate consumption goods using petrol stations in Scotland and Wales as the control in Table 3.8, there is no significant effect of the policy on quantity and revenue but a small, positive and significant effect of 4% on price. This is quite different from columns 1-3 in Table 3.7, leaving the question of which set of results to believe. The results for take-home products are also quite different, with negative but insignificant coefficients on quantity and revenue, and a negative and significant effect of 10% on price in Table 3.8, compared to positive and significant effects in quantity and revenue with a small, insignificant effect on price in Table 3.7. Again, the caveat regarding the structural appropriateness of this particular control group for this product type holds.

Table 3.7: Control 2: Immediate consumption versus take-home products

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	0.011	0.070***	0.072	-0.143	0.012	-0.127
	(0.063)	(0.012)	(0.065)	(0.096)	(0.014)	(0.102)
D: Treated	2.245***	-0.118***	2.122***	6.514***	0.200***	6.695***
	(0.092)	(0.017)	(0.098)	(0.189)	(0.036)	(0.196)
*D*B*	-0.291***	-0.010	-0.291***	0.317*	0.023	0.324*
	(0.086)	(0.016)	(0.091)	(0.167)	(0.026)	(0.174)
N	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Control 3: Immediate consumption versus take-home products

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.147***	0.051***	-0.102**	0.346***	0.138***	0.486***
	(0.048)	(0.014)	(0.050)	(0.091)	(0.011)	(0.095)
D: Treated	5.808***	-0.134***	5.670***	7.796***	0.122***	7.922***
	(0.085)	(0.026)	(0.089)	(0.283)	(0.040)	(0.284)
*D*B*	-0.020	0.041*	0.043	-0.156	-0.105***	-0.263
	(0.058)	(0.022)	(0.060)	(0.237)	(0.027)	(0.238)
N	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Given that each set of results leads us to make slightly different conclusions, which set of results should be considered most reliable? Or rather, under what conditions do we believe each one? The success of non-SSBs in England as an appropriate control hinges on the substitutability of SSBs and non-SSBs. If they are close substitutes then a policy restricting the 'quality' of SSBs will see customers switch into non-SSBs. The policy would then have an impact on our control and confound our DD estimates. To test this substitutability we estimated a simple 2 product almost-ideal demand system (AIDS) model, the full results of which are presented in Table 3.A.2 of the Appendix. The uncompensated cross-price elasticities we obtain between SSBs and non-SSBs are both non-significant which suggests there is little substitutability between SSBs and non-SSBs. In other words consumers are not switching into non-SSBs as a result of price increases in SSBs. This could be purely to do with consumer preferences, but also if consumers have different concerns regarding the negative effects of sugar versus potential negative effects of artificial sweeteners. Given the push from governments to switch from added sugar to added artificial sweeteners there is little research into the negative effects of sweeteners and we are unlikely to see any long-term health implication for several years. With this being the case, we are more confident in the results from control group 1. This group then becomes our preferred control. Petrol stations and large stores are too different, in our opinion, in the types of drinks they primarily stock and the shopping trip purposes to compare effectively even though the analysis supports the conclusion that the policy does reduce consumption of immediate consumption drinks over take home drinks.

3.4.4 Life-stages

Instead of the CAMEO group[5] classifications previously present in the data, the database now categorises consumers into life-stages based on information given when they signed up for a supermarket loyalty card, and their shopping habits. This is the only demographic information available in the data. There are six life-stages, listed below. We are able to see a particular shopper's household

[5]CAMEO is a proprietary consumer classification system that indicates the socio-economic and geo-demographic profiles, created by TransUnion Information Group.

situation but We can observe a shopper's household situation but cannot infer key demographic factors such as household income, geographic location, or exact household size. Nevertheless, applying our DD model to each life-stage subgroup might be useful if consumption of SSBs was different among each of these groups; we could then observe whether there are differential effects on each life-stage from the policy and whether it is targeting those life-stages who have the largest SSB consumption.

Table 3.9: Life-stage summary statistics

	Older Adults	Older Families	Pen-sioners	Young Adults	Young Families	Aggregate
Customers* (000's)	1,836	1,085	777.7	926.8	954.5	5,613
Penetration† (%)	49.5	57.6	39.37	51.03	61.7	
Group size‡ (000's)	3,889	1,838	1,942	1,600	1,265	10,534

A 6th category 'Unclassified' contains negligible customers so we exclude it from this table. All figures are weekly measures. *This is the mean number of unique customers who bought at least one drink in our product set per week. †This is the mean percentage of customers from a group that bought a drink. ‡From the previous two figures we estimate group size.

Table 3.10 shows aggregate results for control group 1, separated into life-stages for large stores only. Broadly speaking, these results corroborate what we saw in Table 3.4. In all lifestages, we see a statistically significant and positive change in the price of SSBs versus the counterfactual of between 2.3% and 8.1% compared to 5% in column 2 of Table 3.4. Both older groups - Older Adults and Older Families - also see a statistically significant and positive change in revenues generated for the supermarket by 9.1% and 6% respectively. All coefficients on quantity and the remaining coefficients on revenue are small in magnitude and crucially not statistically significant. The results suggest that no one life-stage group is affected more than the others.

Next we separate immediate consumption and take-home drinks for each life-stage, given that regardless of the control group, we saw differential effects of the policy depending on the point of consumption of a drink. In Table 3.11, we present the results for Older Families in Panel A and Young Families in Panel B, because these two life-stages demonstrate the largest effects in the interaction term versus the counterfactual. We present the results for the remainder of the life-stages in the appendix, foregoing the Unclassified group. The effects of the policy on immediate consumption drinks from Table 3.5 are most evident for the 'Family' life-stages. Both groups see a relatively large decrease in quantity (13.4% for Older Families and 24.1% for Young Families), an increase in price around 5% for both, and a decrease in revenue (4.8% for Older Families and 16.4% for Young Families), all of which are statistically significant. Given one of the governments aims is to target childhood obesity, the effects on immediate consumption for Families are a positive result. Indeed, the policy is more impactful for Young Families suggesting it is effectively targeting the youngest children. In comparison, for 'Adult' life-stages in Table 3.A.4, we see that there is no significant change in either quantity, price or revenue for immediate consumption drinks, perhaps suggesting that for single people the inconvenience imposed by the policy is not sufficient to change their consumption habits, whereas for busy families, particularly those with very young children, convenience of easy to reach non-SSBs is a much more salient factor in their purchase decision. Based on the assumption that consumer preferences are heterogeneous it would be useful to understand more about the relative sizes, characteristics and consumption behaviour of each group to make informed conclusions about differential impacts, particularly with regards to welfare, on each group. A more detailed, individual level data set would be helpful in this regard.

Table 3.10: Control 1: Large stores, aggregate results by life-stage

	Older Adults			Older Families			Pensioners		
	(1) lnq	(2) lnprice	(3) lnrev	(4) lnq	(5) lnprice	(6) lnrev	(7) lnq	(8) lnprice	(9) lnrev
B: After	-0.210***	0.051***	-0.168***	-0.206***	0.047***	-0.158***	-0.213***	0.059***	-0.135***
	(0.042)	(0.010)	(0.035)	(0.030)	(0.007)	(0.025)	(0.029)	(0.010)	(0.023)
D: Treated	-2.919***	0.663***	-2.262***	-2.867***	0.630***	-2.239***	-3.194***	0.644***	-2.543***
	(0.058)	(0.013)	(0.052)	(0.047)	(0.011)	(0.042)	(0.051)	(0.014)	(0.045)
*D*B*	0.037	0.049***	0.091**	0.013	0.039***	0.060**	-0.034	0.023*	-0.015
	(0.051)	(0.011)	(0.043)	(0.035)	(0.009)	(0.029)	(0.032)	(0.012)	(0.025)

	Unclassified			Young Adults			Young Families		
	(10) lnq	(11) lnprice	(12) lnrev	(13) lnq	(14) lnprice	(15) lnrev	(16) lnq	(17) lnprice	(18) lnrev
B: After	-0.442***	0.049***	-0.406***	-0.184***	0.026*	-0.169***	-0.218***	0.039***	-0.168***
	(0.067)	(0.009)	(0.063)	(0.039)	(0.014)	(0.029)	(0.031)	(0.014)	(0.019)
D: Treated	-2.706***	0.682***	-2.032***	-2.693***	0.763***	-1.935***	-2.593***	0.717***	-1.862***
	(0.089)	(0.019)	(0.084)	(0.054)	(0.018)	(0.043)	(0.046)	(0.019)	(0.035)
*D*B*	-0.045	0.072***	0.047	-0.054	0.081***	0.023	-0.037	0.047**	-0.021
	(0.075)	(0.011)	(0.072)	(0.049)	(0.017)	(0.036)	(0.037)	(0.018)	(0.023)
N	176	176	176	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.11: Control 1: Large stores, immediate consumption versus take-home products for families

Panel A: Older Families

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	0.003	-0.008	-0.004	-0.197***	0.043***	-0.132***
	(0.022)	(0.009)	(0.019)	(0.026)	(0.009)	(0.020)
D: Treated	-0.938***	0.110***	-0.832***	-3.001***	0.556***	-2.437***
	(0.042)	(0.015)	(0.040)	(0.044)	(0.013)	(0.040)
*D*B*	-0.134***	0.069***	-0.048*	-0.040	0.016	-0.023
	(0.028)	(0.014)	(0.026)	(0.029)	(0.012)	(0.023)
N	176	176	176	176	176	176

Panel B: Young Families

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	0.032	-0.002	0.025	-0.238***	0.035***	-0.203***
	(0.023)	(0.010)	(0.025)	(0.036)	(0.013)	(0.025)
D: Treated	-0.548***	0.172***	-0.383***	-2.814***	0.588***	-2.222***
	(0.041)	(0.016)	(0.043)	(0.052)	(0.017)	(0.042)
*D*B*	-0.241***	0.056***	-0.164***	-0.064	0.033**	-0.042
	(0.027)	(0.014)	(0.031)	(0.043)	(0.016)	(0.029)
N	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.5 Conclusions

In assessing the effectiveness of the government's location restriction policy on HFSS products we observed sales of sugar-sweetened beverages in UK stores before and after the policy. Due to potential limitations in our data set, we constructed three potential control groups to help us observe the counterfactual scenario where SSBs were not affected by the policy. Overall we find there is no effect on sales or revenue of SSBs, with a small increase in price using our preferred control group - non-SSBs in the same stores in England. When we break down consumption into immediate consumption goods versus take-home goods we find there is a negative effect on sales of immediate consumption drinks but there is no such effect on take-home drinks. As immediate consumption drinks only account for 7% of total sales however, we ultimately conclude that the policy is largely ineffective in further reducing sugar consumption from SSBs following the large reduction that occurred as a result of SDIL. The effect on immediate consumption drinks makes intuitive sense; it is these drinks that are typically found in refrigerators close to checkouts and store entrances which face the largest reduction in 'quality' as a result of the policy making these drinks more difficult or more inconvenient to purchase. Therefore any policy that aims to truly reduce sugar consumption further needs to target take-home products sold in multi-packs in larger stores. At the same time, the minimum floor size requirement over which stores are subject to the policy, means that shops that sell a greater proportion of immediate consumption drinks are not affected by the policy. Although beyond the scope of this paper, there are potential policies specific to SSBs that the government could enact. These include additional tiers of taxation, advertising and promotion restrictions on television, online and in-print and banning large multipacks.

However, it should be noted that SSBs are only one product that is covered under the incredibly wide reaching HFSS location restriction policy. The policy targets not only sugar, but salt and fat, and the majority of these products save for

SSBs do not have other policies such as the SDIL attached to them. We cannot comment on the effects of the policy on these products, and there has so far been little published work in this domain. Setting a single policy covering so many items is undoubtedly challenging, but if there are significant improvements in salt and fat consumption from other products, then the seemingly marginal improvements in sugar consumption are 'icing on the cake'. In other words, adding SSBs to a policy that effectively targets salt and fatty products costs so little that any reductions in the consumption of sugar are a net positive.

References

Abdelaal, M., C. W. le Roux, and N. G. Docherty (2017). "Morbidity and mortality associated with obesity". *Annals of translational medicine* 5(7).

ACS (2022). *The Forecourt Report 2022*. URL: https://cdn.acs.org.uk/public/acs_forecourt_report_2022_d2_v1_aw_lr_spreads.pdf.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019). "Regressive sin taxes, with an application to the optimal soda tax". *The Quarterly Journal of Economics* 134(3), pp. 1557–1626.

Bilinski, A. and L. A. Hatfield (2018). "Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions". *arXiv preprint arXiv:1805.03273*.

Bokhari, F. A., P. W. Dobson, M. Morciano, and M. Suhrcke (2023). "Banning volume discounts to curb excessive consumption: A cautionary tale". *European Economic Review* 156, pp. 1044–80.

Colchero, M. A., B. M. Popkin, J. A. Rivera, and S. W. Ng (2016). "Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study". *British Medical Journal* 352.

DHSC (July 2020). *Tackling obesity: Empowering adults and children to live healthier lives*. URL: https://www.gov.uk/government/publications/tackling-obesity-government-strategy/tackling-obesity-empowering-adults-and-children-to-live-healthier-lives.

— (Nov. 2022). *New obesity treatments and technology to save the NHS billions*. URL: https://www.gov.uk/government/news/new-obesity-treatments-and-technology-to-save-the-nhs-billions.

— (June 2023). *Restricting promotions of products high in fat, sugar or salt by location and by volume price: implementation guidance*. URL: https://www.gov.uk/government/publications/restricting-promotions-of-products-high-in-fat-sugar-or-salt-by-location-and-by-volume-price/restricting-promotions-of-products-high-in-fat-sugar-or-salt-by-location-and-by-volume-price-implementation-guidance.

DiMeglio, D. P. and R. D. Mattes (2000). "Liquid versus solid carbohydrate: effects on food intake and body weight". *International journal of obesity* 24(6), pp. 794–800.

Dubois, P., R. Griffith, and M. O'Connell (2020). "How well targeted are soda taxes?" *American Economic Review* 110(11), pp. 3661–3704.

Ejlerskov, K. T., S. J. Sharp, M. Stead, A. J. Adamson, M. White, and J. Adams (2018). "Supermarket policies on less-healthy food at checkouts: Natural experimental evaluation using interrupted time series analyses of purchases". *PLoS medicine* 15(12), e1002712.

Fearne, A., N. Borzino, B. De La Iglesia, P. Moffatt, and M. Robbins (2022). "Using supermarket loyalty card data to measure the differential impact of the UK soft drink sugar tax on buyer behaviour". *Journal of Agricultural Economics* 73(2), pp. 321–337.

Freyaldenhoven, S., C. Hansen, and J. M. Shapiro (2019). "Pre-event trends in the panel event-study design". *American Economic Review* 109(9), pp. 3307–3338.

Griffith, R., M. O'Connell, and K. Smith (2017). *Design of optimal corrective taxes in the alcohol market*. Tech. rep. IFS Working Papers.

Griffith, R., M. O'Connell, K. Smith, and R. Stroud (2020). "What's on the menu? Policies to reduce young people's sugar consumption". *Fiscal Studies* 41(1), pp. 165–197.

Griffith, R., M. O'Connell, and K. Smith (2019). "Tax design in the alcohol market". *Journal of Public Economics* 172, pp. 20–35.

Kahn-Lang, A. and K. Lang (2020). "The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications". *Journal of Business & Economic Statistics* 38(3), pp. 613–620.

O'Mara, J. and I. Vlad (Mar. 2023). *Looking back at 5 years of the UK Soft Drinks Industry Levy*. URL: https://www.wcrf.org/looking-back-at-5-years-of-the-uk-soft-drinks-industry-levy/.

PHE (Oct. 2015). *Sugar reduction: The evidence for action.* URL: https://assets.publishing.service.gov.uk/media/5a7f928c40f0b623026904b7/Sugar_reduction_The_evidence_for_action.pdf.

— (Mar. 2017). *Health matters: Obesity and the Food Environment.* URL: https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2.

Rambachan, A. and J. Roth (2019). "An honest approach to parallel trends". *Unpublished manuscript, Harvard University.*

Roth, J. (2022). "Pretest with caution: Event-study estimates after testing for parallel trends". *American Economic Review: Insights* 4(3), pp. 305–322.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023). "What's trending in difference-in-differences? A synthesis of the recent econometrics literature". *Journal of Econometrics* 235(2), pp. 2218–2244.

Vasquez-Nicholson, J. (Dec. 2016). *UK Supermarket Chain Profiles 2016.* URL: https://apps.fas.usda.gov/newgainapi/api/report/downloadreportbyfilename?filename=UK%20Supermarket%20Chain%20Profiles%202016_London_United%20Kingdom_12-13-2016.pdf.

West, T. (2023). *Tesco CCO: Clubcards are now used across 80% of all Tesco transactions.* URL: https://www.marketing-beat.co.uk/2023/07/14/bellini-tesco-clubcard/.

World Health Organization (2015). *Guideline: sugars intake for adults and children.* World Health Organization.

Appendix

3.A Additional figures

Figure 3.A.1: Pre-trends between SSBs in English Extra stores and Petrol Stations in Scotland & Wales (Control 3)

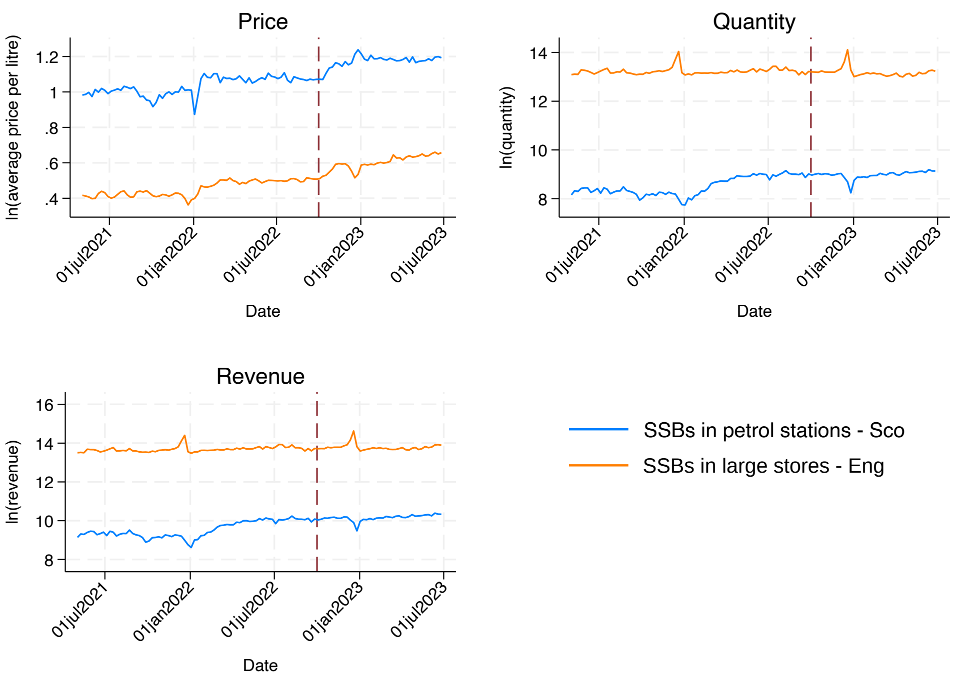


Figure 3.A.2: Yearly comparisons between SSBs and non-SSB in all stores exc petrol in England (Control 1)

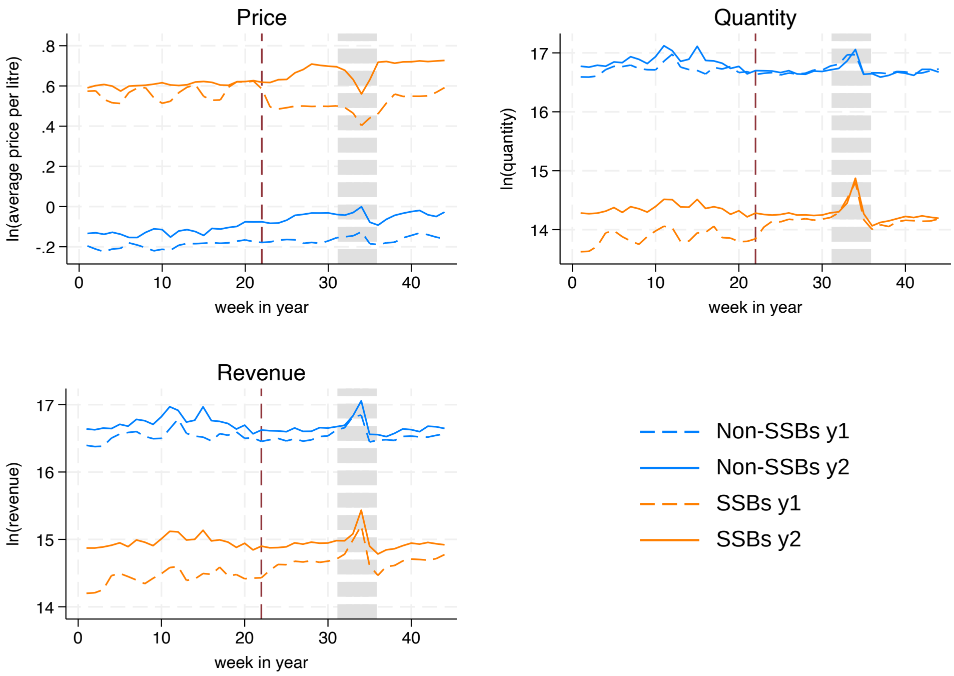


Figure 3.A.3: Yearly comparisons between SSBs in English large stores and Petrol Stations (Control 2)

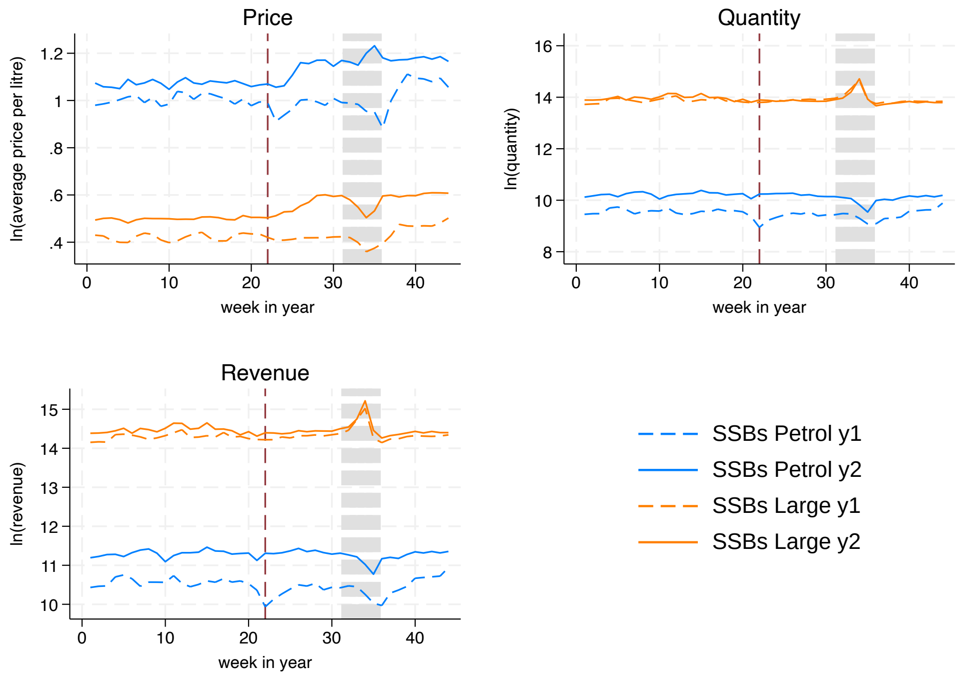


Figure 3.A.4: Yearly comparisons between SSBs in English large stores and Scottish/Welsh Petrol Stations (Control 3)

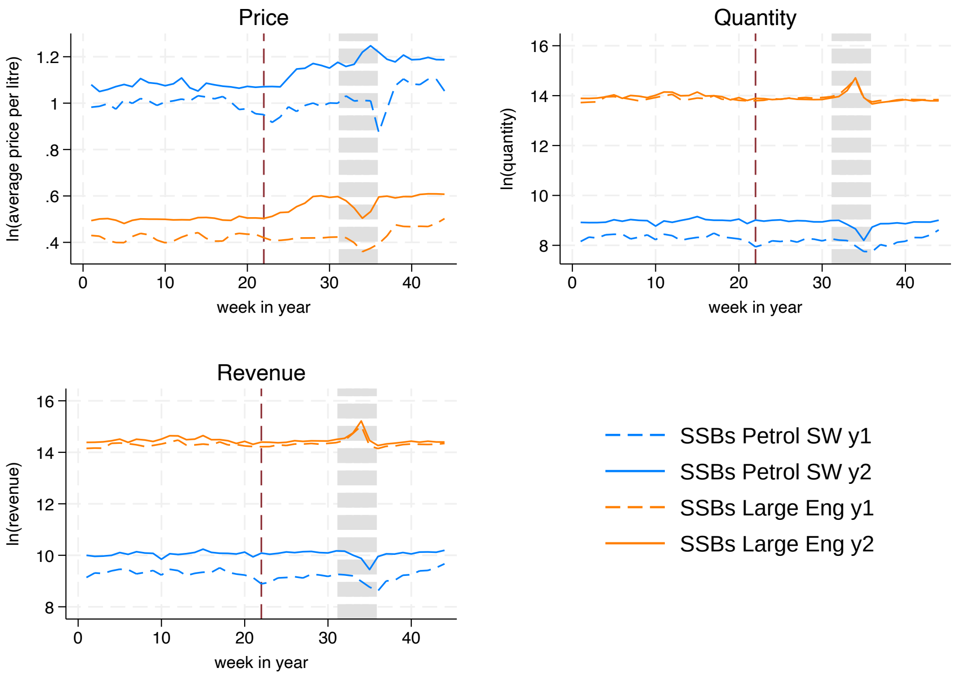


Table 3.A.1: Control 1: Immediate consumption versus take-home products in all stores ex. petrol stations

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.187***	0.030***	-0.151***	-0.244***	0.043***	-0.208***
	(0.034)	(0.011)	(0.031)	(0.040)	(0.012)	(0.031)
D: Treated	-0.802***	0.140***	-0.663***	-3.000***	0.578***	-2.425***
	(0.053)	(0.015)	(0.051)	(0.056)	(0.016)	(0.047)
*D*B*	-0.051	0.024*	-0.026	-0.048	0.030**	-0.017
	(0.041)	(0.014)	(0.038)	(0.047)	(0.014)	(0.036)
N	176	176	176	176	176	176

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes 43 week seasonal dummies, 43 D*week interactions and year dummy

Table 3.A.2: AIDS model results

	Non-SSB's	SSB's
Non-SSB's	-0.997***	0.004
	(0.049)	(0.047)
SSB's	-0.022	-1.026***
	(0.313)	(0.300)

Standard errors in parentheses

Crucially, the cross-price elasticities on the off diagonals are highly insignificant which suggests there is no substitution between SSB's and non-SSB's.

Table 3.A.3: Control 1: All stores, aggregate results by lifestage

	Older Adults			Older Families			Pensioners		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.210***	0.045***	-0.185***	-0.208***	0.038***	-0.177***	-0.233***	0.056***	-0.174***
	(0.052)	(0.011)	(0.048)	(0.041)	(0.007)	(0.038)	(0.036)	(0.010)	(0.030)
D: Treated	-2.758***	0.698***	-2.067***	-2.752***	0.664***	-2.092***	-3.121***	0.674***	-2.447***
	(0.071)	(0.014)	(0.065)	(0.057)	(0.011)	(0.053)	(0.056)	(0.014)	(0.049)
*D*B*	0.059	0.064***	0.123**	0.061	0.061***	0.117**	0.012	0.032***	0.044
	(0.067)	(0.012)	(0.061)	(0.051)	(0.008)	(0.048)	(0.041)	(0.011)	(0.034)

	Unclassified			Young Adults			Young Families		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.396***	0.041***	-0.371***	-0.213***	0.011	-0.227***	-0.249***	0.036**	-0.219***
	(0.068)	(0.009)	(0.064)	(0.049)	(0.017)	(0.041)	(0.042)	(0.017)	(0.029)
D: Treated	-2.569***	0.724***	-1.855***	-2.508***	0.785***	-1.729***	-2.459***	0.755***	-1.696***
	(0.090)	(0.018)	(0.086)	(0.066)	(0.022)	(0.056)	(0.058)	(0.023)	(0.046)
*D*B*	0.026	0.085***	0.132*	-0.014	0.099***	0.089*	0.003	0.052**	0.035
	(0.082)	(0.011)	(0.080)	(0.062)	(0.020)	(0.052)	(0.052)	(0.022)	(0.038)
N	176	176	176	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.4: Control 1: Large stores, immediate consumption versus take-home products for Older Adults, Younger Adults and Pensioners

Panel A: Older Adults

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.105***	0.030***	-0.075***	-0.226***	0.051***	-0.182***
	(0.024)	(0.009)	(0.021)	(0.042)	(0.010)	(0.035)
D: Restricted	-0.991***	0.096***	-0.901***	-3.094***	0.582***	-2.516***
	(0.043)	(0.014)	(0.040)	(0.059)	(0.013)	(0.053)
*D*B*	0.024	0.019	0.062**	0.000	0.035***	0.042
	(0.031)	(0.013)	(0.026)	(0.050)	(0.012)	(0.041)
N	176	176	176	176	176	176

Panel B: Young Adults

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.157***	0.038***	-0.106***	-0.199***	0.033***	-0.195***
	(0.026)	(0.012)	(0.023)	(0.051)	(0.010)	(0.045)
D: Restricted	-0.761***	0.144***	-0.617***	-2.946***	0.639***	-2.320***
	(0.045)	(0.016)	(0.043)	(0.067)	(0.014)	(0.061)
[1em] *D*B*	0.005	0.012	0.017	-0.102	0.072***	-0.007
	(0.031)	(0.015)	(0.027)	(0.063)	(0.012)	(0.057)
N	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel C: Pensioners

	Immediate			Take-home		
	(1)	(2)	(3)	(4)	(5)	(6)
	lnq	lnprice	lnrev	lnq	lnprice	lnrev
B: After	-0.056**	-0.002	-0.063***	-0.217***	0.058***	-0.133***
	(0.024)	(0.010)	(0.021)	(0.029)	(0.011)	(0.024)
D: Treated	-1.268***	0.024	-1.246***	-3.293***	0.604***	-2.680***
	(0.047)	(0.018)	(0.047)	(0.052)	(0.015)	(0.047)
*D*B*	-0.098***	0.060***	-0.033	-0.052*	0.015	-0.039
	(0.031)	(0.016)	(0.027)	(0.031)	(0.013)	(0.026)
N	176	176	176	176	176	176

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4

Assessing the welfare impact of HFSS location restrictions in England

Abstract

This chapter explores the welfare implications of the high in fat, sugar and salt in-store location restriction policy in the market for sugar-sweetened beverages (SSBs). We utilise aggregate sales from one of the UK's largest supermarket chains to estimate a structural demand model, before using these demand estimates to calculate consumer welfare under the policy and compare it to the counterfactual world where the policy was not introduced. We find there is a 1% reduction in consumer welfare as a result of the policy. We also estimate separate models by demographic groups, and find some evidence of heterogeneous effects whereby Families who purchase proportionately more SSBs are worse off. Given we found the policy had no overall effect on the sales of SSBs in Chapter 3, we then extend the policy to all stores in a further simulation. Here we find that sales of SSBs are reduced by 0.5% at the aggregate level. When comparing demographics, again we find Families are worst off. However, we caution against extending the policy as the costs of doing so will likely exceed the minimal benefits.

4.1 Introduction

Reducing sugar from soft-drinks, or sugar-sweetened beverages (SSBs) has, in the last decade or so, been targeted by governments around the world in a bid to reduce obesity and obesity-related illness.[1] As well as the lifestyle costs to the individual of obesity and related illnesses, there are direct costs to society in the form of health expenditure from treating diabetes, heart disease and related cancers.[2] Indirect economic costs arise through absenteeism, lower productivity and infrastructure adjustments to accommodate larger individuals.

Much of the policy response has been in the form of taxation, either based on the volume of drink sold, or the sugar content of the drink (see Table 4.A.1 for list of selected countries and the tax policy for SSBs). Evidence from Mexico, South Africa and the UK show these taxes have reduced consumption from SSBs and resulted in reformulation (changes in recipe) of products to reduce sugar levels, usually by changing the sugar/artificial sweetener composition. Beyond taxation, there are a range of other policies that either restrict or change choice sets, or improve information provision on the dangers of excess sugar consumption.

In this paper, we further evaluate one such policy that restricts choice sets. The UK government introduced a policy in October 2022 that restricted the sale of high in fat, sugar and salt (HFSS) products in certain locations within a store including near checkouts and ends-of-aisles in England. In Chapter 3, we used the market for drinks to test the effects of this policy on sugar purchases because SSBs are among the 'unhealthiest' of grocery products. We show the policy has had minimal effect in reducing the overall quantity of SSBs purchased. However, due to the design of the policy there were heterogeneous effects on products designed for immediate consumption versus those intended for take-home consumption.

[1] The definition of an SSB for our purposes is any drink that contains greater than 5g of added sugar per 100ml of liquid. This includes Coca-Cola but not Diet Coke or Coke Zero which fall under the category of non-SSBs. Fruit juices that contain naturally occurring sugars and dairy products are exempt. Alcoholic beverages and tea, coffee and other drinks intended to be consumed hot are not included in our universe of products.

[2] see Griffith (2023) for a comprehensive analysis of the costs of obesity.

While there was a 10-20% reduction in sales of immediate consumption SSBs there was no significant change to sales of take-home SSBs. Given immediate consumption SSBs only make 6-7% of total sales, we find, assuming there is no stock-piling, the overall consumption of SSBs and hence the consumption of harmful sugars has not been reduced as a result of this policy. We did not document any welfare implications in Chapter 3.

Here we first compute the overall welfare impact of the policy. Then, given the heterogeneous effects by product type, we investigate whether different demographic groups face different welfare effects as a result of the policy. Do some groups see an overall effect in the policy in spite of it only effecting a small percentage of available products? We show that Families purchase a greater percentage of all drinks relative to the number of customers. In Table 4.1 we show that the market share by volume of Families is 42.2% while the share by customers is 34.2%. The market share by volume for immediate consumption drinks is identical. This reflects the fact that Adults are buying for 1-2 people but Families are buying for on average 1.9 people, including children. As a result, we might expect that Families faces a greater change in welfare because they consume a proportionately larger percentage of immediate consumption goods which were more affected by the policy. The results support this idea. Families were the only demographic to see a fall in welfare as a result of the policy regardless of the method we used to calculate this change.

Having conducted a welfare exercise on the current policy application, we ask the natural follow up question - would the policy effect be any larger and would the welfare effects be any different if we extend the policy. This could be either be done by better targeting take-home products or expanding the policy to include all stores. The current policy only applies to stores greater than 2,000 sq ft (185.8m^2) of customer facing floor space. We focus on an expansion to smaller stores with the aim of reducing consumption of immediate consumption drinks in these stores. In theory, this is a straightforward extension. Better targeting take-home SSBs, on the other hand, would require more thought - we have currently not found a method to more target take-home drinks through restricting in-store location only. Nevertheless, increasing the policy scope to all stores may be a costly exercise. The fuzzy delineation between front and back of store in very small stores minimises the inconvenience effect of the location restriction. Alternatives could be prohibitively costly to small, independent retailers creating a situation where only large chains stock these items, causing competition issues and higher prices for consumers. Before considering these alternatives then, we feel it is worth simulating the effects on sales and prices of extending the policy to observe whether it would achieve the governments obesity reduction aims. By extending the policy to all stores we shows the share of SSBs falls by only 0.5%. To put this figure in context, there was a 33.6% reduction in the mean volume of SSBs purchased following the introduction of the Soft Drinks Industry Levy (SDIL) in 2018 (Fearne et al., 2022). We conclude therefore, that even extending the policy by store type, has only a small effect on SSB purchases.

Finally, we reevaluate consumer welfare in our simulated counterfactual scenario at the aggregate level and for each demographic group to observe whether they are heterogeneous effects of the policy extension for different groups. Since the policy aims to reduce consumption of SSBs we think of the limitation of in-store locations as a reduction of the product 'quality' of an SSB. Therefore, for these products we would expect to see a reduction in welfare *ceteris paribus*. Indeed, we see a £0.01-0.02 per litre reduction in consumer and total welfare at the aggregate sample level, reflecting the small reduction in sales above.

To conduct the above we specify an empirical model using weekly sales data from a large UK supermarket, for non-alcoholic beverages (excluding tea and coffee) between May 2021 and June 2023, in the spirit of Berry et al. (1995, henceforth BLP). We use this model to estimate parameters of the utility function. While a reduced form estimation can tell us about the effects on sales and prices, a full demand estimation is required in order to estimate welfare effects. Although the data is at the aggregate level, we do have sales split into five demographic groups. This allows us to comment on whether the policy and our counterfactual scenario affected different groups in different ways and whether this reflects the intended response of the policy. On this data, we estimate a separate random coefficient structural model of demand for SSB consumption in England for each demographic group, to obtain coefficients on price and other product characteristics for the indirect utility function.

We contribute to the general literature on 'sin' goods, which we describe in section 4.1.1. As far as we know, we are the first paper to analyse the welfare impacts of location within a store. This compliments works that has analysed the effects on sales and welfare of changing opening times (Hinnosaar, 2016), changing store opening days (Middleton et al., 2010) and the location of the store itself (Campbell et al., 2009). Most of these papers measure these effects on alcohol. Indeed, legislation and restriction on alcohol consumption are ahead of those in the domain of sugar and SSBs. By shifting our focus to SSBs, we join a growing list of contributors in the sub-field. Additionally, other studies typically use market sales data at an aggregate level, and may include demographic dummies to account for the effects by group. Their reported parameters in the estimation, then, would still reflect the distribution at an aggregate level. By estimating separate models, we are able to report the distributions of each parameter for each demographic group. This adds a richness to our paper; we are able to compare the preferences of each demographic with the aggregate sample, and with each other group. In the absence of data on the supply side, we estimate marginal cost for the products in our sample by solving the system of profit-maximising conditions for the multi-product firms operating under a Nash-Bertrand equilibrium. Combined with the demand-side parameters, we use these estimates to simulate a counterfactual world, by changing the values of certain variables in the model to increase the scope of the restrictions and re-estimating new prices and market shares in this alternative scenario. By removing the floor size cut off we can observe whether total consumption of SSBs has decreased, as per the stated aims of the policy, as well as observe any changes to consumer welfare.

When looking more closely at individual demographic groups we chose to aggregate the sample into Adults, Families and Pensioners, described in more detail in section 4.4. Analysing existing consumption patterns we find that Families disproportionately purchase (and therefore likely consume) drinks and SSBs in particular. They make up 34.2% of the supermarket's customers but purchase 46.1% of SSBs and 42.5% of all drinks. This is reflected in the fact that this is the only group that sees a reduction in consumer welfare due to the policy. This is in contrast to Adults who make up 51% of customers but only purchase 44.4% of SSBs and Pensioners who make up 14.2% of customers but only purchase 8.8% of SSBs. The policy has no real negative impact on these groups because they purchase fewer SSBs on a per capita basis.

From the government's perspective the policy and any extension could also serve to achieve one of the main policy objectives which is to reduce childhood obesity (DHSC, 2020). By construction, Families is the only group to contain households with children and so a policy that targets this group ought to reduce child access to SSBs by reducing the amount Families purchase. In our previous chapter we suggested that the policy had only minimal effect in reducing the consumption of SSBs. That conclusion is supported by our welfare estimations. We find that there is almost no difference in the consumer and total surplus values whether we evaluate the current policy, or the counterfactual extension. Likewise, where prices and shares change, these values are small, often less than 1%. The reason is that SSBs are already only a very small percent of the market and this reduction occurred primarily as the result of reformulation following the introduction of the SDIL. The remaining consumption is likely to come from consumers that have a high preference for SSBs, with low own-price elasticity and shifting demand from these consumers is challenging without an outright ban. On the face of it, our results suggest that extending the policy will marginally reduce sales of SSBs with only a small reduction in welfare. But this must be evaluated against the cost to stores of implementing this policy.

How does one restrict the location of a product in a small store when the distance between front and back is fuzzy? A possible solution would be to place drinks in locked fridges or lock boxes, which has been introduced for alcohol in some stores (Jones, 2024). However, this would be very expensive especially for small and independent stores and difficult to enforce. We feel that the cost will be greater than the reduction in SSB consumption so the HFSS location restriction method of restricting sales is not effective for SSBs. Ultimately, the policy direction needs to be in educating younger consumers on the health effects of sugar and creating long-term cultural change in tastes and preferences because short term limitations on demand have been almost fully realised.

4.1.1 Literature

Currently, the literature surrounding this policy is limited. Aside from our own work in Chapter 3, there are no papers that analyse the effects of this specific policy. Implicitly, there is no work estimating effects from counterfactual versions of the policy and nothing in the realm of welfare changes arising from the actual policy and/or counterfactuals. As we discuss now, there is a large literature on HFSS and alcohol more generally.

SSB consumption can be considered under the topic of 'sin goods' that include alcohol and cigarettes. In fact, of the three, SSBs has had the least focus in the academic literature until recently. Many of the leading papers in this domain focus on alcohol and methods to reduce consumption focus on taxation. But since the overarching aims associated with reducing alcohol consumption and reducing SSB consumption are both to reduce negative externalities and negative internalities, earlier work in the economics of alcohol consumption is extremely relevant to SSBs. Important work by Griffith et al. (2017, 2019) indeed uses the alcohol market in the UK to evaluate optimal tax design for 'sin' type goods. Their main finding is that if consumer preferences are heterogeneous and correlated with marginal externalities then varying tax rates across products by ethanol content can improve welfare in comparison to a flat-rate tax. In particular they find that 'heavy drinkers have systematically different patterns of alcohol demands and welfare gains from optimally varying rates are higher the more concentrated externalities are among heavy drinkers' (Griffith et al., 2019). This notion is directly applicable to SSB legislation; indeed the SDIL is a two-tiered tax dependent on the amount of sugar in a drink.

In Mexico, where a volumetric SSB tax of 1 peso (3.8 pence) per litre was introduced in 2014, Pedraza et al. (2019) find there was a 37% reduction in SSBs purchased in 2016 compared to the year before the tax. The UK introduced the SDIL in 2018. Between 2015 and 2019, the percentage of drinks in supermarkets with sugar content of more than 5 grams per 100 ml, (the threshold for the SDIL) fell from 49% to 15% (Scarborough et al., 2020). By 2023, more than 45,000 tonnes of sugar had been removed from soft drinks in the UK, reducing the sugar content in SSBs by nearly 45% (PHE, 2020). Sugar consumed from SSBs was also reduced by 35.4% in the same period (PHE, 2020).

O'Connell and Smith (2021) show how market power impacts efficiency and redistributive properties of an optimal tax framework in the context of the UK SSB market. They find that ignoring market power when setting optimal SSB tax rates lead to welfare gains 40% below optimum. A move from a single-rate to a multi-rate system can result in further substantial welfare gains. The multi-rate system they propose involves setting different tax rates on different drinks types which enables better targeting of the externality and market power distortions. An additional shift to a sugar tax rather than a volumetric tax can further increase welfare. This is an important point. All SSB levies in US cities are based on a flat x amount per ounce of liquid. In contrast, the UK SDIL has a two tier tax rate based on the amount of sugar present in the drink. As evidenced in the UK, the latter incentivises firms to alter their recipes away from sugar into artificial alternatives to avoid the tax.

Dubois et al. (2020) find that a soda tax does a good job of targeting young consumers and those from lower income households but not individuals with high total dietary sugar. They focus on on-the-go drinks purchased for immediate consumption which they find accounts for almost half of sugar obtained from SSB's. Although their data set does not include information about purchases in bars and restaurants which constitutes a quarter of all on-the-go purchases, it does include data from vending machines, convenience stores, kiosks and larger grocery stores. In comparison our data set does not contain any information about vending machines and kiosk purchases and likely fails to capture a large proportion of convenience store purchases also. Unlike soda taxes which impact a drink regardless of purchase location, the location restriction policy comes into effect only in large stores, greater than 2000 sq ft. As we show in Chapter 3, the proportion of on-the-go drinks sold in large stores is small - around 6-7% - such that despite its apparent success in reducing consumption of immediate consumption SSB's, the overall effect is not statistically different from zero because it fails to impact purchases for take-home consumption which contribute much more to those with high total dietary sugar.

Conlon and Rao (2023) explore an alternative route to curbing negative externalities in regulations that limit competition. They show that a common regulation used to curb alcohol consumption through market power called post-and-hold (PH) leads to substantially lower welfare and government revenue when compared to even simple taxes on the same products because it distorts competition choices away from premium brands towards low-cost alternatives. They estimate that replacing PH with volumetric taxes could reduce consumption by 10-11%, without reducing consumer surplus, while tripling tax revenues. In a complimentary paper, Conlon et al. (2022) find that sin-good purchases as a whole, including alcohol, cigarettes and sugary beverages are highly concentrated; 10% of households pay 80% of the taxes on alcohol and cigarettes. The most taxed households on alcohol and cigarettes are older, less educated and lower income. Taxes on SSBs broaden this base but add to the burden of heavily taxed households. The idea that taxes are more effective than market power solutions to the negative externalities problem, but have their own problems with regards to equity and consumer welfare motivates the design and implementation of additional policies such as the HFSS location restrictions which we analyse here.

Away from taxation, Griffith et al. (2020) consider policies to reduce young people's sugar consumption including advertising restrictions, and potential restrictions on the availability of products or changing the characteristics of products (such as moving from sugar to sweetener). At the time of their work, they were no location restriction polices in place and advertising restrictions had not been made law so they do not conduct any empirical work to test the potential effects of these policies. Several countries have targeted food labelling and advertising to further reduce consumption of SSBs (GFRP, 2024). In the UK the government introduced, simultaneously at the end of 2022, a 9pm TV watershed for high in fat, sugar and salt (HFSS) products and a restriction of paid-for HFSS advertising online (DHSC, 2021). This highlights the importance of the UK HFSS policy as one of the first of its kind globally, providing all stakeholders with an opportunity to empirically understand the effects of these types of policy on purchases, prices and revenues.

While there is little evidence with respect to this location policy, there is some work in more general restrictions on other products, particularly alcohol. Middleton et al. (2010) conduct a meta-analysis of reviews conducted in the US on the effects of restricting days of alcohol sales on alcohol consumption and related harms. They find increasing the days or hours of sale by removing existing restrictions (e.g. full bans on Sunday sales, limits on times high alcohol drinks could be sold) increased excessive alcohol consumption and increased the risk of motor vehicle crashes, incidents of DUI, police interventions against intoxicated people, and, in some cases, assaults and domestic disturbances. Of course, forward looking consumers can circumvent these restrictions by planning their purchases in advance. Although inconsistent time preferences could provide a justification for these sales restrictions, Hinnosaar (2016) find only 3% of all consumers exhibit time inconsistent preferences using scanner data of beer purchases. They also find that although the sales restriction may be welfare improving it is worse than increasing alcohol duties. A further policy restriction on alcohol is to regulate the density of stores selling alcohol. Campbell et al. (2009) unsurprisingly find that increasing the density of stores increases alcohol consumption and associated harms. Beyond philosophical questions about whether governments should behave in such a paternalistic fashion or how consumers might respond to day and time restrictions on the sale of SSBs, together, these studies suggest there are options for policymakers who wish to further target SSB consumption.

The rest of the paper is organised as follows. We provide further detail on the location restriction policy and its introduction in section 4.2.1, alongside a description of our data set. In section 4.3, we outline the elements of the structural model we use to estimate market equilibria including both the demand- and supply-side, a discussion on instruments, how we calculated market size and our general model specification including non-price right-hand side variables. In section 4.4, we present our results. First we discuss the structural model of demand and our post-estimation predictions of marginal costs and consumer surpluses. Then we provide comparisons between a no restriction scenario, a full restriction scenario and the existing policy environment. Finally, we provide our concluding comments in section 4.5, discussing potential improvements to the current legislative framework.

4.2 Background and Data

4.2.1 Background

In 2016 the UK government announced the Soft Drinks Industry Levy (SDIL) to contribute to their plans to reduce childhood obesity by removing added sugar from soft drinks (HMRC, 2016). Discourse surrounding the individual and social effects of sugar intake, especially in children had grown following the publication in 2015, of a World Health Organization (WHO) guideline paper, based on meta-analysis of global randomized controlled trials that recommended a 'reduced intake of free sugar throughout the lifecourse, reducing the intake of free sugars to less than 10% of total energy intake in both adults and children, and a further reduction of free sugars to below 5% of total energy intake' (World Health Organization, 2015). The SDIL followed reports by Public Health England (PHE, 2015) and a Colchero et al. (2016) analysis of an SSB excise tax in Mexico that showed an average purchase decrease of 6% on taxed beverages, and eventually became legislation in 2018. Since its introduction more than 45,000 tonnes of sugar have been removed from soft drinks in the UK and tax revenues of £334 million were raised in 2021-22 in the process (O'Mara and Vlad, 2023). A recent study by Rogers et al. (2023) reports that compared with trends before the SDIL was announced, 1 year after implementation, volume of all soft drinks purchased combined increased by 189 mL, or 2.6% per household per week. However, the amount of sugar in those drinks was 8g, or 2.7%, lower per household per week.

Nevertheless, obesity (DHSC, 2020), the treatment of which is estimated to cost the NHS £6 billion per annum, rising to nearly £10 billion per annum in 2050 (DHSC, 2022), remains a problem to PHE. Wider costs to society from obesity manifest through absenteeism, productivity reduction and accommodations of heavier individuals are estimated to be around £27 billion (PHE, 2017). In a bid to curb consumption of those foods and drinks associated with obesity, DHSC

(2023) introduced legislation 'to restrict the promotion of HFSS products by volume price (for example, 'buy one get one free') and location, both online and in store in England' in 2020. The first of these is due to come into force on October 1st 2025, following two postponements. The location restriction policy, however, came into effect on October 1st 2022.

As stated in Chapter 3, the policy restricts the placement of HFSS products in certain areas of a store for stores that have 185.8m^2 (2,000 sq ft) or greater of relevant floor area including any area within 2m of the checkout facility, any area within 2m of a designated queuing area or queue management system, the ends of aisles, store entrances, and covered external areas. Small businesses with less than 50 employees and specialist retailers are exempt from the location restriction policy (though not from the volume price policy). Although both policies apply to a wide range of HFSS products, again we focus on SSBs in this study. The definition of an SSB for the purposes of the policy are derived from SDIL definitions such that any drink with *added* sugar > 5g per 100ml of liquid is subject to the restrictions. In other words, all beverages that are covered by the SDIL are also impacted by the new policies. Drinks that contain natural sugars, dairy products, and alcoholic beverages are exempt.

4.2.2 Data

Our raw data is the same data we used in Chapter 3. It is taken from a database that contains weekly sales of stock-keeping units (SKU's) by home nation, store type, and lifestage from a random 20% of loyalty program customers of one of the UK's largest supermarket chains. Loyalty transactions represent approximately 80% of the supermarket's sales (West, 2023). The raw sample contains observations from 113 weeks beginning 3rd May 2021 to 2nd July 2023 on over 2,500 SKU's. For each SKU, we have information on the number of units in a pack, the size of each unit, the number of units sold and the revenue generated. From this information we are able to calculate quantity in litres and price per litre as revenue/quantity. We supplement this data with information on the nutritional content of the drinks including energy, protein, sugar per 100ml and a dummy for whether the drink contains any artificial sweeteners.[3] We are not concerned about the many zero sale products in the data set now, as they have 0 shares and drop out during the estimation of the structural model. Therefore we retain all 2,500+ SKU's from the raw data. The only change we make is to split sales into those in large stores (Extra and Super) and those in smaller stores (Express and Petrol). As such, each product in each week has two observations, one in smaller stores and one in larger stores. These are denoted by adding 0.1 to the product IDs of those products in larger stores. Otherwise the product ID remain the same. This separation is relevant for two reasons. Firstly, the policy explicitly restricts the location in large stores but does not apply to stores that are less than 2,000 sq ft (185m^2). A large store dummy variable is therefore, in part, a proxy to whether an individual SKU is 'treated' by the policy.[4] Secondly, we believe these store types attract different population segments. Intuitively, we can imagine that smaller stores are found in more urban locations, close to younger populations whereas large supermarkets, in sub-urban locations are frequented more by older families. Therefore, the storetype dummy adds a further layer of detail about shopping preferences. Unfortunately, by the raw data construction, we are not able to separate the data by demographic group and store type simultaneously to confirm this. Finally, we add a dummy for whether the policy was in effect for that product, in that week, in that store. In the final simulation steps, it is this variable that we change. The aforementioned demographic groups allow us to separate sales into different categories which enables us to estimate demand and subsequently consumer welfare by demographic in section 4.4. Table 4.1 shows the relative sizes of each demographic group in our sample. We create these groups as combinations of

[3]Although in some cases the names of the particular sweeteners are listed on the product, in most instances the amount of sweetener used is not specified.

[4]In part only, because an individual SKU is only treated if it contains $>$ 5mg of sugar per 100ml, in the time-period after the policy was introduced, in stores that are larger than 2,000 sq ft.

demographics from the raw data. Definitions of the raw groups are provided in figure 4.A.1 of the Appendix. From these grouping Adults and Families together makes sense because an obvious distinction is that Families purchase a lot of kids products while Adults do not buy any. The reason we do this is because some raw demographic groups represent a relatively small percentage of the market so we run into problems where market shares of a given product are extremely small which can cause convergence issues during the estimation.[5]

Using the demographic groups in Table 4.1 we can see that Adults and Families are similar in market share by volume but Adults is much greater in size by market share by customers. This reflects the fact that average household size for Families will be larger so each individual shopper in the Family group is purchasing more items. Looking at the market share of SSBs by volume we see a similar tale. Each Family household purchases more SSBs than each Adult unit. As part of the governments aim is to reduce childhood obesity, if they are heterogeneous effects from the policy, it would be preferable if sales for Families were reduced more than the other groups. If this were the case, then we may expect a greater reduction in consumer welfare for Families also. However, as shown by Conlon and Rao (2023) and others, these types of policies often hit the poorest individuals disproportionately which raises issues of equity and fairness.

In Table 4.2 we show the variation in our product characteristics that allows for identification in the demand models. For the key characteristic price, variation comes from both cross-sectional differences in price between products (labeled 'Between' in Table 4.2) and differences in the same product across time (labeled 'Within' in Table 4.2. An example of this is shown in Figure 4.1. Between variation is shown by the fact that the two coke options have different prices per litre i.e. the red and blue line are different. Within variation is shown by the fluctuations of each line by week. Note the other characteristics only have cross-

[5]Given the five raw demographic groups, we can either combine by age to give 'Older' and 'Young' groups or by family unit to give 'Adults' and 'Families'. Pensioners remain a separate group regardless. Although we prefer to use Adults and Families, we also estimate models for Older and Younger groupings.

Table 4.1: Market share summary statistics

	Aggregate	Adults	Families	Pensioners
Mean weekly volume[a]	21,606	9,456	9,168	2,852
SD weekly volume[a]	2,581	1,667	1,056	445
Market share by volume	-	43.8	42.4	13.2
Mean weekly customers[b]	5,388	2,744	1,842	764
SD weekly customer[b]	292	143	77	77
Market share by customers	-	51.0	34.2	14.2
Mean weekly penetration	50.94	50.1	59.5	39.32
SD weekly penetration	3.16	3.21	2.90	3.80
Mean weekly SSB vol[a]	1,548	687	714	137
SD weekly SSB vol[a]	249	155	105	28
Market share SSB by vol	-	44.4	46.1	8.8

[a]measured in 1000's of litres, [b]measured in 1000's.

Penetration measures the number of unique customers that purchased any product in the data set as a share of all unique customers from a given group during the specified time period.

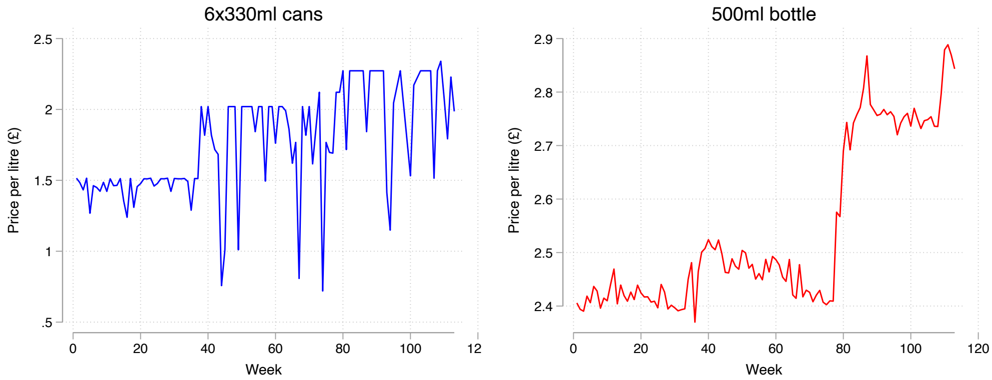
sectional variation because each characteristic is fixed across time for a given product.

Table 4.2: Variation in product characteristics

Variable		Mean	Std dev	Min	Max	Obs
Price (per litre)	Overall	1.679	1.137	0.022	8.491	91708
	Between		1.073	0.08	6.347	930
	Within		0.131	-0.238	0.243	113
Sugar	Overall	3.804	4.571	0	14	91708
Volume	Overall	1.678	1.759	0.2	12	91708
Divisor	Overall	3.082	4.288	1	24	91708

Panel dimensions are product, j by week, t. Variation over time for a given product is called *within (W)*, and variation across products (cross-section) is called *between (B)*. The overall variation is $s_O^2 = \frac{1}{M-1}\sum_j\sum_t(x_{it}-\overline{x})^2$ while $s_B^2 = \frac{1}{m-1}\sum_j(\overline{x}_j-\overline{x})^2$ and $s_W^2 = \frac{1}{M-1}\sum_j\sum_t(x_{jt}-\overline{x}_j)^2$. Observations are listed as overall (M), over number of products (m) for between, and weeks (T) for within.

Figure 4.1: Example of price variation for Coke options



4.3 Model

In this section, we set out the structural model we use to estimate demand in the soft drinks market, and analyse the effects of the policy restriction as well as our stated counterfactual scenarios on the market, firms and consumers.

4.3.1 Demand Specification

Following Berry (1994), Berry et al. (1995) and Nevo (2000) we start with a utility function of

$$U_{ijt} = \alpha_i p_{jt} + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{jt}. \tag{4.3.1}$$

where i indexes individual, j indexes products and t indexes markets. U_{ijt} is the utility that individual i receives from consuming product j in market t. This utility compromises a set of observable product characteristics, namely the price p of product j in market t, and a vector of other product characteristics x that we define later. We also have a set of unobservable (to the researcher) characteristics that are product and market specific, denoted by ξ_{jt} and a separably additive error term, ε_{ijt} that is individual, product and market specific. Finally, α and β

are the terms we seek to estimate that represent consumer preferences on price (marginal utility of income) and a vector of preferences on the remaining product characteristics. In the simplest version of equation 4.3.1, where α and β are not individual-specific we assume that consumer heterogeneity only enters the model through the error term such that consumer preferences are otherwise identical. If we also assume the errors are i.i.d and distributed Type I extreme-value, the model becomes the simple logit that suffers from unrealistic substitution patterns. By including individual-specific preferences characterised by random coefficients (RC) on the product characteristics we seek to avoid these problems. These individual-specific preferences are reflected in equation 4.3.1 by the i subscripts on α and β.

We prefer a random coefficients models to a nested-logit model because a nested-logit requires an artificial grouping of products that imposes restrictions on substitution patterns. Further, in a nested-logit, the order of nests matters which means that researchers must make *a priori* decisions on how consumers construct purchase decisions. Typically, this also assumes that all consumers follow the same internal nesting structure. While not without their own problems, we prefer to use RC models because they tend to offer greater flexibility in substitution patterns and better reflect real market structures. The flexibility of the RC model arises because it allows for consumer preferences on any given characteristic to be spread over a distribution. A standard assumption is that the distribution of these preferences is normal such that we estimate the mean and standard deviation of each of these normal distributions. In our RC models, we restrict preferences on price to be distributed log-normal. This forces all values of α to be negative which ensures that all demand curves in our estimations are downward-sloping and saves us from running into potential divide by zero errors during the counterfactual simulations.

In all models, the vector x_{jt} contains the sugar content of the drink in grams per 100ml and a dummy to indicate whether the product contains sweetener. The final model we use also includes a dummy to indicate whether the store was restricted by the policy, a dummy to indicate whether the product was restricted by the policy (i.e. is an SSB) and a policy term which indicates whether the product was restricted in that particular store in that particular week. This is exactly equivalent to a three-way interaction between the large store dummy, the SSB dummy and the post-restriction dummy.

From here we can predict market shares, $\tilde{s}_{jt}$ for a given set of parameters as a function of product characteristics, prices and structural errors where set A_{jt} is the set of individuals who choose product j in market t (because it gave them the highest utility),

$$\tilde{s}_{jt} = \int_{A_{jt}} dP^{*}_{\varepsilon}(\varepsilon), \tag{4.3.2}$$

where $P^{*}(.)$ denote the population distribution function. A natural approach would then be to find $\theta = \{\alpha, \beta\}$ that minimises the distance between the observed market shares s_{jt} and our model predicted market shares, $\min \sum_{j=1}^{J}[s_{jt} - \tilde{s}_{jt}]^2$. Crucially, ξ enters the predicted market shares and is not additively separable so this non-linear least squares estimator is inconsistent. To avoid this, we use the 2-step GMM procedure set out by BLP (1995). We carried out our analysis using pyBLP, a publicly available package in Python designed specifically for structural demand estimation using BLP's methods (Conlon and Gortmaker, 2020).

4.3.2 Supply side

Having obtained demand side parameters, we use a standard model of Nash-Bertrand oligopolistic competition between multi-product firms to estimate the marginal cost associated with each product. Joint-profit of firm f, producing a subset $\mathcal{F}_{fm} \in J_m$ products is given by

$$\Pi_{ft} = \sum_{k \in \mathcal{F}_{ft}} (p_{kt} - c_{kt}) s_{kt}(p), \tag{4.3.3}$$

where c is marginal cost and s is market share. Assuming the existence of a pure-strategy static Bertrand–Nash price equilibrium with strictly positive prices, each of the prices, p_{jt}, satisfies the following first-order conditions as in Dubé (2005):

$$s_{jt}(p) + \sum_{k \in \mathcal{F}_{ft}} (p_{kt} - c_{kt}) \frac{\partial s_{kt}(p)}{\partial p_{jt}} = 0. \tag{4.3.4}$$

Constructing a $(J \times J)$ ownership matrix, Φ, where element ϕ_{jk} is 1 if j, k are produced by the same firm and 0 otherwise and stacking the system of equations from 4.3.4 we obtain $(J_t \times 1)$ vectors $\mathbf{p}, \mathbf{c}$ and $\mathbf{s}$ in matrix form. Rearranging gives the classic markup equation

$$\mathbf{p} - \mathbf{c} = \boldsymbol{\Omega}^{-1}\mathbf{s}, \qquad \text{where} \qquad \Omega_{jk} = -\phi_{jk} \frac{\partial s_k(p)}{\partial p_j}, \tag{4.3.5}$$

Since we can estimate the markup, $\eta = \boldsymbol{\Omega}^{-1}\mathbf{s}$ from our demand parameters, $\mathbf{c} = \mathbf{p} - \eta$. We note here that the separate estimations for each demographic group produce demographic group specific marginal costs. Of course in reality such a thing does not exist. There is only a single marginal cost per product per market for the population as whole. Nevertheless estimating consumer and producer surplus, and running simulations for a demographic group requires the costs predicted for that particular group. We discuss this issue further in section 4.4.1.

4.3.3 Market Size

Market size, or market definition M is crucial to logit-type demand models particularly in the specification of the share of the outside good which gives consumers the option to purchase a product outside of the set of products in our sample. For a given M the share of the outside good is $s_0 = M - \sum s_j$ in a given market, where $\sum s_j$ is the total of our observed sales. The challenge then is to correctly predict M or specify a model that is robust to different methods of estimating M. Utilising two variables in the dataset; 'customers' and 'customer penetration', we calculate market size, specifically the share of the outside good as a best estimate from the information available. The variable Customers measures the unique number of customers in a specified group that purchased at least one item from our product set during a given time period. Customer penetration (hereafter penetration) expresses the number of customers relative to the number of total unique customers from that group that purchased any item from Tesco during the same time period, as a percentage. From these two figures we are then able to back out the total number of unique customers in the particular time period. However, we only have customer and customer penetration figures from week 53 in the sample onwards. We express market size in volume terms. For each market, we calculate the potential volume as the total volume purchased in the market over the penetration multiplied by 100. As such we are estimating the demand for beverages within Tesco stores in the UK specifically and the outside good is Tesco customers that do not purchase a drink in a given week. Although it may be possible to extrapolate this to estimate demand nationally, we feel there are too many assumptions required to accurately predict this.

$$\text{potential volume}_t = \frac{\text{total volume}_t}{\text{penetration}_t} * 100. \tag{4.3.6}$$

Since we do not have penetration figures for weeks 1-52, for week t in this period we substitute in the penetration figure from week $t + 52$.

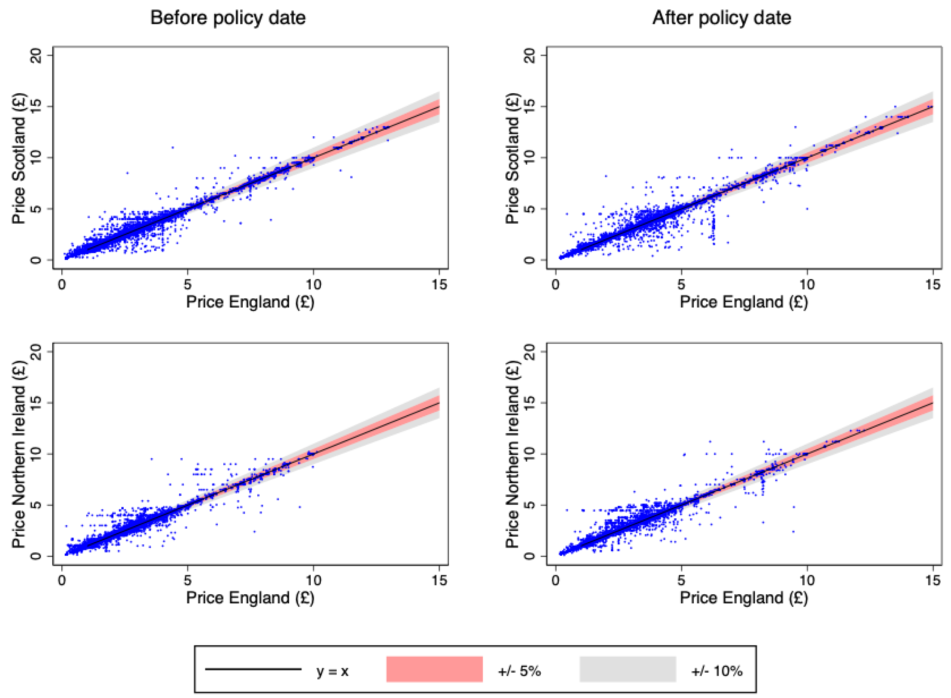
4.3.4 Instruments

One of the primary challenges associated with random coefficient models is the selection of instrument variables. The presence of the structural error term ξ in the utility function leads to an endogeneity problem because ξ is often correlated with price. Therefore, we must find a set of instruments to include in the model. Further, as stated by Berry and Haile (2021, p.24) 'in a parametric model, identification requires at least as many moment conditions as parameters, and the parameters of the model include not only the coefficients $\theta_1 = (\alpha_0, \beta_0)$ on x_{jt} and p_{jt}, but also the parameters θ_2 governing the variation in the random coefficients'. Limitations in the data prevent us from using Hausman type instruments - the price of a product in other geographical markets - at the aggregate level. Although we have information on prices across four separate countries in the UK, and could theoretically use prices outside England as instruments for prices in England, we find that prices in any store type in any other country are almost a perfect predictor of prices in England. This is not surprising; supermarket chains are likely to negotiate prices and face cost-shifters at the UK-wide level. Further, the supermarket has a policy of price parity within store types across the UK for the same product. While we were unable to find official documentation of this policy, Figure 4.2 shows that for most products prices are close to the 45 degree line when prices in England are plotted against prices in Scotland and Wales. With a plus/minus 10% window, 95% of prices are equal. Secondly, we do not have information regarding prices in other grocery store chains.

Instead we follow the canonical instruments from Berry et al. (1995), supplemented by the differentiation instruments of Gandhi and Houde (2019) to make our initial estimates of θ.[6] Table 4.B.1 in the appendix list these instruments and their descriptive statistics. We use these to then estimate

[6] BLP instruments create excluded instruments by summing a product characteristics' values over non-rivals (i.e. same brand/producer) and separately over rival goods. Differentiation are sums over functions of differences between non-rival goods and sums over difference between rival goods.

Figure 4.2: Evidence of price parity across the UK



optimal instruments in the spirit of Amemiya (1977) and Chamberlain (1987). Reynaert and Verboven (2014) show that these instruments can reduce bias, improve efficiency and increase the stability of estimates. In testing best practices for estimating BLP models with PyBLP, Conlon and Gortmaker (2020) show empirically this assertion holds true. Section 4.B in the Appendix is dedicated to further discussion of instruments.

4.3.5 Other considerations

Including product-specific dummy variables can improve the fit of the model when a researcher is unsure that the observed characteristics are the true determinants of utility (Nevo, 2000). Additionally, product-dummies capture those characteristics that do not vary by market so that the correlation between prices and product specific mean of ξ_j is accounted for and does not require an instrument (Nevo, 2000). Of course the major problem with introducing product-specific dummies is that β cannot be identified directly from the estimation. Instead one must use Chamberlain's (1982) minimum-distance procedure to retrieve β.[7]

However, this poses a challenge when using the estimates to run a counterfactual simulation. With this in mind, we introduce brand-specific dummy variables, one-level up from the product. One can think of these as advertising brands. For example, all Coca-Cola products including Diet and Zero would be under one brand. We believe this allows us to capture a significant portion of product-specific unobserved characteristics without losing the ability to directly estimate β. Some practitioners also include market-specific dummies in BLP-type models. The challenge we face is that although we model multiple markets in actual fact these are the same geography (England) measured across consecutive time periods. From Berry and Haile (2021, p. 9) 'if the price of a given product is the same for all consumers in a given market (often this is the definition of "market"), a fixed effect for each product×market will leave no variation in price, making it impossible to measure demand elasticities or to connect demand to standard notions of aggregate welfare.' Additionally, market dummies, being as they are week dummies are co-linear with a pre/post policy dummy which we include as a right-hand side variable that forms part of the dummy variable that indicates whether a particular product in a given store in a given market (week) is restricted by the policy. Having tested both the inclusion of market-dummies and the pre/post dummy, we opt for the latter given the warning of Berry and Haile (2021), and indeed this model appears to fit the data better.

[7] This is because these product characteristics are constant within a particular brand. Using Chamberlain (1982) we first estimate a $J*1$ vector of brand dummy coefficients, $d = (d_1, ..., d_j)'$ using the previously described mixed logit procedure. From the original indirect utility equation 4.3.1 it follows that $d = X\beta + \xi$, where X is a $J*K$ matrix of product characteristics that are fixed and ξ' is a vector of $J*1$ unobserved product characteristics. Assuming that $E(\xi|X) = 0$ then $\hat{\beta} = (X'{V_d}^{-1}X)^{-1}(X'{V_d}^{-1}\hat{d})$ and $\hat{\xi} = \hat{d} - X\hat{\beta}$.

4.3.6 Counterfactual simulations

In order to comment on whether the policy could be improved, we run several counterfactual simulations based upon our structural model estimates. In a full restriction world, all restricted drinks, regardless of store type are subject to the location restriction. There may be physical challenges associated with this in the real world. For example, in a very small store how does one place a sufficient restriction on location to reduce consumption? Anecdotally, some supermarkets in the UK have been trialling electronic security cabinets in alcohol aisles to fight theft (Jones, 2024) so something similar could be applied to SSB's. However, the point of our exercise is to see whether this would reduce the consumption of SSBs rather than suggesting how to achieve it. The second extreme is a no restriction world, whereby all location restrictions are removed and SSBs can be placed anywhere in stores. This would have been the state of the world if the policy had not come into effect. We would expect sales of SSBs to fall further than the status quo in the full restriction model and sales to increase without any restrictions. We would also expect consumer surplus to move in the same direction. Together these allow us to observe the effects of the current policy versus no action and whether extending the policy is a useful exercise.

4.4 Results

4.4.1 Demand Estimation

We present the results from the full demand estimation model in Table 4.3, using both family type and age groupings.[8] As mentioned before, the counterfactual simulations we do involve changing the restrictions applicable to store types. Therefore each product has two observations - one for large stores and one for

[8] We also present results using the demographic groups from the raw sample in Table 4.A.2. In some cases, over 40% of marginal costs predicted by these model are greater than zero. As.a crude test of fit then, these demographic groupings are not appropriate, in part because the smaller percentage shares of some demographics results in some very small shares for some products which leads to computational problems in the estimation.

small stores. We differentiate between the same product in different stores by adding 0.1 to the product IDs of products in large stores. We report parameters for the aggregated sample in column 1 and then parameters for groups aggregated by family type in columns 2 and 3, aggregated by age in columns 4 and 5, and then finally Pensioners in column 6. In helping to decide which aggregation is the most appropriate we use the number of marginal cost estimates below zero as a crude measure of model fit. Using these figure it is clear that the aggregate model, the two models aggregated by family type and Pensioners are fitted well by the current specification with mc<0 of 1.33%, 1.24% and 0.837% respectively; and the two groups aggregated by age fit less well with mc<0 of 22.11% and 18.37% respectively.

As expected the coefficient on price is negative for all groups. However, it is not statistically different from zero in the aggregate sample and for the Older group. The policy term, which indicates whether a particular product in a particular store type is restricted or not is also negative among all groups. This supports the notion that the restriction negatively effects the 'quality' of a good and creates dis-utility. However, the mean value is not statistically significant for Older and Pensioners. Interestingly the sugar variable is positive for all groups except for Younger. This variable represents all sugar in a drink including both natural and added and so indicates that in general the groups prefer some sweet flavour to their drinks. The exception is the Younger group. This could be for health reason associated with sugary beverages or just a preference in taste. This notion is borne out by the facts that the coefficient on SSB, which is a dummy for whether the drink contains *added* sugar, is negative. Interestingly, in the aggregate sample the various tastes are balanced out to leave the coefficient on sugar not statistically different from zero. The coefficient on large store is notably positive and significant for all groups, suggesting that the preference is to shop at large supermarket type store over a convenience store wherever possible. This makes sense as these large stores have greater variety, and often cheaper prices.

Table 4.3: Random coefficient demand estimates

Age		-	-	-	Older	Young	Pensioner
Type		Aggregate	Adults	Families	-	-	-
Variable	Parameter	(1)	(2)	(3)	(4)	(5)	(6)
log Price	Mean	-1.637	-1.822^c	-0.940^c	-0.396	-0.514^c	-1.936^c
		(2.957)	(0.240)	(0.143)	(0.934)	(0.117)	(0.100)
	SD	0.673	0.505^c	0	0	0	1.048^c
		(1.835)	(0.125)	(0)	(0)	(0)	(0.177)
Total volume	Mean	-1.058^c	-1.588^c	-1.128^c	-0.601^c	-0.639^c	-0.537^c
		(0.151)	(0.504)	(0.216)	(0.206)	(0.111)	(0.194)
	SD	0.986^b	1.301^c	1.272^c	0.806^c	0.853^c	0.476^c
		(0.388)	(0.349)	(0.172)	(0.159)	(0.082)	(0.119)
Policy	Mean	-0.767^c	-0.557^c	-2.281^c	-0.127^c	-0.623^c	-1.372
		(0.100)	(0.128)	(0.698)	(0.465)	(0.063)	(3.170)
	SD	0	0	2.029	7.564^c	0	1.225
		(0)	(61.77)	(50.37)	(0.250)	(0)	(3.49)
Sugar	Mean	0.090	0.222^c	0.126^c	0.076	-0.065^c	0.125^c
		(0.414)	(0.075)	(0.018)	(0.066)	(0.011)	(0.015)
	SD	0.192^a	0	0	0	0.243^c	0
		(0.194)	(0)	(0)	(0)	(0.007)	(0.848)
Large store	Mean	2.285^c	1.736^c	2.199^c	2.470^c	1.710^c	2.237^c
		(0.078)	(0.153)	(0.051)	(0.196)	(0.030)	(0.032)
SSB	Mean	-0.492	-1.033^b	-0.212^c	-0.068^c	-0.015^c	-0.471^c
		(1.429)	(0.500)	(0.168)	(0.410)	(0.074)	(0.093)
Pack size	Mean	0.140	0.170^c	0.119^c	0.042	0.033^c	0.105^c
		(0.182)	(0.066)	(0.025)	(0.087)	(0.013)	(0.025)
Post	Mean	0.582	0.763^c	0.422^c	0.296^c	0.253^c	0.462^c
		(0.845)	(0.176)	(0.156)	(0.065)	(0.040)	(0.072)
Summary Statistics							
Mean own-ped		-3.98	-5.58	-4.31	-2.50	-2.82	-2.96
Median own-ped		-4.00	-5.51	-3.58	-2.07	-2.34	-3.07
mc< 0(%)		1.33	1.24	6.96	22.11	18.37	0.837
Mean markup		33.88	25.56	47.30	78.57	71.10	41.22

Standard errors in parentheses: ^{a} $p < 0.10$, ^{b} $p < 0.05$, ^{c} $p < 0.01$

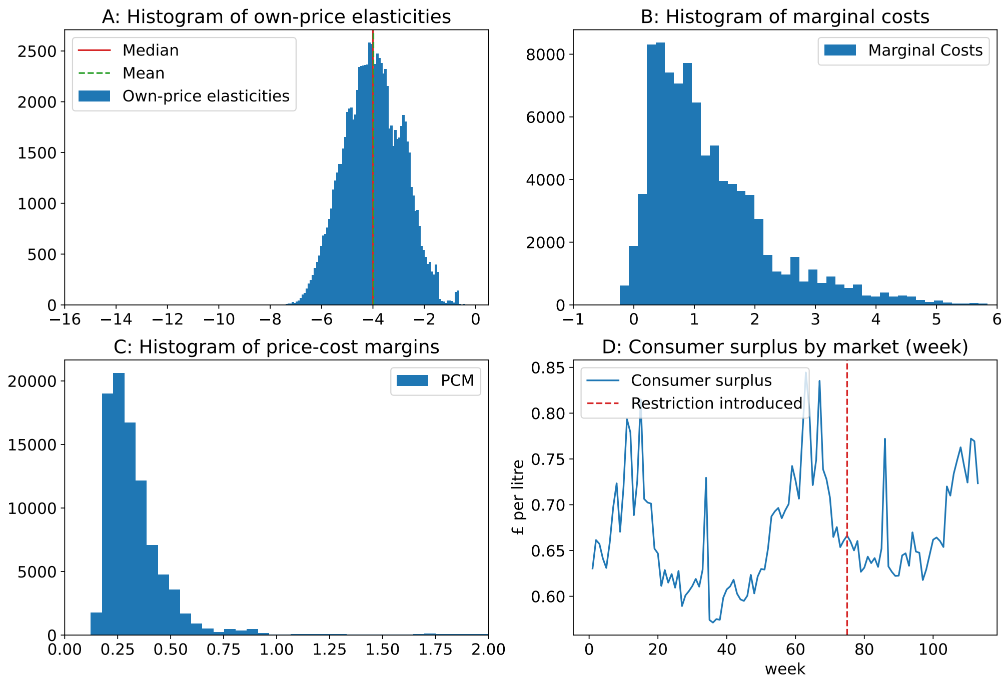
In the summary statistics we show mean and median own-price elasticities all of which are believable levels. Although we generate marginal cost values for each demographic, in reality there are not separate marginal costs for each group. Instead, the markup can change due to second degree price discrimination effects. Therefore, although the retail price should be the same for each demographic group, the transaction price or average price per litre can be different because of factors including use of vouchers/coupons, take up of promotional offers such as '3 for 2' or 'buy one get one free', shopping at different times of the day e.g. some products are discounted at the end of the day or close to expiry dates.

Additionally, there may be third degree price discrimination in the form of student or pensioner discounts. Nevertheless, these separate marginal costs are required when we evaluate consumer and producer surplus (profit) later so we include them in this table for completeness. Since marginal cost greater than zero are high for Older and Younger aggregated groups we focus the remainder of our analysis on aggregation by family type.

4.4.2 Supply side estimates

Using the demand estimates from column 1 Table 4.3, Figure 4.3 illustrates elements of the supply-side on the aggregate sample, as well as elasticities and consumer surpluses. This is a natural progression having estimated the demand side parameters and a necessary step before we properly analyse the effects of the policy on welfare. The aim is to check that the model(s) we have estimated allow for believable marginal costs and price-cost margins. Panel A is a histogram of all own-price elasticities in all markets. The distribution is relatively symmetrical despite the log-normal specification on price, indicated by the fact that median and mean are almost identical, although there is a slight left-skew. In panel B, we see the equivalent histogram for marginal cost. The distribution is now right-skewed, with a mode between £0-£1. There are a small number of products with marginal cost below zero as noted in Table 4.3. Panel C plots the histogram of price-cost margins (PCM) defined as $\frac{p-c}{p}$ for all products in all markets. Where the PCM is greater than 1, it indicates that marginal cost is negative. Again this represents only a small percentage of the sample however, and corresponds with those products in panel B. Finally panel D plots average consumer surplus by market. There certainly appears to be some seasonality in this graph. The red-line marks the week the legislation came into force in October 2022. The upward trend immediately before the red line is therefore summer 2022, an extremely hot year in the UK (MetOffice, 2023) resulting in high sales and increased consumer surplus. The spike immediately after the red line coincides with Christmas and New Year, again a period where we might expect higher consumer surplus. In Figure 4.4 we overlay the consumer surplus estimates according to the particular week of the year to show our predictions follow close cyclical trends, spiking in the summer months and at Christmas/New Year. The post-restriction period is from around week 32 onwards of the red line and the blue line. From this graph we can see that there appear to be small yearly trends of increasing consumer surplus as the line is slightly higher for each year moving from 2021 to 2023.

Figure 4.3: Post-estimation summary for aggregated sample



In Figure 4.5, we plot share-weighted average marginal costs and markups chronologically by market. The two panels are stark in that they show a clear upward (downward) trend for marginal costs (markups) before the restrictions were introduced. Ordinarily, we might be sceptical of such a pronounced upward (downward) trend in marginal costs (markups). However, our sample period coincides with a period of significant global inflationary pressure in most input and product markets (see Figure 4.6 as one illustration). These measures support the idea that marginal costs of SSBs increased significantly over our sample period and markups correspondingly fell before stabilising somewhat.

Figure 4.4: Consumer surplus by year

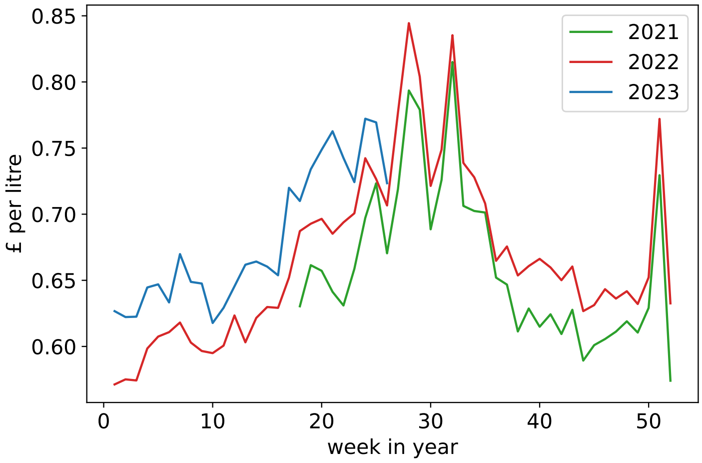
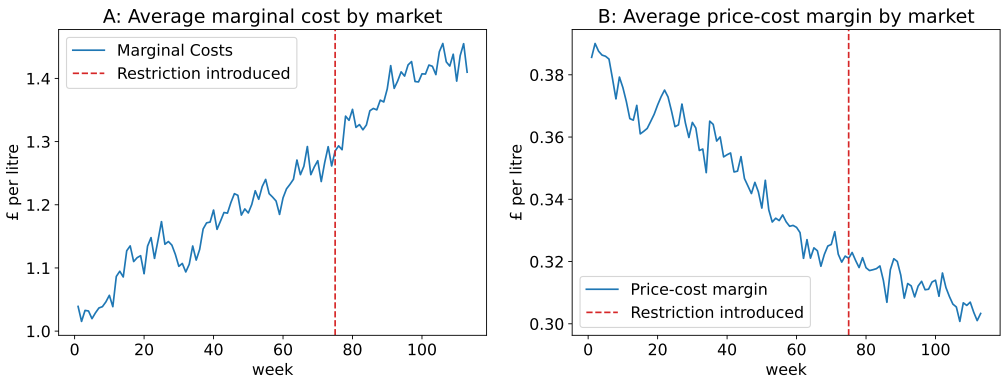


Figure 4.5: Share-weighted average supply-side by week

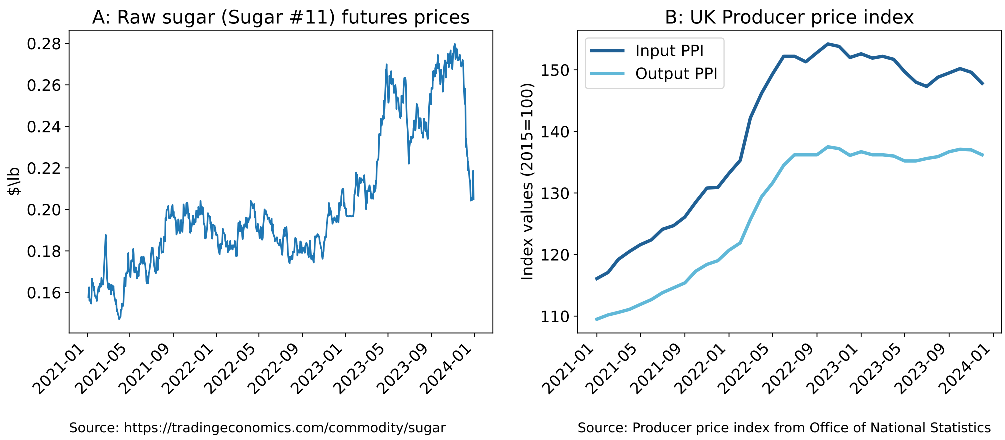


4.4.3 Consumer surplus effects of policy

Having estimated consumer surplus on the aggregated sample as a result of the policy, the next step is to compare this consumer surplus to a scenario where the policy was not introduced. There are several ways of doing this. We present two alternatives here.

The first method involves simulating a counterfactual scenario in which the policy does not exist. To do this, we subset the sample to only the weeks following the introduction of the policy (week 75 onwards). For these weeks, we set the value for 'Policy' in the demand function to zero for all products in all these markets. Then we re-estimate prices and shares, given the demand parameters

Figure 4.6: Evidence of rising input prices



from column 1 of Table 4.3. Finally, we use these new prices and shares to calculate consumer surplus in this no restriction world. This is presented in Figure 4.7. In Panel A, we overlay the simulated no restriction weekly consumer surplus over the previously calculated consumer surplus under the current policy, for week 75 onwards. We see that other than a large spike around week 80 for the no restriction scenario, the two lines generally follow the same trend. Comparing the mean value of weekly consumer surplus for each scenario in Panel B, we see that the no restriction world has slightly higher mean weekly consumer surplus, around £0.01-0.02 which translates to a 2-3% difference. This is not surprising, given that we would expect something that reduces perceived quality of a product to result in a lower consumer surplus.

Figure 4.7: Comparison of mean weekly consumer surplus pre and post restriction

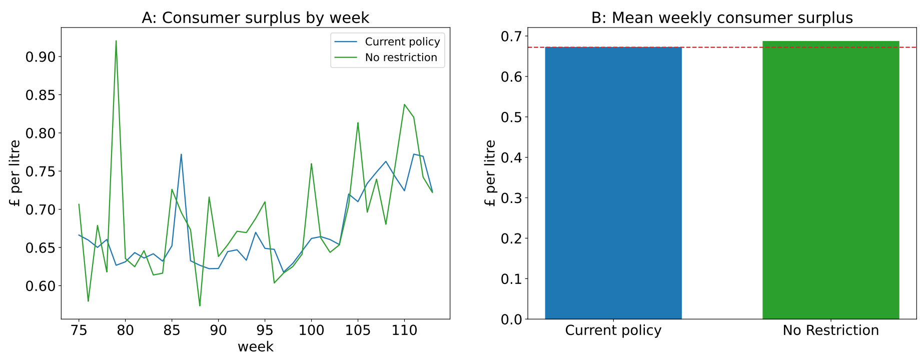
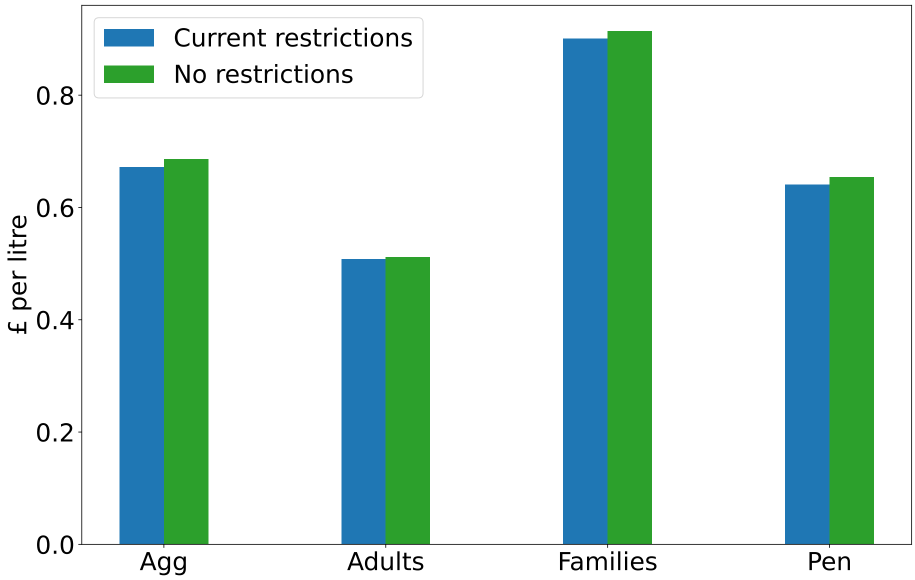


Figure 4.8 extends Panel B from Figure 4.7 to all demographics. Consumer

Figure 4.8: Comparison of mean weekly consumer surplus pre and post restriction for all demographics



welfare for each group is again marginally higher without any restrictions and the effect is broadly the same across groups. Given the heterogeneous effects on types of drinks we found in Chapter 3 and differences in spending habits between these different groups, we might have expected some heterogeneous effects in the changes to consumer welfare. This appears not be the case. Instead the overriding observation is that the overall minimal effect on sales of SSBs from the policy we show in Chapter 3 is accompanied by a fall of consumer welfare of around 1-2% for all demographic groups.

In Figure 4.9 we show our alternative method to compare CS. Using the CS values we previously calculated in Panel D of Figure 4.3, we can calculate the mean weekly CS in the post-policy period as the mean of the values in the weeks after the red line (in fact this calculation is the same as for the blue bar in Figure 4.7, Panel B). To obtain the mean weekly CS in the pre-policy period, we subset the equivalent weeks for which we have post-policy data but exactly 12 months prior, and then calculate the mean weekly CS of these weeks. This ensures that we account for any seasonal effects. Using this method, we see that the post-period in orange is higher than the pre-period. However, we did say that consumer surplus is around £0.05 per litre higher in 2023 than in 2022, from Figure 4.4. If we take this into account, then the difference between the pre and post period goes away and may even reverse, supporting the conclusion of the first method we used to calculate the differences in Figure 4.7.

The remaining bars in Figure 4.9 do the same for each demographic group. For all groups consumer surplus is higher in the post period, which is likely due to the year effects already mentioned. However for families consumer surplus is lower following the restrictions. If we refer back to Table 4.1, we can see that although Families (OF and YF) together make up 34.2% of customers they purchase 42.5% and 46.1% of SSBs. Therefore, any policy that reduces the quality of SSBs is likely to effect them more negatively. Given their disproportionate purchasing it appears the policy is effective in targeting those groups that consume more of these drinks and by effecting Families more this also achieves the governments stated aims of targeting childhood obesity (DHSC, 2020).

Figure 4.9: Comparison of mean weekly consumer surplus pre and post restriction

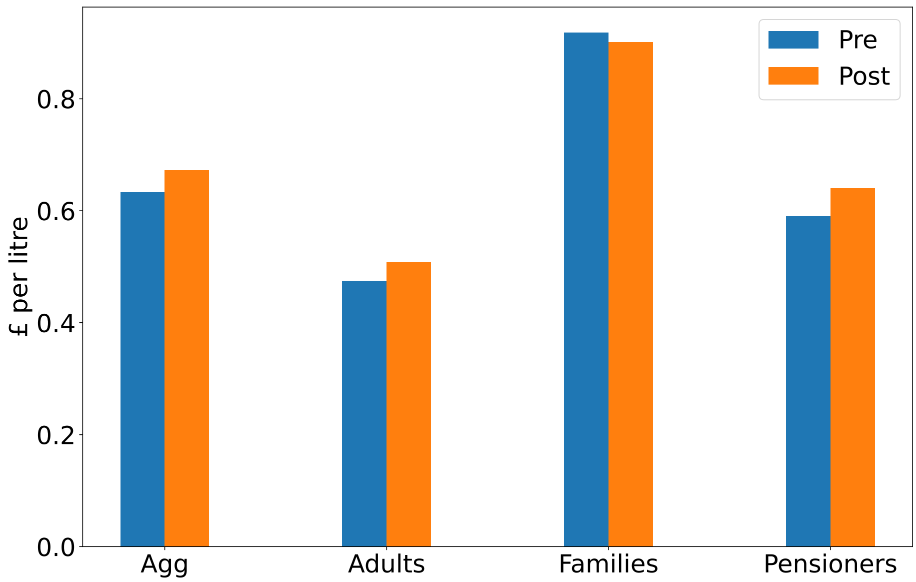
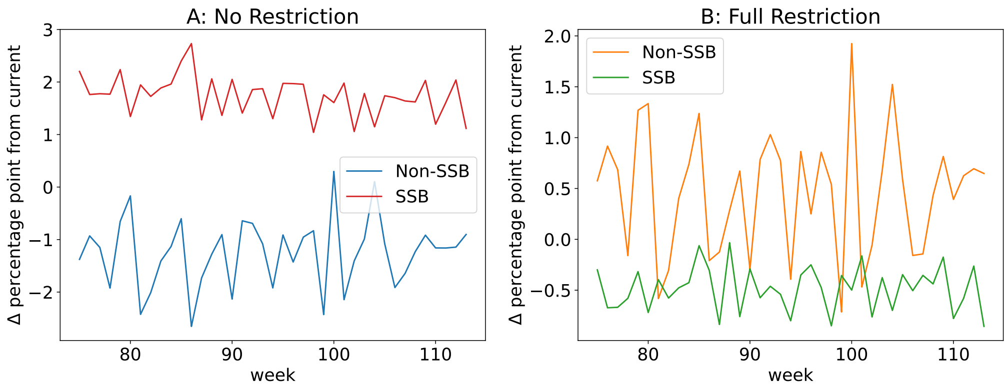


Figure 4.10: Weekly difference in shares from status quo

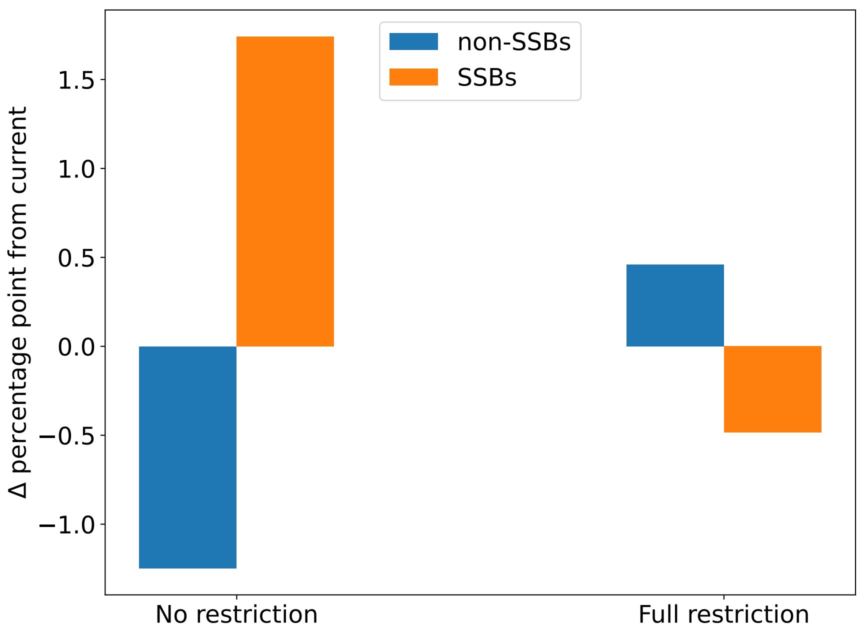


4.4.4 Counterfactual Simulations

In this section we analyse the market outcomes for our no restriction scenario, alongside a full restriction scenario, in which we extend the policy to apply to all stores, against the current policy regime.

In Chapter 3 we concluded that the policy in its current guise was ineffectual in that there was no overall statistically significant change in sales of SSBs. This is because the policy targeted those SSBs purchased for immediate consumption but then exempted the stores in which immediate consumption drinks held the largest share of purchases. In Figure 4.10, we show our predictions of shares under the full restriction and no restriction simulations. Following each simulation, we separate SSBs from non-SSBs and calculate the combined weekly share of each. From these values we subtract the equivalent combined weekly shares from the observed data and plot this difference against weeks. Panel A illustrates a no restriction world. In the absence of any restrictions we see that shares of SSBs, in red, are greater than the current observed shares under the status quo policy. While there are significant fluctuations in the blue non-SSB line, the mean change in shares is negative suggesting that consumers either divert away from non-SSBs into SSBs or the outside-good. In Chapter 3, we found a simple AIDS model showed little cross-elasticity between SSBs and non-SSBs i.e. people who are buying non-sugary drinks are not substituting into buying sugary beverages, so the latter of these is more likely. The opposite is true in the full restriction scenario. Now SSBs see a decline in their combined share when compared to current observed shares, illustrated by the green line in Panel B. Together these graphs suggest that the government could improve the effectiveness of its current policy by extending it to apply to all stores. In Figure 4.A.2 we plot the same data but combine SSBs into one graph in Panel B to better illustrate the difference in the change in shares between no restrictions and full restrictions. We summarise these changes in shares neatly in Figure 4.11. Under no restrictions the mean weekly change in the share of SSBs is around +1.8 percentage points. With full restrictions this falls to -0.5 percentage points.

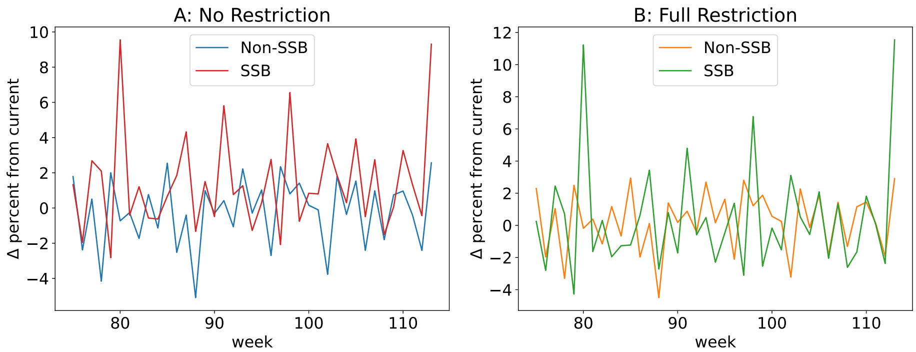
Figure 4.11: Mean weekly difference in shares from status quo



Next, we compare simulated prices against current observed prices. To calculate this we take the mean share-weighted weekly price per litre for SSBs and non-SSBs respectively, and then plot the percentage difference from the equivalent group and week in the observed data.

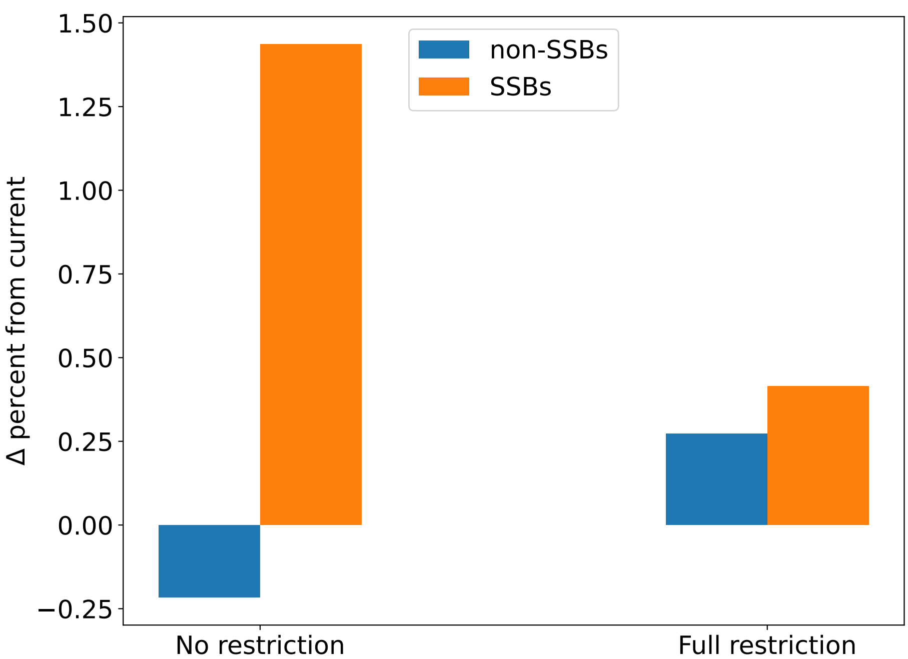
In Panel A of Figure 4.12, we plot the prices for SSBs under the no restrictions in red and for non-SSBs in blue. We can see the price fluctuations are in general higher for SSBs. We believe this occurs because producers do not have to reduce

Figure 4.12: Percentage change in prices from status quo



the price to compensate for a perceived lack of quality. Price changes are more similar under full restrictions. However, the green line representing SSBs does fluctuate more. We can confirm this by plotting the mean weekly difference in prices as we did previously for shares, shown in Figure 4.13. Mean prices for SSBs are around 0.5% higher under full restrictions for the reasons we suggested prior. In a no restriction world, SSBs prices are much higher, while non-SSBs are cheaper as a result of the competition from strong brands within the SSB category.

Figure 4.13: Mean weekly difference in prices from status quo



Finally, we estimate changes in welfare as a result of the simulations in Figures 4.14, 4.15, and 4.16.

Figure 4.14: Consumer surplus

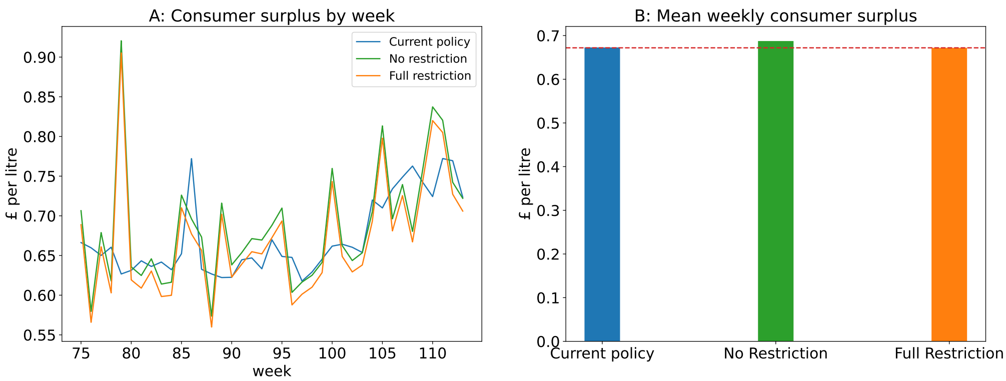


Figure 4.15: Profit

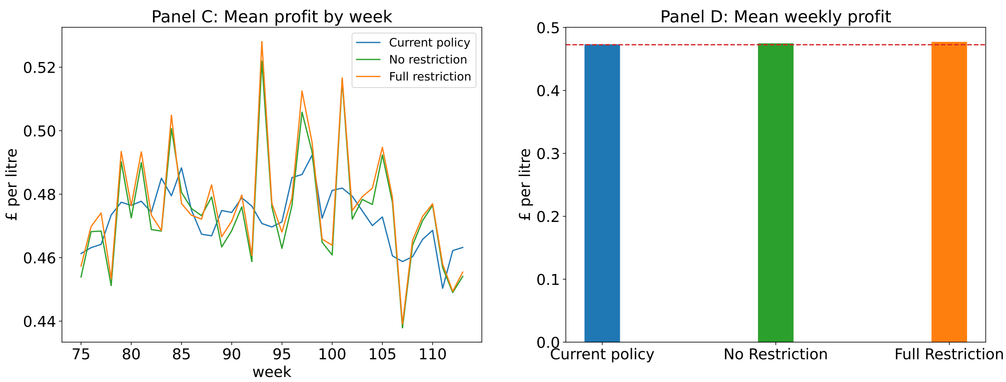
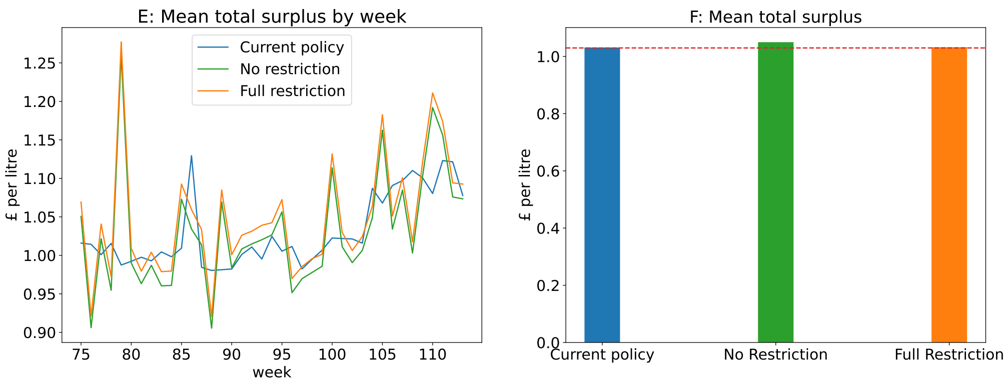


Figure 4.16: Total surplus



For each figure, the left panel plots the mean welfare figure by week for each of the simulations - no restrictions in green and full restrictions in orange - alongside the observed values in blue. In all cases, the simulations track the observed values in a general trend sense, albeit with more week to week volatility. For consumer surplus in panel A, the green line of no restriction is marginally higher in all cases than under full restrictions indicating that consumer surplus is indeed and as expected higher with no restrictions. However, when comparing the mean values under each scenario in panel B, we can see the average differences are marginal, no more than £0.01 per litre or approximately 1.2%. A similar story plays out for profits but now profits. Therefore, when we estimate total surplus as a combination of consumer welfare and profits, the differences in consumer surplus and profits offset each other so that mean total surplus is largely unchanged regardless of the policy environment. If these results hold, this has important implications for the future direction of the location restriction policy. We suggest that the policy should be extended to cover all store types because we ought to observe a greater decline in the share of SSBs sold than in changes in consumer welfare while total welfare remains unchanged.

4.4.5 Demographic comparisons

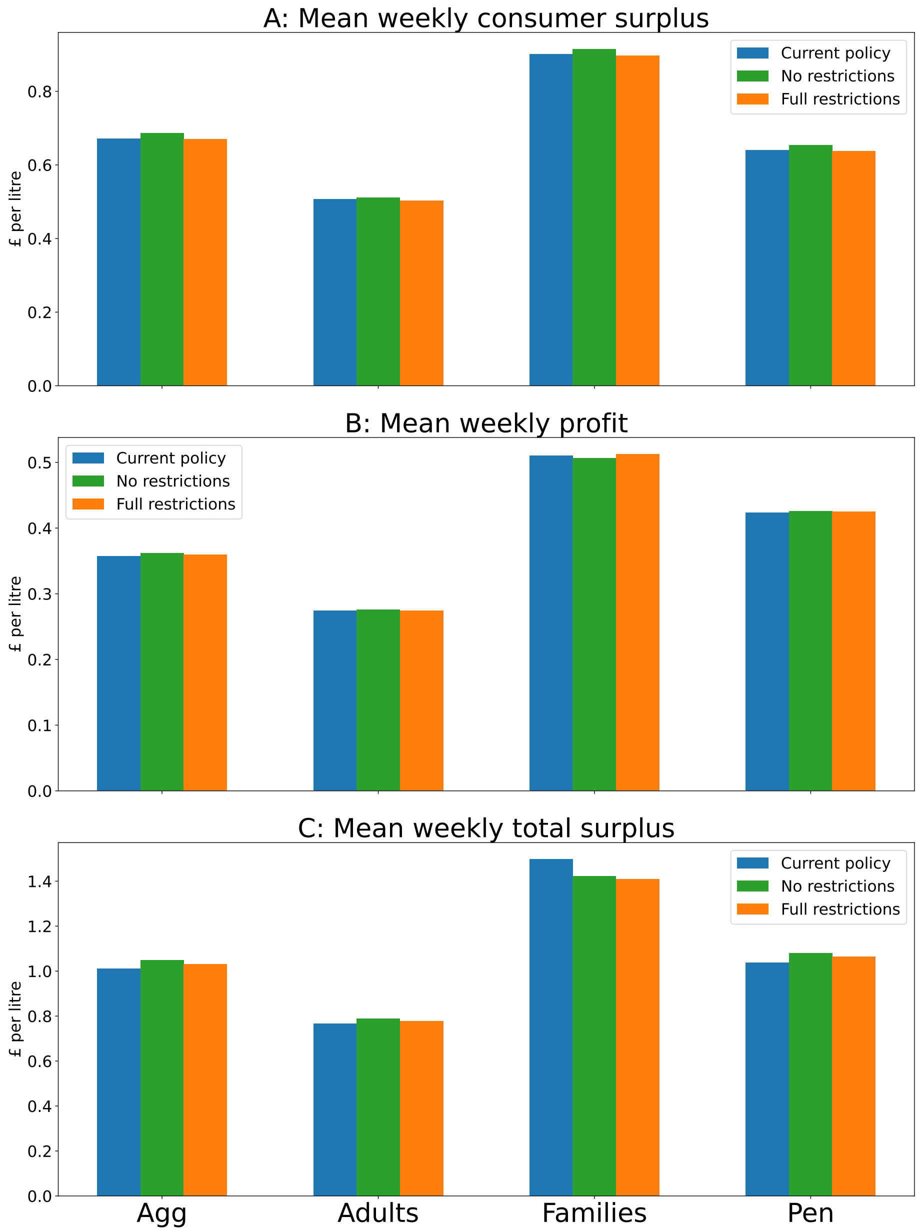
The final part of our analysis involves comparing welfare in a potential full restriction world, with the current policy regime. For completeness we also include results from our no restriction simulation to easily compare across all three possible scenarios. When looking at welfare statistics by demographic in Table 4.4, where the values at £/litre, we see a similar story in that there are no great changes for any one group. Regardless of the policy environment, Adults see almost identical consumer surplus, profits and total surplus. For Families however, the total surplus is reduced. Again this adds support to the notion that the policy is targeting those groups that are more likely to purchase SSBs and extending the policy to cover all store types would reduce sales without damaging consumer surplus overtly. Pensioners follow a similar pattern to Adults in that this is no great change in surpluses whatever the policy environment.

Table 4.4: Welfare comparisons by demographic

	Agg	Adu	Fam	Pen
Consumer Surplus				
Observed data	0.672	0.508	0.901	0.640
No restrictions	0.687	0.511	0.914	0.654
Full restrictions	0.671	0.504	0.897	0.638
Profit				
Observed data	0.358	0.274	0.511	0.423
No restrictions	0.362	0.276	0.507	0.426
Full restrictions	0.360	0.275	0.512	0.425
Total Surplus				
Observed data	1.012	0.767	1.497	1.037
No restrictions	1.048	0.788	1.421	1.080
Full restrictions	1.031	0.778	1.409	1.063

Figure 4.17 is a visual representation of the information presented in Table 4.4 where it is easier to see the largely constant values in surpluses/welfare for most groups and the fall in total surplus for families. Interestingly we can see that total surplus for the Family group is highest during the current policy regime i.e. there is a transfer of welfare from consumers to producers. Of course, this is not useful in achieving the policy's stated aims. At the aggregate level though, we conclude that whatever the policy environment, the policy has little impact on consumer welfare, profits and total welfare.

Figure 4.17: Simulated surpluses by demographic



4.5 Conclusion

In Chapter 3 we concluded that the location restriction policy was not very effective in reducing added sugar consumption for two reasons. Firstly, the introduction of the SDIL had served to greatly reduce the number of SSBs produced and therefore consumed by incentivising producers to reformulate their drinks to reduce the sugar content. Secondly, the location restriction policy itself had exemptions for smaller stores. However, we found a large percentage of SSB purchases were from these stores and thus were not subject to any restrictions. The aim of this paper was to see whether we could improve the effectiveness of the policy by extending the restrictions to all stores. In order to do this, we estimated a structural demand model on the same data as in Chapter 3. Using these estimates we were able to simulate counterfactual scenarios in two extreme alternative worlds; one in which the restrictions were completely removed and another in which they applied to all stores. We found that extending the restrictions did reduce sales of SSBs at the aggregate level by around 0.5%.

We also wanted to observe what would happen primarily to consumer welfare but also total welfare if the restrictions were extended. We find that although there is a reduction in both consumer and total welfare the effect is marginal so we argue that it makes sense for the government to extend the policy to achieve their objectives. However, we temper this recommendation with the following. In our analysis, we have not considered the cost of, or ability to, implement full restrictions. As stated earlier, how does one apply a location restriction when the distinction between front and back of a small store is negligible in terms of the inconvenience it causes? One solution could be to introduce locked fridges or individual lock-boxes for items. Yet this could be prohibitively expensive especially for smaller and independent stores. By reducing profitability we would like observe a greater reduction in total welfare than any potential gains from reduced sugar consumption.

The policy, and indeed, the government's focus on sugar does not reflect potential concerns about the effects of artificial sweeteners. Reformulation of SSBs in a post-SDIL world has seen the proliferation of various artificial sweeteners in popular beverages. There is not enough data to evaluate the long-term health effects of consuming large quantities of these sweeteners. By effectively incentivising manufacturers to shift from the known effects of sugar to the unknown effects of artificial sweeteners, we may be creating further public health issues in the future.

The topic of HFSS in general, however, remains at the forefront of public health economics and we suggest several areas to continue the research into SSBs and related policy. This paper serves demographic analysis by separating the data into graphs and analysing each individually. The advantage of this method is that we obtain a separate coefficient on each parameter for each demographic easily. This is helpful if we believe that preferences differ greatly between demographics and allows us to make quick comparisons. The problem is that when we estimate separate models, we obtain a marginal cost for each model. Of course the marginal cost to the firm does not change depending on which group purchases the product. Ultimately, any subsequent calculations related to total surpluses are effected by this demographic marginal cost. An alternative strategy would be to carry out a joint estimation whereby we stack the demographic sales data on top of each other so that we have a row for each product, in each store type, for each demographic. The disadvantage of this is that preferences for each demographic are less obvious, although we can interact a demographic dummy with characteristics we believe to differ between demographics. The advantage is the model would internalise a fixed marginal cost so that our markup calculations include this correct marginal cost.

Our conclusions are also based on the current UK threshold of 5mg/100ml to be considered an SSB. Some countries (see 4.A.1) have much stricter thresholds already. Further research can be done to observe the effects of reducing the threshold in the UK. If the reaction is similar to the introduction of the SDIL, then as more products become classified as SSBs, manufacturers will change recipes to remove even more sugar from the market. However, we caution policy makers to consider that further increasing the amount of artificial sweeteners is not necessarily positive either. The alternative then is not policy to restrict demand in store but demand before consumers enter the store. Of course, changing cultural tastes and preferences takes much longer but through education into the harmful effects of sugar and limiting advertising to children, we argue the government would be better placed to achieve its goals without creating new problems in the future.

References

Amemiya, T. (1977). "A note on a heteroscedastic model". *Journal of Econometrics* 6(3), pp. 365–370.

Berry, S., J. Levinsohn, and A. Pakes (1995). "Automobile prices in market equilibrium". *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Berry, S. T. (1994). "Estimating discrete-choice models of product differentiation". *The RAND Journal of Economics*, pp. 242–262.

Berry, S. T. and P. A. Haile (2021). "Foundations of demand estimation". *Handbook of industrial organization*. Vol. 4. 1. Elsevier, pp. 1–62.

Campbell, C. A., R. A. Hahn, R. Elder, R. Brewer, S. Chattopadhyay, J. Fielding, T. S. Naimi, T. Toomey, B. Lawrence, J. C. Middleton, et al. (2009). "The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms". *American journal of preventive medicine* 37(6), pp. 556–569.

Chamberlain, G. (1982). "Multivariate regression models for panel data". *Journal of econometrics* 18(1), pp. 5–46.

— (1987). "Asymptotic efficiency in estimation with conditional moment restrictions". *Journal of econometrics* 34(3), pp. 305–334.

Colchero, M. A., B. M. Popkin, J. A. Rivera, and S. W. Ng (2016). "Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study". *British Medical Journal* 352.

Conlon, C. and J. Gortmaker (2020). "Best practices for differentiated products demand estimation with PyBLP". *The RAND Journal of Economics* 51(4), pp. 1108–1161. DOI: https://doi.org/10.1111/1756-2171.12352. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12352. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12352.

Conlon, C., N. Rao, and Y. Wang (2022). "Who pays sin taxes? understanding the overlapping burdens of corrective taxes". *Review of Economics and Statistics*, pp. 1–27.

Conlon, C. and N. L. Rao (2023). *The cost of curbing externalities with market power: Alcohol regulations and tax alternatives*. Tech. rep. National Bureau of Economic Research.

DHSC (July 2020). *Tackling obesity: Empowering adults and children to live healthier lives*. URL: https://www.gov.uk/government/publications/tackling-obesity-government-strategy/tackling-obesity-empowering-adults-and-children-to-live-healthier-lives.

— (Nov. 2022). *New obesity treatments and technology to save the NHS billions*. URL: https://www.gov.uk/government/news/new-obesity-treatments-and-technology-to-save-the-nhs-billions.

— (June 2023). *Restricting promotions of products high in fat, sugar or salt by location and by volume price: implementation guidance*. URL: https://www.gov.uk/government/publications/restricting-promotions-of-products-high-in-fat-sugar-or-salt-by-location-and-by-volume-price/restricting-promotions-of-products-high-in-fat-sugar-or-salt-by-location-and-by-volume-price-implementation-guidance.

DHSC, D. (June 2021). *Introducing further advertising restrictions on TV and online for products high in fat, salt and sugar*. URL: https://www.gov.uk/government/consultations/further-advertising-restrictions-for-products-high-in-fat-salt-and-sugar/outcome/introducing-further-advertising-restrictions-on-tv-and-online-for-products-high-in-fat-salt-and-sugar-government-response).

Dubé, J.-P. (2005). "Product differentiation and mergers in the carbonated soft drink industry". *Journal of Economics & Management Strategy* 14(4), pp. 879–904.

Dubois, P., R. Griffith, and M. O'Connell (2020). "How well targeted are soda taxes?" *American Economic Review* 110(11), pp. 3661–3704.

Fearne, A., N. Borzino, B. De La Iglesia, P. Moffatt, and M. Robbins (2022). "Using supermarket loyalty card data to measure the differential impact of the UK soft drink sugar tax on buyer behaviour". *Journal of Agricultural Economics* 73(2), pp. 321–337.

Gandhi, A. and J.-F. Houde (2019). "Measuring substitution patterns in differentiated-products industries". *NBER working paper* (w26375).

GFRP (2024). *Policy research - Marketing Regulations.* URL: https://www.globalfoodresearchprogram.org/policy-research/marketing-regulations/ (visited on 06/01/2024).

Griffith, R. (2023). *The costs of obesity.* R266. IFS Report.

Griffith, R., M. O'Connell, and K. Smith (2017). *Design of optimal corrective taxes in the alcohol market.* Tech. rep. IFS Working Papers.

Griffith, R., M. O'Connell, K. Smith, and R. Stroud (2020). "What's on the menu? Policies to reduce young people's sugar consumption". *Fiscal Studies* 41(1), pp. 165–197.

Griffith, R., M. O'Connell, and K. Smith (2019). "Tax design in the alcohol market". *Journal of Public Economics* 172, pp. 20–35.

Hinnosaar, M. (2016). "Time inconsistency and alcohol sales restrictions". *European Economic Review* 87, pp. 108–131.

HMRC (Dec. 2016). *Soft Drinks Industry Levy.* URL: https://www.gov.uk/government/publications/soft-drinks-industry-levy/soft-drinks-industry-levy.

Jones, S. (May 2024). *Sainsbury's trialling electronic security cabinets in booze aisles to fight theft.* URL: https://www.thegrocer.co.uk/sainsburys/sainsburys-trialling-electronic-security-cabinets-in-booze-aisles-to-fight-theft/688538.article.

MetOffice (July 2023). *Record breaking 2022 indicative of future UK climate.* URL: https://www.metoffice.gov.uk/about-us/news-and-media/media-centre/weather-and-climate-news/2023/record-breaking-2022-indicative-of-future-uk-climate.

Middleton, J. C., R. A. Hahn, J. L. Kuzara, R. Elder, R. Brewer, S. Chattopadhyay, J. Fielding, T. S. Naimi, T. Toomey, B. Lawrence, et al. (2010). "Effectiveness of policies maintaining or restricting days of alcohol sales on excessive alcohol consumption and related harms". *American journal of preventive medicine* 39(6), pp. 575–589.

Nevo, A. (2000). "A practitioner's guide to estimating random-coefficient logit models of demand". *Journal of Economics and Management Strategy* 9(4), pp. 513–548.

O'Connell, M. and K. Smith (2021). *Optimal sin taxation and market power*. Tech. rep. IFS Working Paper.

O'Mara, J. and I. Vlad (Mar. 2023). *Looking back at 5 years of the UK Soft Drinks Industry Levy*. URL: https://www.wcrf.org/looking-back-at-5-years-of-the-uk-soft-drinks-industry-levy/.

Pedraza, L. S., B. M. Popkin, C. Batis, L. Adair, W. R. Robinson, D. K. Guilkey, and L. S. Taillie (2019). "The caloric and sugar content of beverages purchased at different store-types changed after the sugary drinks taxation in Mexico". *international journal of behavioral nutrition and physical activity* 16, pp. 1–11.

PHE (Oct. 2015). *Sugar reduction: The evidence for action*. URL: https://assets.publishing.service.gov.uk/media/5a7f928c40f0b623026904b7/Sugar_reduction_The_evidence_for_action.pdf.

— (Mar. 2017). *Health matters: Obesity and the Food Environment*. URL: https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2.

— (Oct. 2020). *Sugar reduction: progress report, 2015 to 2019*. URL: https://www.gov.uk/government/publications/sugar-reduction-report-on-progress-between-2015-and-2019.

Reynaert, M. and F. Verboven (2014). "Improving the performance of random coefficients demand models: The role of optimal instruments". *Journal of econometrics* 179(1), pp. 83–98.

Rogers, N. T., D. Pell, O. T. Mytton, T. L. Penney, A. Briggs, S. Cummins, C. Jones, M. Rayner, H. Rutter, P. Scarborough, et al. (2023). "Changes in soft drinks purchased by British households associated with the UK soft drinks industry levy: a controlled interrupted time series analysis". *BMJ open* 13(12), e077059.

Scarborough, P., V. Adhikari, R. A. Harrington, A. Elhussein, A. Briggs, M. Rayner, J. Adams, S. Cummins, T. Penney, and M. White (2020). "Impact of the announcement and implementation of the UK Soft Drinks Industry Levy on sugar content, price, product size and number of available soft drinks in the UK, 2015-19: A controlled interrupted time series analysis". *PLoS medicine* 17(2), e1003025.

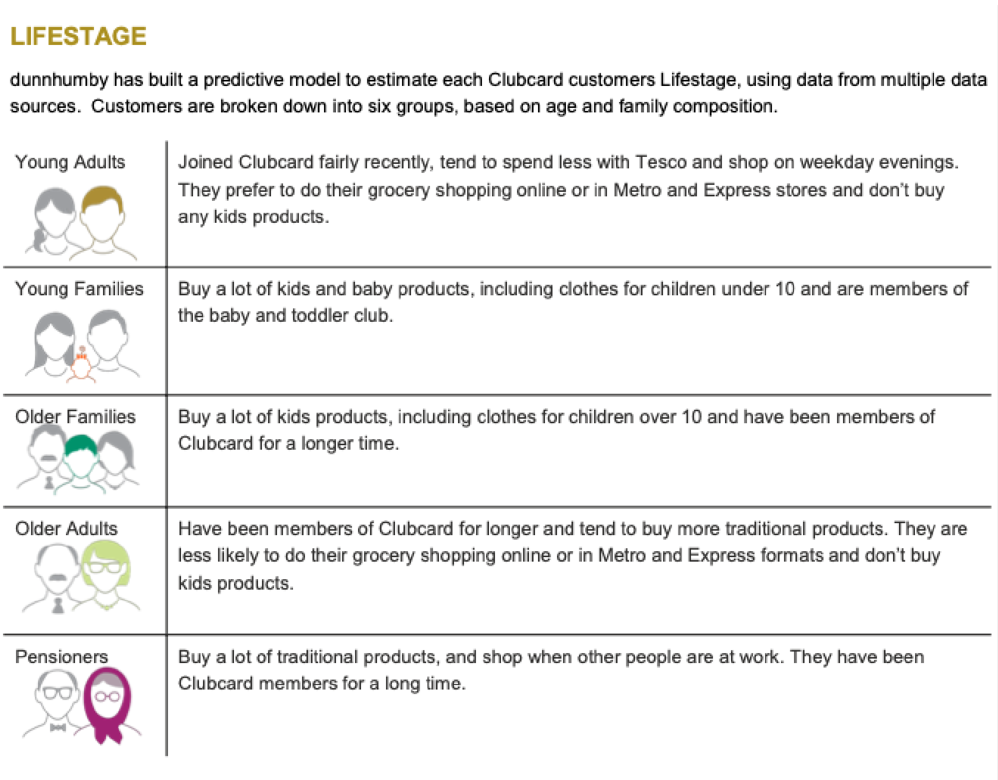
West, T. (2023). *Tesco CCO: Clubcards are now used across 80% of all Tesco transactions*. URL: https://www.marketing-beat.co.uk/2023/07/14/bellini-tesco-clubcard/.

World Health Organization (2015). *Guideline: sugars intake for adults and children*. World Health Organization.

Appendix

4.A Additional Figures

Figure 4.A.1: Dunhumby lifestages



LIFESTAGE

dunnhumby has built a predictive model to estimate each Clubcard customers Lifestage, using data from multiple data sources. Customers are broken down into six groups, based on age and family composition.

Young Adults	Joined Clubcard fairly recently, tend to spend less with Tesco and shop on weekday evenings. They prefer to do their grocery shopping online or in Metro and Express stores and don't buy any kids products.
Young Families	Buy a lot of kids and baby products, including clothes for children under 10 and are members of the baby and toddler club.
Older Families	Buy a lot of kids products, including clothes for children over 10 and have been members of Clubcard for a longer time.
Older Adults	Have been members of Clubcard for longer and tend to buy more traditional products. They are less likely to do their grocery shopping online or in Metro and Express formats and don't buy kids products.
Pensioners	Buy a lot of traditional products, and shop when other people are at work. They have been Clubcard members for a long time.

Table 4.A.1: Sugar tax by country

Country	Introduced	Type	Tax Design	Value†
Belgium	2016	Specific	Volumetric	\$0.081
Chile	2014	Ad valorem	Sugar content	18% >6.25g
				10% <6.25g
Ecuador	2016	Specific/Ad valorem	Sugar content	10% <25g/L
				0.0018/g <25g/L
Finland	2011	Specific	Volumetric	\$0.26
France	2012	Specific	Volumetric	\$0.13/1.5l
	2018	Specific	Volumetric/Sugar content	\$0.24 >11g
Hungary	2011	Specific	Volumetric	\$0.02
India	2017	Goods & service tax		12% SSBs
				28% aerated SSBs
R. Ireland	2018	Specific	Volumetric/Sugar content	\$0.24 >5g
				\$0.36 >8g
Latvia	2004	Specific	Volumetric	\$0.08
Mexico	2014	Specific	Volumetric	\$0.06
Peru	2018	Ad Valorem		12% on <0.5g
				17% on 0.5-6g
				25% on >6g
Philippines	2018	Specific	Volumetric	\$0.11
Poland	2017	Specific	Volumetric/Sugar content	\$0.11 base
				+ \$0.11 per g >5g
Portugal	2017	Specific	Volumetric/Sugar content	\$0.01 <25g/L
				\$0.06 25-50g/L
				\$0.09 50-80g/L
				\$0.21 >80g/L
Qatar	2019	Ad Valorem		50%
Saudi Arabia	2017	Ad Valorem		50%
South Africa	2018	Specific	Sugar content	\$0.001/g
Thailand	2017	Specific/Ad Valorem	Sugar content	\$0.14 >10g
United Kingdom	2018	Specific	Sugar content	\$0.21 5-8g
				\$0.30 >8g

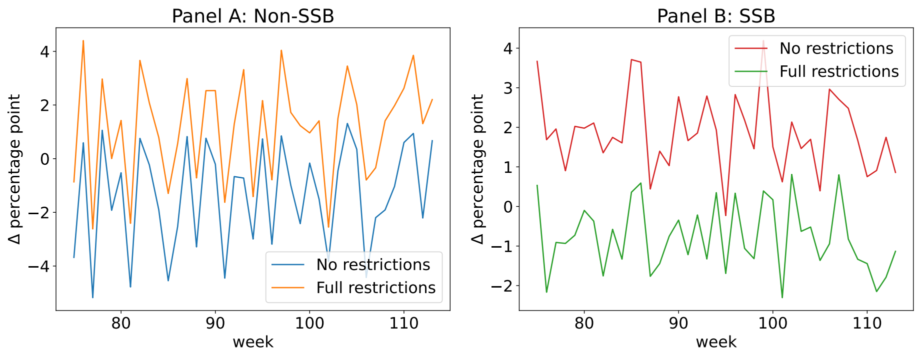
†Value measured in USD (2020 prices) per litre for specific excises or percent on sugar level per 100ml of liquid for ad valorem taxes unless otherwise stated.

Table 4.A.2: Random coefficient demand estimates - raw demographic groups

Age		-	Older	Older	Pensioner	Young	Young
Type		Aggregate	Adults	Families	-	Adults	Families
Variable	Parameter	(1)	(2)	(3)	(4)	(5)	(6)
log Price	Mean	1.637	2.136^c	2.009^c	1.936^c	0.339	0.016
		(2.957)	(0.078)	(0.238)	(0.100)	(0.240)	(0.023)
	SD	0.6725	1.796^c	2.808^c	1.048^c	0	0
		(1.835)	(0.207)	(0.385)	(0.178)	(0)	(0)
Total volume	Mean	-1.058^c	-0.531^c	-0.157	-0.537^c	-0.427^c	-0.192
		(0.151)	(0.104)	(0.082)	(0.194)	(0.115)	(0.113)
	SD	0.986^b	0.618^c	0.335^c	0.476^c	0.646^c	0.412
		(0.3878)	(0.061)	(0.043)	(0.119)	(0.068)	(0.319)
Policy	Mean	-0.767^c	-3.088	-8.347	-1.372	-0.557^c	-0.646^c
		(0.100)	(6.877)	(28.25)	(3.170)	(0.066)	(0.058)
	SD	0	2.838^c	6.758^c	1.225^c	0	0
		(0)	(5.441)	(19.01)	(3.490)	(0)	(32.76)
Sugar	Mean	-0.090	0.103^c	0.046^a	0.125^c	-0.041	-0.273
		(0.414)	(0.016)	(0.023)	(0.015)	(0.049)	(0.326)
	SD	0.192^a	0.117^c	0.143^c	0	0.204^c	0.412
		(0.193)	(0.012)	(0.011)	(0.848)	(0.011)	(0.319)
Large store	Mean	2.285^c	2.133^c	2.495^c	2.237^c	1.250^c	1.848^c
		(0.078)	(0.027)	(0.025)	(0.032)	(0.048)	(0.030)
SSB	Mean	-0.492	-0.531^c	-0.296	-0.471	-0.099	0.356^c
		(1.429)	(0.104)	(0.158)	(0.093)	(0.137)	(0.081)
Pack size	Mean	0.140	0.085^c	0.059^c	0.105^c	0.002	0.006
		(0.182)	(0.017)	(0.017)	(0.025)	(0.021)	(0.014)
Post	Mean	0.582	0.354^c	0.054	0.462^c	0.173^c	0.165^c
		(0.845)	(0.080)	(0.200)	(0.072)	(0.066)	(0.045)
Summary Statistics							
Mean own-ped		-3.98	-2.44	-1.41	-2.96	-2.36	-1.75
Median own-ped		-4.00	-2.41	-1.26	-3.07	-1.97	-1.42
mc<0(%)		1.33	0.40	39.17	0.84	24.46	40.75
Mean markup		33.88	50.95	55.2	41.22	83.33	111.7

Standard errors in parentheses: ^{a} $p < 0.10$, ^{b} $p < 0.05$, ^{c} $p < 0.01$

Figure 4.A.2: Weekly difference in shares from status quo



4.B Instruments

4.B.1 BLP instruments

Traditional 'sums of characteristics' Berry et al. (1995) instruments, BLP are

$$Z^{\text{BLP}}(X) = [Z^{\text{BLP, Other}}(X), Z^{\text{BLP, Rival}}(X)], \tag{4.B.1}$$

where X is a matrix of product characteristics, $Z^{\text{BLP, Other}}(X)$ is a matrix that consists of sums of product characteristics over goods produced by the same firm (i.e. non-rival goods). Finally, $Z^{\text{BLP, Rival}}(X)$ is matrix that consists of sums of product characteristics over goods produced by rival firms. All three matrices have the same dimensions (Conlon and Gortmaker, 2020).

4.B.2 GH instruments

Differentiation instruments, GH, of Gandhi and Houde (2019) are

$$Z^{\text{GH}}(X) = [Z^{\text{GH, Other}}(X), Z^{\text{GH, Rival}}(X)], \tag{4.B.2}$$

where X is again a matrix of product characteristics. $Z^{\text{GH, Other}}(X)$ is now the sums over function of differences between non-rival goods and $Z^{\text{GH, Rival}}(X)$ is the sum over the same function of difference between rival goods.

There are several ways to calculate this function of differences. We use the 'local' version of $Z^{\text{GH}}(X)$. If x_{jkt} in X is characteristic k in market t of product $j \in J_{ft}$ where f is firm, then

$$Z_{jkt}^{\text{Local, Other}}(X) = \sum_{i \in J_{ft} \setminus j} 1(|d_{ijtk}|, \text{SD}_k), \tag{4.B.3}$$

$$Z_{jkt}^{\text{Local, Rival}}(X) = \sum_{i \notin J_{ft}} 1(|d_{ijtk}|, \text{SD}_k), \tag{4.B.4}$$

where $d_{ijtk} = x_{itk} - x_{jtk}$ is the difference between products i and j over characteristic k, and SD_k is the standard deviation of these pairwise differences computed across all markets. $1(|d_{ijtk}|, \text{SD}_k)$ indicates that products i and j are similar in terms of characteristic k (Conlon and Gortmaker, 2020). The intuition is that demand for good j is most heavily influenced by a small number of other goods that have similar characteristics (Gandhi and Houde, 2019).

4.B.3 Optimal Instruments

Reynaert and Verboven (2014) showed that optimal instruments reduce bias, improve efficiency and enhance stability of Berry et al. (1995) instruments. As defined by Chamberlain (1987), optimal instruments are

$$Z_{jt}^{\text{opt}} = \sum_{\xi}^{-1} E\left[\frac{\partial \xi_{jt}}{\partial \theta}\middle| Z\right], \tag{4.B.5}$$

where Z are all exogenous variables.

PyBLP estimates feasible optimal instruments by evaluating this expression at an estimated value of $\hat{\theta}$. The expectation is taken by approximating an integral over the density of ξ. For each error term realization, if not already estimated, equilibrium prices and shares are computed by iterating over the ζ-markup contraction, defined by Morrow and Skerlos (2011), as an alternative to the standard markup equation (Conlon and Gortmaker, 2020).

Table 4.B.1: Variation in instruments

Variable	Mean	sd^2_{O}.	sd^2_{B}.	sd^2_{W}	Min	Max
BLP Instruments						
Sum across other products by produced by own firm						
Count	133.58	97.44	98.04	4.63	0	278
Sugar (mg)	543.61	438.22	436.72	17.46	0	1106.2
Sweetener (dummy)	76.52	54.82	55.21	3.47	0	169
Total volume (litres)	220.98	160.74	162.05	7.74	0	463.65
Pack size (units)	375.81	263.17	259.16	15.78	0	816
GH Instruments						
Sum across other products by produced by own firm						
Sugar (mg)	87.04	61.69	42.43	3.97	0	272
Sweetener (dummy)	74.12	46.14	33.32	3.37	0	168
Total Volume (litres)	116.08	91.54	62.75	5.5	0	267
Pack size (units)	111.28	96.13	67.79	6.15	0	278
Sum across other products by produced by own firm						
Sugar (mg)	435.93	150.01	51.46	9.45	134	780
Sweetener (dummy)	340.26	69.63	34.08	7.3	220	489
Total volume (litres)	565.31	172.57	74.89	12.72	0	788
Pack size (units)	516.65	182.31	76.92	13.78	6	810