How well targeted are soda taxes?

Pierre Dubois, Rachel Griffith and Martin O'Connell*

May 8, 2017

[PRELIMINARY AND INCOMPLETE; PLEASE DO NOT CITE WITHOUT AUTHORS' PERMISSION]

Abstract

Soda taxes aim to reduce externalities (including costs to the consumer themselves in the future) from excess sugar consumption. Their effectiveness depends on how consumers respond, and crucially how demand responsiveness correlates with marginal harm. We estimate demand using novel longitudinal data, which allows us to identify individual preference parameters. We study demand for drinks on-the-go, i.e. for immediate consumption. We show that heavy sugar consumers are the least willing to switch away from sugar in drinks, and that sugar taxes are regressive.

Keywords: preference heterogeneity, discrete choice demand, sugar tax

JEL classification: D12, H31, I18

Acknowledgments: The authors gratefully acknowledge financial support from the European Research Council (ERC) under ERC-2015-AdG-694822, the Economic and Social Research Council (ESRC) under the Centre for the Microeconomic Analysis of Public Policy (CPP), grant number RES-544-28-0001, the Open Research Area (ORA) grant number ES/I012222/1, ANR-10-ORAR-009-01 and ES/N011562/1 and the British Academy under pf160093. Data supplied by TNS UK Limited. The use of TNS UK Ltd. data in this work does not imply the endorsement of TNS UK Ltd. in relation to the interpretation or analysis of the data. All errors and omissions remained the responsibility of the authors.

*Correspondence: Dubois: Toulouse School of Economics, pierre.dubois@tse-fr.eu; Griffith: Institute for Fiscal Studies and University of Manchester, rgriffith@ifs.org.uk; O'Connell: Institute for Fiscal Studies, martin_o@ifs.org.uk

1 Introduction

Corrective taxes have long been seen as a tool to improve social welfare when consumption imposes costs on others in the form of externalities (Pigou (1920)); taxes on fuel, alcohol, tobacco, gambling are examples. More recently corrective taxes have been advocated to reduce consumption that imposes costs on your future self in the form of "internalities" (Gruber and Koszegi (2004), O'Donoghue and Rabin (2006), Haavio and Kotakorpi (2011), Allcott et al. (2014)). Soda taxes are a leading example. Key to evaluating the effects of such taxes in practice is knowing how the shape of demand varies across individuals, and how this relates to the internalities and externalities associated with their consumption. An effective tax will lead to a greater reduction in demand amongst those consumers with the highest marginal harm from consumption.

Our contribution in this paper is twofold. First, we estimate demand for drinks purchases that are made on-the-go (i.e. for immediate consumption outside of the home); this represents an important and relatively understudied part of the market. Second, we exploit longitudinal data to estimate individual specific preference parameters over price, soda and sugar. These can be used to directly assess how well targeted a tax is by relating individual consumers' responsiveness to the tax to other characteristics that proxy for marginal harm. We can also study the distributional consequences of the tax.

We estimate demand for drinks purchased on-the-go, i.e. for consumption outside of the home. Consumption of soda outside the home is common – for instance, in the US around half of sugar-sweetened beverages are consumed outside the home (Han and Powell (2013)) – and yet there are few studies of this important part of the market. Behaviour in this segment of the market is also of interest because it is where we might think consumers are most tempted and so most likely to make choices with the highest marginal harm. Conveniently, estimating demand in the on-the-go market also avoids the well known issues of consumer stockpiling (see, for example Hendel and Nevo (2006) and Wang (2015)) and intra-household allocation (see Browning and Chiappori (1998)). An advantage of our data is that they contain many repeated observations for each individual, which enables us to estimate consumer level parameters. We have information on the on-the-go purchases of over 5000 UK individuals, along with information on their grocery purchases for home from the Kantar Worldpanel.[1] The on-the-go survey was introduced in 2009 and to

[1]The Worldpanel data is used in Dubois et al. (2014) and is similar to the AC Nielsen data used for instance in Aguiar and Hurst (2007) and many other papers.

our knowledge is one of the few data sources available to study consumer behaviour at the individual level using data on the on-the-go segment of the market.

We estimate consumer specific preference parameters over important product attributes – price, soda and sugar. We can relate the parameters to demographic and other (potentially endogenous) features of consumer behaviour. This allows us, for example, to relate the preference parameters and the outcomes of counterfactual simulations to proxies for the marginal externalities and internalities a consumer generates, such as a measure of how much sugar that specific consumer has in their total diet.

Our estimation approach contrasts with the approach typically taken to allow for consumer level preference heterogeneity in choice models. The standard approach is to treat the individual specific preferences as random draws from a mixing distribution, independent of other variables included in the model, so that the resulting choice model is a mixed (or random coefficient) logit. These random draws are then integrated out and cannot be related to information about consumers outside of the model. Often papers will allow the mean of the random coefficient distributions to shift with some demographics. However, this only allows for very restrictive dependence of random coefficients on other consumer information, it requires ex ante knowledge of which variables interact with the random coefficients and the interacting variables must be exogenous (both independent of the demand shocks and random coefficient draws). Our approach avoids all of these limitations.

Taxes targeted at reducing soda consumption have been implemented in a number of locations, including Berkeley, Philadelphia and a number of other US cities, in France, Mexico and the UK and are planned for in many more. The World Health Organisation (WHO (2015)) has urged countries to tax sugary drinks to reduce sugar consumption, especially in children, due to the growing body of evidence that sugar is over consumed, and that this contributes to rising obesity, type 2 diabetes, heart disease, cancers, and other diseases. There is also evidence that excess sugar consumption is associated with poor mental health and poor school performance in children, and poor childhood nutrition is thought to be an important determinant of later life health, social and economic outcomes and of presistent inequality. The Centre for Disease Control and Prevention in the US has highlighted the consumption of sugar-sweetened drinks as a key area of public health concern (CDC (2016)), based on evidence that soda and other sugar-sweetened drinks are a major contributor to added sugars in diets, particularly for young people.[2]

[2]For the US see Han and Powell (2013), Welsh et al. (2011), andWoodward-Lopez et al. (2010) Figure 2); for the UK see Griffith et al. (2016).

We use our demand estimates to explore the implications of introducing a soda tax. We compare the two forms of tax that have recently been introduced – a volumetric tax levied on all soda and a tax applied only to soda that contains added sugar. We assess the effects of these taxes on sugar consumption, how well they target consumers with the highest marginal harm, and how the burden of the tax is distributed with income. Crucial to this analysis is our ability to flexibly relate consumer preferences to other aspects of their behaviour.

We find that the joint distribution of preferences over product attributes departs from the independent normal distribution commonly imposed in applied analysis using random coefficient models. We find important correlations between drink preference parameters and other aspects of consumer behaviour separate from the model. Consumers with a high share of added sugar in their total annual grocery baskets tend to have strong preferences for sugar when purchasing drinks on-the-go and tend also to be relatively insensitive to price. Relatively poor consumers (with low equivalised total grocery expenditure over the year) tend to have strong preferences for soda and tend to be sensitive to price. These correlations have important implications for the effect of soda taxes.

The sugary soda tax achieves larger reductions in sugar consumption across the distribution of total added sugar in diet, compared a soda tax with the same tax rate, because it leads to great switching from sugary to diet soda. However, those consumers with high added sugar across their entire (annual) shopping basket have particularly strong estimated sugar preferences for on-the-go drinks. They are therefore less willing to switch to diet alternatives than consumers with less added sugar in their diets. A consequence of this is that the sugary soda tax (to a greater extent than the soda tax) fails to achieve larger percentage reductions in sugar consumption among people that consume the most sugar (relative to more moderate sugar consumers). We also show that the welfare burden of soda taxes is concentrated on the poorest consumers.

Our work is related to a large literature that highlights the importance of allowing for consumer specific preference heterogeneity in consumer demand models. A number of papers (see, for instance, Berry et al. (1995), Nevo (2001) and Berry et al. (2004)) show that incorporating parametric random coefficients into logit choice models is important for enabling them to capture realistic aggregate switching patterns. Lewbel and Pendakur (2017) show similar results apply in nonlinear continuous choice models, with the incorporation of random coefficients resulting in their model much more effectively capturing the distributional impacts of taxation. A contribution of our work relative to this literature is that we not only incorporate consumer specific preferences but we capture arbitrary correlation between it any other information on the consumers.

A few papers have developed non-parametric methods that relax parametric restrictions on random coefficients, while maintaining the assumption of the independence of random coefficients from other variables in or outside the choice model. For instance Burda et al. (2008) exploit Bayesian Markov Chain Monte Carlo techniques and Train (2008) uses an expectation-maximization algorithm to estimate the random coefficient distribution. Train (2008) applies the method either with a discrete random coefficient distribution or with mixtures of normals. Bajari et al. (2007) discretize the random coefficient distribution and use linear estimation techniques to estimate the frequency of consumers at each fixed point of the preference distribution. Our approach allows the entire joint distribution of parameters to vary across consumers, and it allows us to assess the relationship between consumer preferences and consumer attributes that are likely to be endogenous to soft drink choice. For instance, we can examine whether consumers with strong preferences for sugary soda are also observed purchasing relatively large quantities of sugary products in other markets.

Our approach relies on the large time (T) and cross-sectional (N) dimension in our data. However, our estimates may be subject to an incidental parameter problem that is common in non linear panel data estimation. Even if both $N \to \infty$ and $T \to \infty$ an asymptotic bias may remain, although it shrinks as the sample size rises (Arellano and Hahn (2007)). To deal with this we employ the split sample jackknife bias correction procedure suggested in Dhaene and Jochmans (2015), showing our conclusions are robust to this correction.

Our work is also related to a number of papers that estimate the effects of soda and broader nutrient taxes. Wang (2015) uses household scanner data to estimate the impact of soda taxes on consumer welfare. She specifies a model of dynamic demand that explicitly accounts for consumer stockpiling and shows estimates of the impact of soda taxes that ignore stockpiling behaviour when it is present overestimate the effectiveness of the taxes. We use data on purchases of soda for instantaneous consumption, obviating the need to specify a structural model of stockpiling behaviour. Interestingly, our estimate of the own price elasticity of sugary soda demand is similar to that in Wang (2015).

Bonnet and Réquillart (2013) study taxation in the French soda market. They model demand using a random coefficient logit model, and include non-correlated normally distributed random coefficients on the price and sugar attributes of products. They use their demand estimates to consider supply-side behaviour when taxes are introduced and place less focus than us on assessing consumer level heterogeneity in response to price changes, beyond allowing the mean of their random coefficients to shift with whether any people in the purchasing household are classified as overweight. They show that tax pass-through is incomplete with ad valorem taxes but more than one with excise taxes. Their prediction of over-shifting of the excise tax contrasts with evidence of the Berkeley tax on sugar sweetened beverages, which based on a difference in difference approach with San Francisco as the control, suggests less than half of the tax was passed through to consumers (Cawley and Frisvold (2016)).

Harding and Lovenheim (2014) estimate a continuous demand model over nutritional clusters (aggregates of products based on nutrient content). They use this to estimate switching across grocery products in response to a range of taxes on soda, sugar-sweetened beverages, and other processed food and on nutrients such as fat, salt, and sugar. Their main finding is taxes levied on nutrients are more effective at changing diet than product specific taxes. In our main specification we model the possibility that consumers will respond to soda taxes by switching to alternative non-taxed drinks. We also consider the possibility consumers switch to other forms of sugar (e.g. chocolate). We find evidence that switching to sugar in food is a much less important margin of response than switching to non-taxed alternative drinks such as fruit juice.

The rest of this paper is structured as follows. In Section 2 we discuss the design of sugar taxes and provide evidence that individuals that over-consume sugar typically get a high share of their sugar from soda. In Section 3 we discuss our data and model of demand. Section 4 presents our estimation results. Section 5 presents the results of soda tax simulations and a final section concludes.

2 The Effects of a Tax on Soda

In this section we consider the effects of a soda tax using a simple welfare criterion; this helps to organise the discussion of our empirical results below and highlights the key forces at play. The rationale for such a tax is that excess sugar consumption is associated with externalities, in the form of public costs of funding healthcare systems, and internalities, in the form of unanticipated future health and well-being costs to individuals themselves. This simple framework clarifies that for a sugar tax to be effective it should target the consumption of those with the largest marginal externalities and internalities, and that such a measure will be most effective when the tax leads to larger reductions in sugar amongst those with the large externalities and internalities, while leaving consumers with smaller externalities and internalities relatively undistorted. Additionally, if the burden of the tax falls on poorer consumers the tax might also have undesirable redistributive properties. This motivates the need to obtain demand estimates that can be related to individual characteristics that are proxies for the marginal externalities and internalities generated by consumption and the marginal utility of income.

We document the extent of over-consumption of sugar, relative to government guidelines, and show that those that consume sugar in excess tend to buy relatively large quantities of sugary soda. We show descriptive evidence that a tax levied on the sugar in soda looks potentially promising, since sugar in soda accounts for a large share of sugar consumption by consumers likely to have the highest marginal harm. We use two data source for the descriptive evidence. First, we use a sample of 36,189 adults and children from the National Health and Nutrition Examination Study (NHANES) over 2007-2014. Second, we use a sample of 3073 adults and children in the National Diet and Nutrition Survey (NDNS) over 2008-2011. Both surveys combine interviews and physical examinations to assess the health and nutritional status of participates.

2.1 A tax on the sugar in soda

Let $i \in \{1, ..., N\}$ index consumers, each with income y_i and let $j = \{1, ..., j', j' + 1, ..., J\} \in \Omega$ index food and drink products. Products $j \in \{1, ..., j'\} = \Omega_w$ are sodas and products $j \in \{j' + 1, ..., J\} = \Omega_{nw}$ are non soda products that contain sugar. Products are available at post tax prices $\mathbf{p} = (p_1, \ldots, p_J)'$; each product contains z_j sugar. We consider a tax, τ, levied on the sugar in soda. Suppose consumers have indirect decision utility functions given by $v_i(\mathbf{p}, y_i)$. The consumer's demand for product j is given by $q_{ij}(\mathbf{p}, y_i) = -\frac{\partial v_i / \partial p_j}{\partial v_i / \partial y_i}$. $v_i(\mathbf{p}, y_i)$ governs the choice the consumer makes over which food and drink products to purchase. However it may not reflect the consumer's long run welfare.

In particular, sugar consumption may give rise to future costs that consumers do not take account of at the point of consumption. Much of these costs will be internalities, like future health costs that the consumer may underweight at the point of consumption, although they may also include externalities such as the public health care costs of treating diet related disease. We refer to both of these as internalities for ease of exposition (and since there is evidence that internalities are likely to be particularly important with respect to sugar consumption).

Denote the total sugar in a consumer's diet $\mathcal{S}_i(\mathbf{p}, y_i) = \sum_{j \in \Omega} z_j q_{ij}(\mathbf{p}, y_i)$. Also denote the total sugar from soda and non-soda products in the consumers diet by

$\mathcal{S}_i^w(\mathbf{p}, y_i)$ and $\mathcal{S}_i^{nw}(\mathbf{p}, y_i)$, where $\mathcal{S}_i(\mathbf{p}, y_i) = \mathcal{S}_i^w(\mathbf{p}, y_i) + \mathcal{S}_i^{nw}(\mathbf{p}, y_i)$. Suppose the internality from a consumer's sugar consumption is given by the positive, convex function $\phi_i(\mathcal{S}_i(\mathbf{p}, y_i))$. Consumers ignore these internality costs when making their choices. Tax policy has the potential to improve welfare by inducing consumers to internalise these costs. However, whether tax policy can indeed improve welfare will depend on how successfully it averts internalities by lowering sugar consumption of those prone to suffer them, how much it distorts the behaviour of those that do not suffer from internalities and to what extent it has undesirable distributional consequences.

Consider a utilitarian social welfare function, which is a function of $v_i(\mathbf{p}, y_i) - \phi_i(\mathcal{S}_i(\mathbf{p}, y_i))$, thereby taking account of consumers' long run welfare. Given a soda tax rate τ, the after tax prices of soda products ($j \in \Omega_w$) are $p_j = \tilde{p}_j + \tau z_j$ while for non-soda products ($j \in \Omega_{nw}$) they are $p_j = \tilde{p}_j$, where $\tilde{p}_j$ denotes the pre-tax price. Denote by r_i a rebate that consumer i gets from the tax revenue raised through the soda tax.

How efficiency considerations balance with redistributive effects of the tax will depend on the nature of the tax rebate r_i. Suppose consumer i gets a rebate $r_i = \beta_i \tau \sum_i S_i^w$ where $\beta_i \geq 0$ and $\sum_i \beta_i \leq 1$, meaning tax revenue is redistributed back to consumers with β_i determining the share of revenue consumer i receives.

With a welfare function equal to

$$W = \sum\nolimits_i \left[v_i(\mathbf{p}, y_i + r_i) - \phi_i(S_i(\mathbf{p}, y_i + r_i)) \right], \tag{2.1}$$

the effect of a marginal change in the soda tax on welfare is:

$$\frac{dW}{d\tau} = \underbrace{\sum\nolimits_i (\phi_i' - \tau\bar{\lambda})|S_i'^w|}_{\text{direct efficiency}} - \underbrace{\sum\nolimits_i \phi_i' S_i'^{nw}}_{\text{indirect efficiency}} - \underbrace{\sum\nolimits_i (\lambda_i - \bar{\lambda}) S_i^w}_{\text{redistribution}} \tag{2.2}$$

where λ_i denotes the marginal (decision) utility of income of consumer i , $\bar{\lambda}$ is the weighted average marginal utility of income, $\bar{\lambda} = \sum_i \beta_i \lambda_i$, $\phi_i' \equiv \phi_i'(\mathcal{S}_i(\mathbf{p}, y + r_i))$ is the marginal internality of consumer i and $S_i' \equiv \sum_{j \in \Omega} z_j \frac{dq_{ij}}{d\tau}$ denotes the impact of a marginal change in the tax rate on the consumer's sugar demand.[3]

The effect of tax on welfare depends on the sum of three intuitive terms.

The first term is the direct efficiency effect of the tax. For consumers with a marginal internality that exceeds the tax rate (converted into utils by multiplication of the average marginal utility of income) this term is positive (if the tax on the sugar

[3] In general S_i' depends on tax rate both through dependency of prices on the tax and the impact the tax has on the rebate; $S_i' = \sum_{j \in \Omega} z_j \frac{dq_{ij}}{d\tau} = \sum_{j \in \Omega} z_j \left(\sum_{k \in \Omega_w} \frac{\partial q_{ij}}{\partial p_k} z_k + \frac{\partial q_{ij}}{\partial r_i} \frac{\partial r_i}{\partial \tau} \right)$.

in soda weakly lowers demand for sugar in soda $S_i^{\prime w} \leq 0$). For these consumers the reduction in sugar from soda that results from an increased tax rate leads, through this channel, to a welfare gain. The size of this gain is proportional to how response the consumer's demand for sugar in soda is to the tax instrument. Conversely, for consumers with marginal internalities below the tax rate this term is negative.

The second term is an indirect efficiency effect associated with how taxing the sugar in soda affects demand for sugar in non soda products. If $S_i^{\prime nw} > 0$, so taxing the sugar in soda increases demand for untaxed sugar, the indirect efficiency effect will be negative. If those with large marginal internalities strongly switch to other forms of sugar this inefficiency cost from only taxing a subset of sugar will be large.

The final term reflects redistributive concerns. If those consumers with high marginal utility of incomes tend to have high demands for sugary soda products any tax will be incident on the group the planner would most like to redistribute towards. In this case the third term would act to reduce welfare.

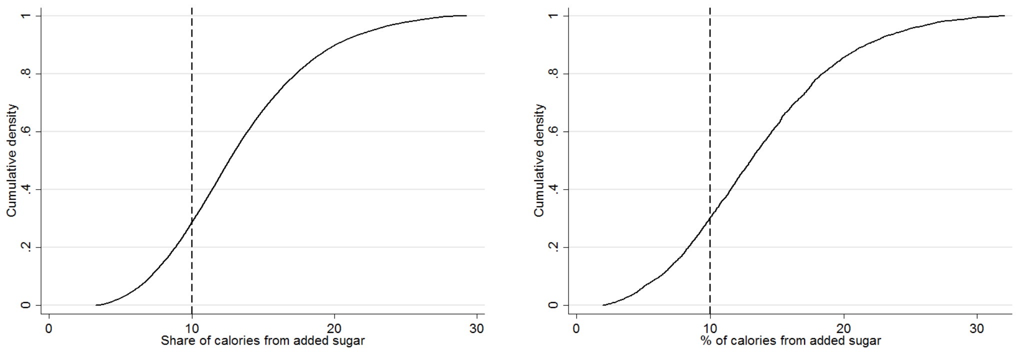
To assess empirically the likelihood of taxes on soda being effective, we require estimates of how strongly consumers will switch away from the sugar in soda ($S_i^{\prime w}$) and how strongly they will switch to alternatives ($S_i^{\prime nw}$) and we require measures of marginal internalities ($\phi_i^{\prime}$) and the marginal utility of income (λ_i). Importantly though, we also need to know the correlation between all these variables. In Section 3.3 we develop a demand model which allows us to estimated these correlations.

2.2 Over consumption of sugar and soda

Excess sugar consumption is a widespread phenomenon. The World Health Organization recommends people should obtain no more that 10% of their daily calories from added sugar, and that ideally they should get less than 5% from added sugar (WHO (2015)). Figure 2.1 shows that most people in the US obtain more of their calories from added sugar than these recommended amounts – over 70% of consumers are above the 10% threshold and over 95% are above the 5% threshold. The picture is very similar for the UK.

Figure 2.1: *Cumulative density of share of calories from added sugar*

(a) US (b) UK



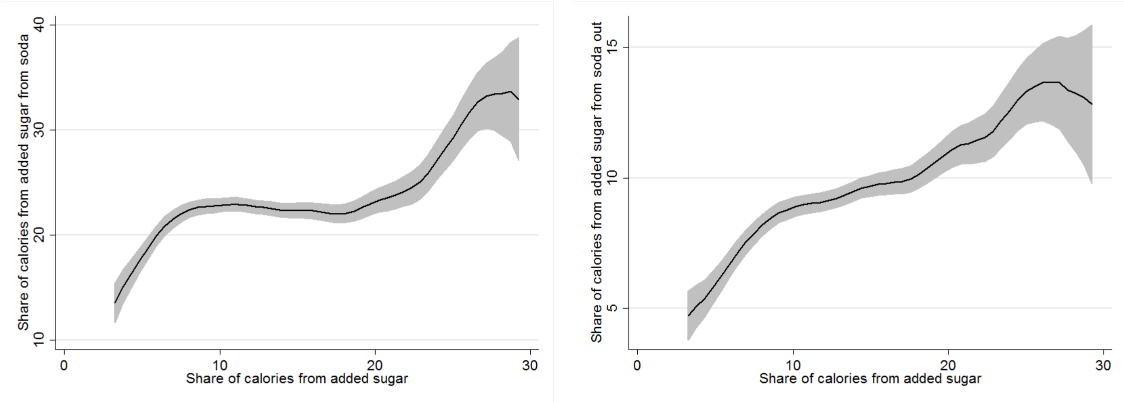
Source: NHANES and NDNS. Notes: Vertical dashed line is WHO recommended maximum.

Policy markers have specifically targeted the sugar in soda with introduction of soda taxes. One reason for focusing tax policy on this form of sugar, rather than levying a more broad based sugar tax, is soda accounts for a substantial share of sugar and does not contain other nutritionally beneficial nutrients. Thus a soda tax serves to increase the price of a popular form of sugar whilst limiting distortions to other nutrients in consumers' food baskets.

Figures 2.2 and 2.3 points to a second advantage to taxing the sugar on soda. The figure, for all soda purchases (panel (a)) and soda purchased for consumption on-the-go (panel (b)), describes the relationship between share of added sugar calories from soda and share of calories from added sugar. It shows that those consumers with high added sugar in their diet systematically obtain more of their added sugar from soda (both overall and on-the-go). Therefore taxing the sugar in soda affects a larger share of the added sugar of those consumers with the most added sugar in their diets (and that create most of the social harms from excess consumption).

Figure 2.2: *Relationship between soda and added sugar - US*

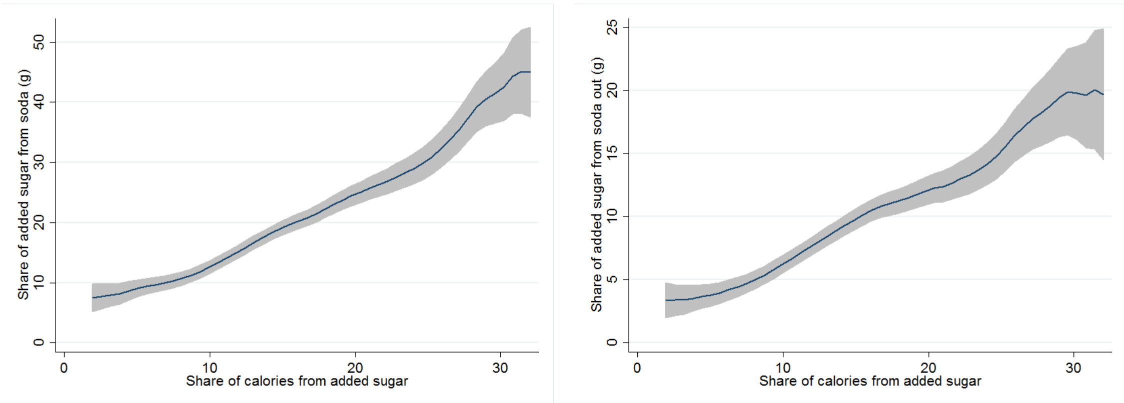
(a) All soda (b) Soda on-the-go



Source: NHANES.

Figure 2.3: *Relationship between soda and added sugar - UK*

(a) All soda (b) Soda on-the-go



Source: NDNS.

While this provides descriptive evidence in favour of tax on the sugar in soda being reasonably well targeted, how effective the measure will be in improving welfare (specifically, reducing the externalities from sugar consumptions while minimising the direct welfare costs to consumers) will depend on how demand responses vary across the added sugar distribution. We develop a model of demand for drinks that incorporates very rich consumer level heterogeneity and that allows us to relate this consumer level heterogeneity to other information about consumers (including how much of the their total calories is provided by added sugar). This allows us to assess the effectiveness of different forms of soda taxes implemented in practice.

3 On-the-go drinks demand

We model drinks (including soda) demand using longitudinal data on purchases made by a sample of consumers on-the-go. A key focus in modelling demand is to choose a specification that allows us to flexibly capture the distribution of preferences, and to relate characteristics of demand behaviour to consumer level information outside of the model. This enables us to assess the impacts of soda taxes, including whether they would be successful at lowering sugar consumption of the individuals who generate the largest marginal externalities, as well as enabling us to look at the distributional consequences of the policies.

3.1 On-the-go data

We exploit novel panel data that records purchases of foods and drinks made by a sample of individuals while on-the-go (i.e. foods and drinks purchased and consumed outside of the house, not including restaurant or canteen meals). Participants record all purchases of snacks and non-alcoholic drinks at the barcode (UPC) level using their mobile phones. The data contains product and store information, transaction level prices and demographic information of the consumer. The data are collected by the market research firm Kantar and are a random sample of individuals that live in households that participate in the Kantar Worldpanel.

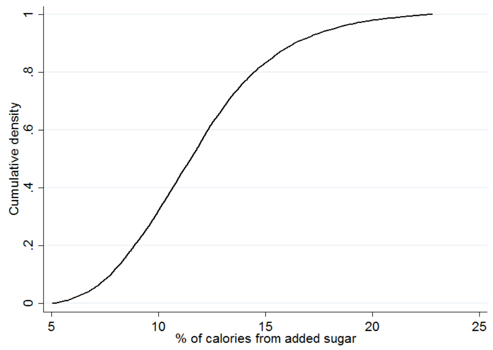
The Kantar Worldpanel is a longitudinal data set that tracks the grocery purchases made and brought into the home by a sample of households representative of the British population. Worldpanel households scan the barcode of all grocery purchases made and brought into the home. This means that we have comprehensive information on the total grocery baskets of the households to which the individuals in our on-the-go panel belong. The Kantar Worldpanel (and similar data collected in the US by AC Nielsen) have been used in a number of papers studying consumer grocery demand (see, for instance, Aguiar and Hurst (2007) and Dubois et al. (2014)). Data on food purchased on-the-go have, so far, been much less exploited.

The two most important measures of overall food purchasing behaviour we use in our analysis are the share of calories from added sugar in consumer grocery baskets and total equivalised grocery spending. By relating our consumer specific preferences estimates and estimates of the effects of soda taxes to this information, we can assess both whether soda taxes achieve reductions in sugar among consumers that have a large amount of sugar in their total diet and to what extent the taxes are regressive. In Figure 3.1 we show the cumulative density functions of both these variables.

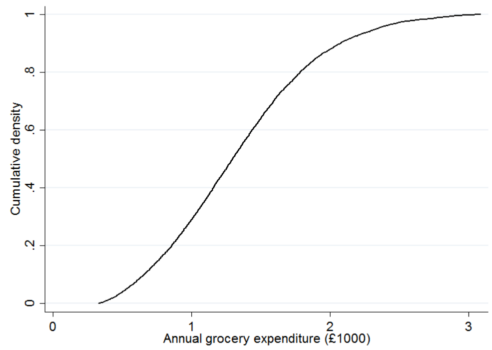
In Figure 3.1 we show the cumulative density functions of both these variables. Panel (a) is for share of calories from added sugar in consumer grocery baskets. For each consumer in our sample, we compute this as the share of the calories of the grocery basket their household purchases over the course of a calendar year that is comprised of sugar. The figure shows the distribution of the share of calories from added sugar in our data is similar to that NDNS intake data (see Figure 2.1), with the majority of households purchasing more than 10% of total energy intake from added sugar – the World Health Organization recommendation. Panel (b) is for total equivalised grocery spending. For each consumer we compute the equivalised annual grocery expenditure for the household that they belong to.[4]

Figure 3.1: *Distribution across households of:*

(a) Share of calories from added sugar



(b) Equivalised annual expenditure



Source: Kantar Worldpanel and Kantar on-the-go panel. Notes: In each case we trim the top and bottom percentiles of the distribution.

We use information on 5,373 individuals over the period June 2010-October 2012. We observe each person making purchases on a minimum of 25 days and on 81 days on average. We model demand for cold drinks – including both sugary and diet soda as well as fruit juice, flavoured milk and mineral water. This enables us to capture both switching towards diet soda and switching towards other drinks in response to a soda tax. In Section 6 we consider switching to sugar in chocolate and show **XXX**.

In Table 3.1 we describe the distribution of consumers by their participation in the market. We distinguish consumers into those that we never observe purchasing drinks (27.5%), that are observed purchasing only non-soda drinks (24.8%) and

[4]We use the OECD modified equivalence scale, see Hagenaars et al. (1994). It assumes for every additional adult (beyond the first) the household needs 0.5 times the resources of the first adult, and for every person younger than 14 a household needs 0.3 times the resources of the first adult.

that are observed purchasing soda (47.7%). We focus on modelling demand of the soda purchasing consumers – individuals that never purchase soda have zero soda demands and would be unaffected by a tax on soda. We observe these 2,563 soda purchasing consumers making 180,675 separate drinks purchases. Table 3.1 also shows that males and females under the age of 40 are more likely to purchase soda than older people.

Table 3.1: *Participation in market*

	Female		Male		Total
	<40	40+	<40	40+	
Never purchase drink	230	456	280	512	1478
	18.7	*30.5*	*23.5*	*35.3*	*27.5*
Only purchase non-soda drinks	321	430	242	339	1332
	26.0	*28.7*	*20.3*	*23.3*	*24.8*
Purchase soda	682	611	669	601	2563
	55.3	*40.8*	*56.2*	*41.4*	*47.7*
Total	1233	1497	1191	1452	5373
	100.0	*100.0*	*100.0*	*100.0*	*100.0*

Notes: Purchases by 5,373 individuals on-the-go over the period June 2010-October 2012. Column percent are shown in italics.

In Table 3.2 we show the main products in the UK on-the-go drinks market, including both soda products and non-soda options such as fruit juice, and flavoured milk. We model consumer choice between these products and the outside option (purchasing mineral water). The soda market is dominated by a set of well known brands. Most brands are available in both a sugary and diet variety, and often in two different container sizes. We omit small brands with market shares below 4% and some rarely purchased products. The products included in our analysis account for over 70% (by both volume and expenditure) of on-the-go drinks sales.

Table 3.2 also shows the share of transactions accounted for by each product, the mean prices and sugar content. There are seven soda brands. The most popular brand is Coca Cola, with a market share (by transactions) of 38.1%. In the food on-the-go market there are four Coca Cola products – regular (which are sugary) and diet varieties each available in a 330ml can and the more popular 500ml bottle. The bottles are more expensive than the smaller cans, even in per litre terms (£2.16 vs. £1.88), reflecting differences in costs and the fact that consumers have different preferences for bottle versus can. We include the composite option fruit juice and flavoured milk, which allows us to capture the possibility that consumers might respond to a soda tax by switching to alternative non-soda (but high sugar) drinks.

Table 3.2: *Drinks products*

	Product					
	Brand	Variety	Size	Market share	Price (£)	g sugar per 100ml
Soda options						
	Coca Cola			*38.1%*		
		Regular	330ml can	6.2%	0.62	10.6
		Regular	500ml bottle	11.2%	1.08	10.6
		Diet	330ml can	7.1%	0.63	0.0
		Diet	500ml bottle	13.6%	1.09	0.0
	Fanta			*5.9%*		
		Regular	330ml can	0.9%	0.60	6.9
		Regular	500ml bottle	4.5%	1.08	6.9
		Diet	500ml bottle	0.5%	1.07	0.6
	Cherry Coke			*4.3%*		
		Regular	330ml can	0.8%	0.63	11.2
		Regular	500ml bottle	2.4%	1.08	11.2
		Diet	500ml bottle	1.1%	1.08	0.0
	Oasis			*6.5%*		
		Regular	500ml bottle	5.9%	1.07	4.1
		Diet	500ml bottle	0.5%	1.06	0.5
	Pepsi			*15.1%*		
		Regular	330ml can	1.6%	0.61	11.0
		Regular	500ml bottle	3.5%	0.96	11.0
		Diet	330ml can	1.9%	0.62	0.0
		Diet	500ml bottle	8.2%	0.95	0.0
	Lucozade			*7.4%*		
		Regular	380ml bottle	3.8%	0.93	13.8
		Regular	500ml bottle	3.6%	1.13	13.8
	Ribena			*4.3%*		
		Regular	288ml carton	1.1%	0.65	10.5
		Regular	500ml bottle	2.4%	1.12	10.5
		Diet	500ml bottle	0.9%	1.10	0.5
Non-soda options	Fruit juice		330ml	4.0%	1.39	10.6
	Flavoured milk		500ml	2.2%	0.96	10.6
Outside option				12.3%		

Notes: Regular varieties are sugary. Market shares are based on transactions. Prices are the mean across all choice occasions.

Of the 2,563 consumers with positive soda demands, we can distinguish between those that always choose soda and those that sometimes choose an alternative drink (i.e. fruit juice, flavoured milk or the outside option). We can also distinguish between consumers who, when buying an inside option, always, sometimes or never choose a sugary drinks. Table 3.3 shows that 24.6% of consumers always choose soda and that, when purchasing a drink (other than the outside option), 5.1% of consumers buy only diet soda and 21.1% of consumers buy only sugary drinks. We will build this feature of behaviour into our demand model.

Table 3.3: *Soda consumers*

	Purchase:		
	Soda and non soda	Only soda	Total
Only diet	66	64	130
	2.6	*2.5*	*5.1*
Both diet and sugary	1492	399	1891
	58.2	*15.6*	*73.8*
Buy only sugary	375	167	542
	14.6	*6.5*	*21.1*
Total	1933	630	2563
	75.4	*24.6*	*100.0*

Notes: Percent of consumers shown in italics.

3.2 Prices

Product prices vary over time and across retail outlets. We compute the mean monthly price for each product in each retail outlet and use this in demand estimation. For each product we compute six price series. These include the price in the largest national retailer, Tesco, and the price in vending machines. Tesco prices nationally and vending machine prices do not vary much geographically. We therefore compute national price series for Tesco and vending machines.

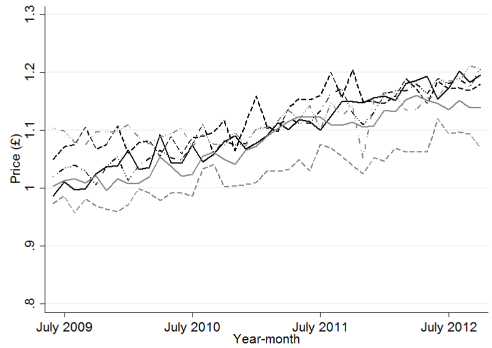
The other four price series are based on prices set by mainly smaller local stores, which make up around **XX**% of on-the-go purchases of soda. These vary geographically. We compute regional prices for the North, Midlands, South and London. On each choice occasion we observe where an individual shops, we assume that this is independent of demand shocks (see Section 3.4), and we assume that the consumer faces the vector of prices for products in the retailer that we observe them shopping in.

To illustrate the variation in prices that we use, in Figure 3.2 we plot the evolution of prices over time for the 330ml can (panel (a)) and 500ml bottle (panel (b)) of Coca Cola. We control for time varying brand effects in the demand estimates, so this means that we exploit differential time series variation in prices across the two container sizes and across stores. In panel (c) we plot the evolution of the ratio of the price of the can to the price of the bottle. The graph shows over time and stores that there is considerable variation in the ratio of the two prices.

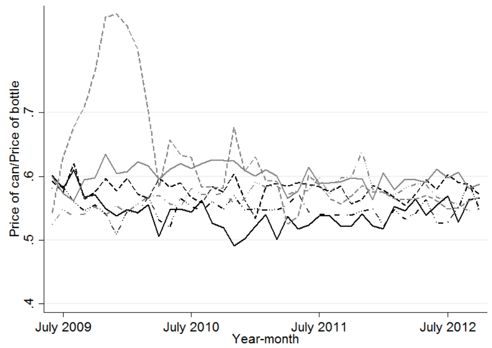
Figure 3.2: *Price variation for Coca Cola*

(a) 330ml can

(b) 500ml bottle



(c) Within brand price variation



Notes: Each line corresponds to a different retailer.

3.3 Demand model

We consider the decisions that consumers, indexed $i \in \{1, ..., N\}$, make over which drink to purchase when choosing for immediate consumption on-the-go. We observe consumer i on many choice occasions, indexed by $t = \{1, ..., T\}$. A choice occasion refers to a consumer visiting a store and purchasing a drink. We index the "inside" products by $j \in \{1, ..., j', j' + 1, ..., J\}$. Products $j \in \{1, ..., j'\} = \Omega_w$ are the set of soda products, and $j \in \{j' + 1, ..., J\} = \Omega_{nw}$ denotes alternative sugary drinks (fruit juice and flavoured milk). We denote the outside option of choosing water rather than juice by $j = 0$.

Each product $j > 0$ is associated with a vector of product attributes. These attributes include the price, p_{jrt}, which varies over time (t) and cross-sectionally across retail outlets (indexed by r), a dummy variable for whether the product is a sugary variety (rather than a diet variety) – denoted by s_j – and a dummy variable for whether the product is a soda – denoted by w_j. We allow for consumers to have heterogeneous (and potentially) correlated preferences for each of these attributes. In addition, we include a set of additional attributes (denoted by $\mathbf{x}_{jt}$), including size, carton type and time-varying brand effects. We allow the influences of these attributes on utility to vary by gender and age (whether the consumer is younger than 40) – we index the gender-age group with $d = (1, ..., D)$.

One convenient feature of considering soda purchased on-the-go for immediate consumption is that we do not have to worry about stockpiling (as in Wang (2015)); by definition the consumption occasions that we are modelling do not involve storage. These consumers might also have purchased soda and stored it at home; we assume that any inventories that they hold do not affect their decision over immediate consumption. We consider the robustness of our results to this assumption in Section 6.

We assume the pay-off associated with purchasing a product, $j > 0$, takes the form:

$$U_{ijt} = \alpha_i p_{jrt} + \beta_i s_j + \gamma_i w_j + g(\mathbf{x}_{jt}; d, \eta) + \epsilon_{ijt}, \qquad (3.1)$$

where ϵ_{ijt} is an idiosyncratic shock distributed type I extreme value.

$\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_N)'$ and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_N)'$ are vectors of individual preference parameters over which we make no distributional assumptions. We use the large T dimension of our data to recover estimates of individual specific parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and the large N dimension to construct the nonparametric estimate of the joint probability distribution function $f(\alpha_i, \beta_i, \gamma_i)$. We can also construct the distribution of preferences conditional on observable consumer characteristics, X; $f(\alpha_i, \beta_i, \gamma_i|X)$. These observable characteristics can be demographic variables or measures of the overall diet or grocery purchasing behaviour of the consumer.

Our estimates may be subject to an incidental parameter problem that is common in non linear panel data estimation. Even if both $N \to \infty$ and $T \to \infty$ an asymptotic bias may remain, although it shrinks as the sample size rises (Arellano and Hahn (2007)). The long T dimension of our data is helpful in lowering the chance that the incidental parameter problem leads to large biases in our case. We implement the split sample jackknife bias correction procedure suggested in Dhaene and Jochmans (2015) and in Section 6 show that the bias correction does not impact our main conclusion.

In addition to the individual specific preference parameters for price, soda and sugar we estimate a set of gender-age group specific parameters that capture the effects of other product attributes on utility, η are additional preference parameters that appear in the function $g(.)$. We assume this function takes the form:

$$g(\mathbf{x}_{jt}; d, \eta) = \delta_d z_j + \xi_{db(j)t} + \zeta_{db(j)r}, \tag{3.2}$$

where z_j denotes a set of fixed effects capturing size and carton type and $\xi_{db(j)t}$ denotes a set of time varying gender-age group-brand effects. $b(j)$ denotes the brand that product j belongs to. Each product belongs to one of B brands, shown in the first column of Table 3.2. There are more products than brands ($B < J$), since most brands come in at least two different sizes and in sugary and diet varieties. Exceptions to this are the composite brands fruit juice and flavoured milk, which we only allow to come in one variety. The time varying brand effects will capture any brand shocks to demands through, for example, any effects of national advertising or promotion campaigns. We assume that preferences over product size and carton are fixed over time, but might vary across demographic group (gender and age). We discuss our identification strategy in detail in Section 3.4.

The payoff associated with choosing the outside option, $j = 0$, is given by:

$$U_{i0t} = \zeta_{d0rt} + \epsilon_{i0t}, \tag{3.3}$$

where ζ_{d0rt} are gender-age, retail outlet specific deviations in the mean outside option pay-off

We are able to use the long time dimension of our data to identify consumers that have infinite preferences for some characteristics. For instance, assuming that the unobservable error term has "large" support (we assume infinite support with an extreme value distribution), a consumer that always chooses one of the non-soda options, (fruit juice, flavoured milk or the outside option) can be thought of as having a negatively infinite soda preference parameter $\gamma_i = -\infty$. Such consumers have purchase probabilities given by $P_{it}(j) = 0$ for $j \in \Omega_w$ and $\sum_{j \in \Omega_{nw}} P_{it}(j) = 1$. Consumers that always purchase soda can be thought of as having positively infinite soda preferences $\gamma_i = \infty$ and those that sometimes purchase soda have finite soda preferences $\gamma_i \in (-\infty, \infty)$.

With cross-sectional data, or panel data with only a few observations per consumer, it would not be possible to identify infinite regions of the distribution of soda preferences: a consumer may be observed never purchasing soda simply because it got a series of high draws of $(\epsilon_{i0t}, \epsilon_{ij'+1t}, ..., \epsilon_{iJt})$ over time. However, with many purchases for each consumer, getting such a series of draws becomes a zero probability event, allowing for identification of infinite soda preferences. Our identification of infinite soda or sugar preferences relies on the fact that we have a large T dimension that allows us to use asymptotic results on conditional choice probabilities. However, it is possible that a consumer that we observe, when purchasing a drink, always chooses a soda (and hence is modelled as having an infinite soda preference), would switch to alternatives to soda if the price of all sodas were increased by a sufficiently large amount (as a result of the asymptotic error associated with using a large but finite sample when identifying the soda preference). We consider soda taxes that are similar to those currently proposed and that do not involve very large prices increase (price increases of around 10%). It is unlikely that such price increases would induce a consumer who has never chosen non-sodas over dozens of choice occasions to switch towards them. We test the robustness of our results to this assumption in Appendix A.

A similar argument applies for sugar preferences; consumers that only buy diet soda (or the outside option) have negatively infinite sugar preferences ($\beta_i = -\infty$) and consumers that only buy sugary products (or the outside option) have positively infinite sugar preferences ($\beta_i = \infty$). Those consumers observed purchasing both diet and sugary soda across their choice occasions have finite sugar preferences ($\beta_i \in (-\infty, \infty)$).

To express formulae for consumer choice probabilities it is convenient to both distinguish between the set of soda options, Ω_w and non-soda juices Ω_{nw} and also between the set of sugary sodas, $\Omega_s = \{j | j \in \Omega_w, s_j = 1\}$, and diet sodas $\Omega_{ns} = \{j | j \in \Omega_w, s_j = 0\}$. For notation simplicity we use the following notation to denote the union of two sets $\Omega_{s,nw} = \Omega_s \cup \Omega_{nw}$ (i.e. the set of sugary sodas plus the set of non-soda juices that contain sugar). Our assumption that ϵ_{ijt} is an idiosyncratic shock distributed type I extreme value means the consumer level choice probabilities are given by the multinomial logit formula. The exact formula depends on whether the consumer has infinite or finite preferences for soda and sugar (see Table 3.4).

Table 3.4: *Logit choice probabilities ($P_{it}(j)$)*

	Soda preference	
Sugar preference	$\gamma_i \in (-\infty, \infty)$	$\gamma_i = \infty$
$\beta_i = -\infty$	$\frac{\exp(\zeta_{d0rt})1_{j=0} + \exp(\alpha_i p_{jrt} + g(\mathbf{x}_{jt}; d, \eta))1_{j \in \Omega_{ns,nw}}}{\exp(\zeta_{d0rt}) + \sum_k \exp(\alpha_i p_{krt} + g(\mathbf{x}_{kt}; d, \eta))1_{k \in \Omega_{ns,nw}}}$	$\frac{\exp(\alpha_i p_{jrt} + g(\mathbf{x}_{jt}; d, \eta))1_{j \in \Omega_{ns}}}{\sum_k \exp(\alpha_i p_{krt} + g(\mathbf{x}_{kt}; d, \eta))1_{k \in \Omega_{ns}}}$
$\beta_i \in (-\infty, \infty)$	$\frac{\exp(\zeta_{d0rt})1_{j=0} + \exp(\alpha_i p_{jrt} + \beta_i s_j + g(\mathbf{x}_{jt}; d, \eta))1_{j \in \Omega}}{\exp(\zeta_{d0rt}) + \sum_k \exp(\alpha_i p_{krt} + \beta_i s_k + g(\mathbf{x}_{kt}; d, \eta))1_{j \in \Omega}}$	$\frac{\exp(\alpha_i p_{jrt} + \beta_i s_j + g(\mathbf{x}_{jt}; d, \eta))1_{j \in \Omega_w}}{\sum_k \exp(\alpha_i p_{krt} + \beta_i s_k + g(\mathbf{x}_{kt}; d, \eta))1_{k \in \Omega_w}}$
$\beta_i = \infty$	$\frac{\exp(\zeta_{d0rt})1_{j=0} + \exp(\alpha_i p_{jrt} + g(\mathbf{x}_{jt}; d, \eta))1_{j \in \Omega_{s,nw}}}{\exp(\zeta_{d0rt}) + \sum_k \exp(\alpha_i p_{krt} + g(\mathbf{x}_{kt}; d, \eta))1_{k \in \Omega_{s,nw}}}$	$\frac{\exp(\alpha_i p_{jrt} + g(\mathbf{x}_{jt}; d, \eta))1_{j \in \Omega_s}}{\sum_k \exp(\alpha_i p_{krt} + g(\mathbf{x}_{kt}; d, \eta))1_{k \in \Omega_s}}$

Notes:

If we denote $y_i = (y_{i1}, ..., y_{iT})$ consumer i's sequence of choices across all choice occasions. The probability of observing y_i is given by:

$$\mathcal{P}_i(y_i) = \prod_t P_{it}(y_{it}) \tag{3.4}$$

and the associated log-likelihood function is:

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \eta) = \sum_i \ln \mathcal{P}_i(y_i). \tag{3.5}$$

3.4 Identification

The principal identification challenge we face relates to separating the causal impact of price on demand from shocks to demands. If there are demand shocks that we do not control for and that are correlated with product prices this will lead to inconsistent estimates of the price (and other) preference parameters. Our identification strategy exploits the rich very granular nature of the food on-the-go data and institutional features of the UK grocery market that allow us to isolate exogenous price variation.

We measure product prices, p_{jrt} at the retail outlet level (indexed by r). In a given time period the price of Coca Cola varies across the 330ml can and 500ml bottle version of the product and across retail outlets (there is little price variation across sugary and diet varieties of the same brand-size). The inclusion of time varying brand effects, $\xi_{db(j)t}$, in utility means we control for aggregate time varying shocks to demand. These will absorb the effects of seasonality and national brand advertising on demands.

The price variation we exploit to identify slopes of demands is i) cross-retailer variation in the relative prices of different drinks and ii) time series variation in products price *within* brands that is driven by factors other than shocks to consumers' soda demands. We address each source in turn.

The retail outlets include a set of large supermarket chains that price nationally and a set of smaller outlets with regionally varying prices. In demand we control for retailer effects (including in the outside option). We exploit time series variation in the relative price of soda products across retail outlets relative to the average difference. The identifying assumption is that differential changes in the prices of different sodas across retailers are not driven by retailer-time varying demand shocks for soda products. We think this is a plausible assumption. In the UK soda market **x**% of soda advertising is done nationally and by the manufacturer. There is very little retailer or regional advertising. Differential price movements across retail outlet are likely to be driven by differences in vertical contracts with manufacturers (or, in the case of the many small stores, proximity of nearest large wholesale store) and promotions related to excess stock. As we study goods that are purchased for immediate consumption, retailer level promotions are a useful source of price variation that are unlikely to give rise to the usual stocking up concerns (see Hendel and Nevo (2006)).

In exploiting cross retailer price variation we also assume that individual level demand shocks to specific soda products do not drive store choice for the on-the-go market; for instance, a violation of this assumption would occur if a consumer that has a demand shock that leads them to want Coca Cola visits a retailer that happens to temporarily have a low price for that product, and, if instead they had a demand shock that led them to want Pepsi they would have selected a retailer with a relatively low Pepsi price. Such behaviour would occur either if consumers could predict fluctuating relative prices across retailers or if they visited several retailers in search of a low price draw for the product they are seeking. We find either scenario highly unlikely in the case on-the-go soda (which makes up just **x**% of total household spending).

The second source of price variation is due to nonlinear pricing across container sizes that is common in the UK (prices are linear for a fixed container size but nonlinear across different container sizes of the same brand). This price variation is not collinear with the size fixed effects. In addition, the extent of nonlinear pricing varies over time and retailers. What would invalidate this as a source of identification is if there were systematic shocks to consumers' valuation of sizes that were differential across brand after conditioning on time varying brand effects and container size and type effects. Rather, it is more plausible that such tilting of brand price schedules is driven by cost variations that are not proportional to pack size, differential pass-through of cost shocks and differences in how brand advertising affects demands for different pack sizes. This identification argument is similar to that in Bajari and Benkard (2005). In an application to the computer market, they assume that, conditional on observables, unobserved product characteristics are the same for all products that belong to the same model. We assume that conditional on time varying brand characteristics, unobserved size characteristics do not vary differentially across brand.

4 Parameter estimates

4.1 Preference heterogeneity

In Table 4.1 we summarise the parameter estimates – obtained by maximising the likelihood function (equation 3.5). The top panel summarises the estimates of the consumer specific preference parameters for the price, soda and sugar attributes, reporting moments on the distribution. These are based on the finite portion of the joint preference distribution. The bottom panel reports the estimates of the size and brand effects. These vary across consumer gender and age group (based on whether the consumer is below 40 years old or not). We normalise the mean effect of the outside option, the 330ml can effect and the Coca Cola brand effect to zero, meaning that included container size/type and brand effects are estimated relative to these omitted groups.[5] The reported brand effects are for the first period in the data (June 2010). We allow each of them to vary through time (from month-to-month).[6]

The mean of the distribution of price preference parameters is -1.79, with a standard deviation of 4.35. On average, consumers dislike higher prices, with the large standard deviation indicating considerable heterogeneity in how important prices are in the purchase decisions of different consumers. The soda preferences capture, conditional on a consumer's preferences over other product attributes, the desirability of purchasing soda over fruit juice, flavoured milk or the outside option; a more positive soda preference implies a higher baseline utility from soda. The standard deviation (2.86) in soda preferences indicates considerable preference heterogeneity. A consumer's sugar preference captures the desirability of purchasing a sugary product over a diet one; a more positive sugar preference implies a higher baseline taste for sugary drinks over diet soda. Like preferences for soda, preference for sugar are dispersed (with a standard deviation of 2.05).

We do not need to impose any distributional assumption on consumer preferences over price, soda and sugar and in particular we do not assume the marginal distributions are normal as is common in random coefficient models. The skewness and kurtosis of the price and sugar preference distributions indicate departures from normality – price preferences are positively skewed and leptokurtic (i.e. kurtosis above 3 indicating fatter tails than a normal distribution) and the finite sugar preferences are negatively skewed and leptokurtic. The finite soda preferences have

[5] In most applications of discrete choice demand models, if one normalises the mean utility from the outside option to zero, it is not necessary to also drop one of the brand effects. The difference in our case is due to the fact we include the soda characteristic.

[6] We do not report the time varying brand effects or the retailer effects in Table 4.1. These are available upon request.

kurtosis close to 3 and skewness close to 0 (like a normal distribution). However both the sugar and soda preferences distributions have infinite portions too (see Figure 4.2 below).

The covariance matrix of consumer preferences over price, soda and sugar is unrestricted (we only assume that individual preferences are stable over time, for the 28 months period of data), allowing consumers' preferences for sugar to be related to the price sensitivity as well as to the taste for soda. We find that price preferences are strongly negatively correlated with soda preferences and negatively correlated with sugar preferences. This means that consumers that are relatively price sensitive (have a more negative price parameter) tend to have relatively strong preferences for both soda and sugar compared to less price sensitive consumers. Soda and sugar preferences are positively correlated. In Figure 4.1 we show contour plots of the bivariate distribution of consumer specific preferences – these graphically illustrate the pattern of correlation in preferences (based on the finite portion of the distributions).

Table 4.1: *Model estimates*

Moments of distribution of consumer specific preferences

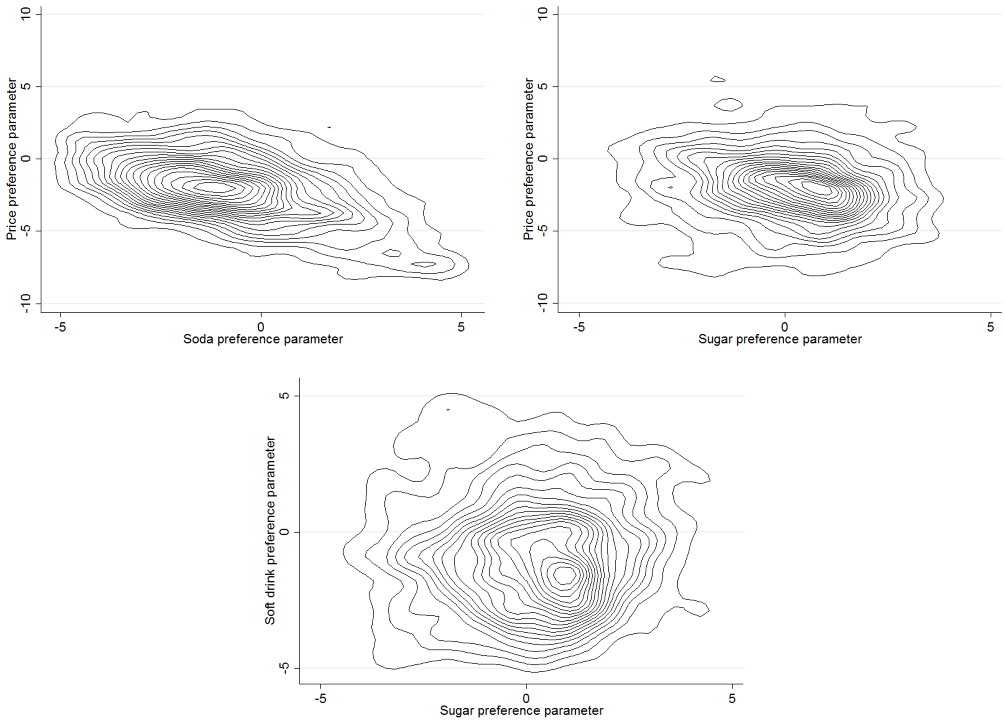
Variable		Estimate	Standard error
Price	Mean	-1.7856	0.0727
	Standard deviation	4.3500	0.0898
	Skewness	0.6999	0.1892
	Kurtosis	6.6126	0.9868
Soda	Mean	-0.6297	0.0938
	Standard deviation	2.8590	0.0622
	Skewness	0.1663	0.1994
	Kurtosis	5.5891	1.0256
Sugar	Mean	-0.0027	0.0218
	Standard deviation	2.0513	0.0836
	Skewness	-0.8439	0.5754
	Kurtosis	9.3688	4.3680
Price-Soda	Covariance	-5.6252	0.3427
Price-Sugar	Covariance	-1.0102	0.2236
Soda-Sugar	Covariance	0.4631	0.2928

Consumer group specific preferences

Variable	Estimate	Standard error	Estimate	Standard error
	Female - <40		*Female - 40+*	
288ml carton	1.2407	0.0469	0.6903	0.0698
380ml bottle	2.3507	0.0532	2.2498	0.0568
500ml bottle	2.2819	0.0577	2.2436	0.0669
Fanta	-1.7720	0.1469	-1.7825	0.1479
Cherry Coke	-1.6180	0.1408	-2.5570	0.1922
Ribena	-1.1724	0.1177	-1.3078	0.1191
Pepsi	-0.9854	0.0957	-0.9844	0.1003
Lucozade	-2.0206	0.1710	-1.5247	0.1418
Oasis	-2.1829	0.1527	-1.9083	0.1391
Fruit juice	-0.1399	0.2867	1.4397	0.3273
Flavoured milk	-3.5404	0.2616	-2.9586	0.3426
	Male - <40		*Male - 40+*	
288ml carton	-0.1527	0.0634	-0.1002	0.0658
380ml bottle	2.1269	0.0452	2.3867	0.0502
500ml bottle	2.5041	0.0544	2.2437	0.0587
Fanta	-1.8289	0.1238	-1.3700	0.1059
Cherry Coke	-2.1613	0.1489	-1.9045	0.1428
Ribena	-2.3484	0.1506	-1.1716	0.1091
Pepsi	-1.5897	0.1046	-0.9880	0.0873
Lucozade	-1.6734	0.1323	-1.8384	0.1177
Oasis	-2.3463	0.1582	-2.8384	0.1747
Fruit juice	1.4702	0.3233	-0.6937	0.3473
Flavoured milk	-2.3191	0.2501	-3.7719	0.3130
Time-demographic-brand effects			Yes	
Retailer-demographic-brand effects			Yes	

Notes: Estimates based on a sample of 2,563 soda consumers and 180,675 choice occasions. Moments of distribution of heterogenous preferences are computed using estimates of consumer specific preference parameters. These moments are based on consumers with finite parameters and omit the top and bottom percentile of each distribution. Standard errors for moments are computed using the delta method.

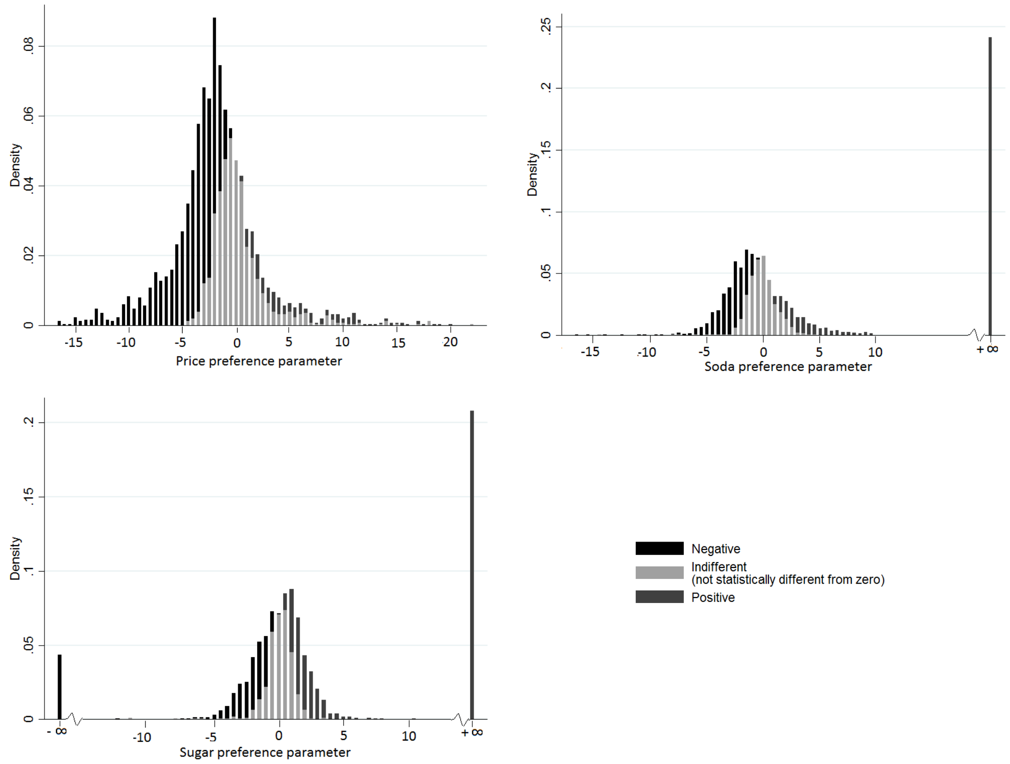
Figure 4.1: *Bivariate distributions of consumer specific preference parameters*



Notes: Distribution plots are based on consumers with finite preference parameters.

In Figure 4.2 we plot the marginal distribution of preferences over price, soda and sugar. The shading represents consumers with negative, positive and indifferent (i.e. not statistically significantly different from zero) preferences for each attribute. The first figure shows the coefficient on price – 52.1% of consumers have a negative and statistically significant coefficient; for 40.1% of consumers the price preference parameter is not statistically different from zero – for these consumers price does not weigh heavily on their selection of soda. A small fraction of consumers are estimated to have positive and statistically significant price coefficients. This, at least to some extent, is likely to reflect sampling uncertainty. 29% of consumers have negative soda preferences and 58% of consumers have positive soda preferences (including around 24% with an infinite preference for soda). For sugar, there are 25% of consumers with negative preferences and 68% with positive sugar preferences.

Figure 4.2: *Univariate distributions of consumer specific preference parameters*



Notes: The top and bottom percentiles of (the finite) part of the distribution are omitted from these figures.

In random coefficient models, preference heterogeneity is typically specified to be orthogonal to any other consumer varying aspect of the model. For instance, if prices or choice sets vary cross-sectionally random coefficient models impose that this cross consumer variation is statistically independent from preference heterogeneity. We do not need to make this assumption. This means, for example, that consumers with finite soda preferences and consumers with infinitely positive soda preferences (which in effect means the non-soda options are in the former set of consumers' choice sets but not the latter) may have different distributions of price and sugar preferences. Similarly consumers with infinitely negative, finite and infinitely positive sugar preferences (corresponding to choice sets with only diet sodas, diet and sugary drinks, and sugary drinks) may have different price and soda preference distributions.

Tables 4.2 and 4.3 show that we find evidence for this in practice. Table 4.2 show the 25th, 50th and 75th percentiles of the price and sugar preferences distribution for consumers with finite and infinite soda preferences. Table 4.3 show the 25th, 50th and 75th percentiles of the price and soda preferences distribution for consumers with finite and infinite sugar preferences. In each case 95 percent confidence intervals are given in brackets.[7] Consumers that choose between the soda and non-soda options ($\beta_i \in [-\infty, \infty]$) have more compressed price and sugar preference distributions than those consumers that only choose between the set of soda options ($\beta_i = \infty$). Consumers that never select sugary drinks ($\gamma_i = -\infty$) have a soda preference distribution shifted rightwards to those with finite and positive infinite sugar preferences, while consumers with finite sugar preferences have price and soda preferences distributions with less dispersion than consumers with infinite sugar preferences.

Table 4.2: *Variation in preferences between consumers with finite and infinite soda preferences*

	Percentile of distribution					
	Price preference			Sugar preference		
Soda preference	25th	50th	75th	25th	50th	75th
$\beta_i \in [-\infty, \infty]$	-3.5	-1.9	-0.3	-1.1	0.2	1.3
	[-3.8, -3.5]	[-2.1, -1.8]	[-0.4, 0.0]	[-1.2, -1.0]	[0.1, 0.2]	[1.3, 1.4]
$\beta_i = \infty$	-5.1	-2.2	2.1	-1.4	0.0	1.3
	[-6.0, -5.3]	[-2.7, -2.0]	[1.5, 2.6]	[-1.6, -1.3]	[-0.2, 0.1]	[1.1, 1.5]

Notes: Consumers with a soda preference parameter $\beta_i \in (-\infty, \infty)$, when purchasing a drink, choose between sodas and non-soda option. Consumers with $\beta_i = \infty$, when purchasing a drink choose between sodas.

Table 4.3: *Variation in preferences between consumers with finite and infinite sugar preferences*

	Percentile of distribution					
	Price preference			Soda preference		
Sugar preference	25th	50th	75th	25th	50th	75th
$\gamma_i = -\infty$	-4.6	-1.5	2.4	-4.9	-0.1	2.3
	[-6.1, -4.1]	[-2.3, -1.1]	[1.6, 4.2]	[-7.6, -3.5]	[-0.9, 0.8]	[1.8, 3.5]
$\gamma_i \in [-\infty, \infty]$	-3.6	-2.0	-0.2	-2.3	-0.9	0.6
	[-4.0, -3.7]	[-2.2, -1.8]	[-0.4, 0.0]	[-2.6, -2.2]	[-1.1, -0.7]	[0.6, 1.0]
$\gamma_i = \infty$	-4.2	-1.8	0.8	-2.9	-0.9	1.5
	[-4.9, -4.2]	[-2.0, -1.5]	[0.5, 1.1]	[-3.3, -2.7]	[-1.2, -0.5]	[1.2, 2.0]

Notes: Consumers with a sugar preference parameter $\gamma_i = -\infty/\gamma_i = \infty$, choosing between drinks that are diet/sugary. Consumers with $\gamma_i \in (-\infty, \infty)$ choose between both diet and sugary drinks.

[7] We calculate confidence intervals by first obtaining the variance-covariance matrix for the parameter vector estimates using standard asymptotic results. We then take 100 draws of the parameter vector from the joint normal asymptotic distribution of the parameters and, for each draw, compute the statistic of interest, using the resulting distribution across draws to compute Monte Carlo confidence intervals (which need not be symmetric around the statistic estimates).

The consumer preferences have distributions with infinite sections, non-normal finite portions and rich correlations. Our demand models is sufficiently flexible to capture these rich effects and to allow us to credibly uncover heterogeneity in the effects of soda taxes.

4.2 Relationship between preferences, total sugar consumption and grocery expenditure

As well as allowing us to nonparametrically characterise the joint distribution of preferences, our model enables us to describe how consumer level preference parameters relate to consumer demographics or other aspects of their behaviour. This relies on being able to recover consumer level parameters (rather than the parameters governing the preference distribution, as in random coefficient models). We relate preferences to the two measures of consumers broader grocery demand outlined in Section 3.1 – the share of their total calories from added sugar and equivalised total grocery expenditure.

In Table 4.4 we summarise how price, soda and sugar preference parameters and predicted annual sugar consumption from drinks varies across the distribution of share of total grocery basket calories from added sugar. The first column shows the mean value of each variable and the subsequent four shows the average deviation in each variable from the mean in each quartile of the added sugar distribution. In Figure 4.3 we also show the relationship graphically as kernel weighted local polynomial regressions.

Consumers with a relatively low amount of added sugar in their diet tend to be relatively price sensitive – those in the bottom quartile of the added sugar distribution have an average price parameter 0.18 lower than the mean, while those in the top quartile have an average 0.27 above the mean. There is little variation in soda preference parameters across this added sugar distribution – with the exception that consumers at the very bottom of the added sugar distribution have relatively low soda preferences. In contrast, sugar preference parameters and share of total calories from added sugar show a strong relationship; consumers with a higher share of added sugar in their total grocery baskets systematically have stronger estimated preferences for sugar based on their on-the-go drinks purchases. The relationship is very intuitive. However, it is important to recognise we do not impose this; we find that sugar preferences estimated off of individual level on-the-go drinks demand are strongly positively related to the total share of sugar in household level diets across the year based on all grocery purchases that are brought into the home, a measure that is completely separate from our model. This is evidence that the model captures features of consumers' drinks demands. The final row of Table 4.4 and Figure 4.3 (d) show that the model recovers the positive relationship between total sugar consumption from drinks and the share of added sugar across all groceries evident in the data (see Section 2.2).

Table 4.5 and Figure 4.4 repeat the analysis of Table 4.4 and Figure 4.3, but instead focus on how price, soda and sugar preference parameters and predicted sugar consumption from on-the-go drinks varies across the equivalised total annual grocery distribution. Price and soda preferences are strongly related to equivalised expenditure; people from low spending households are typically relatively price sensitive and have relatively strong preferences for soda. The correlation between price preferences and equivalised expenditure, like the correlation between sugar preferences and share of calories from added sugar, is intuitive and evidence that our demand estimates recover realistic correlations in behaviour (including between drink preferences and measures completely outside the model). The correlation between equivalised expenditure and sugar preferences is weaker (consumers from low spending households have somewhat stronger sugar preferences than consumers from higher spending households), however their is strong negative relationship between equivalised grocery expenditure and sugar consumption from drinks.

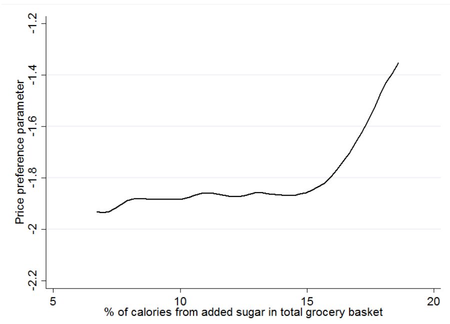
Table 4.4: *Relationship between preference parameters and share of calories from added sugar*

	Mean	Average deviation from mean preference parameter for quartile of added sugar distribution: 1	2	3	4
Price preference parameter	-1.83	-0.12	-0.05	0.01	0.14
	[-1.96, -1.71]	[-0.22, -0.01]	[-0.12, 0.02]	[-0.06, 0.10]	[0.01, 0.26]
Soda preference parameter	-0.66	0.02	-0.03	-0.03	0.05
	[-0.84, -0.49]	[-0.06, 0.09]	[-0.09, 0.02]	[-0.08, 0.02]	[-0.01, 0.12]
Sugar preference parameter	0.00	-0.33	-0.08	0.05	0.33
	[-0.03, 0.05]	[-0.39, -0.27]	[-0.13, -0.02]	[0.01, 0.08]	[0.28, 0.37]
Predicted sugar consumption (kg)	1.62	-0.25	-0.22	-0.01	0.38
	[1.59, 1.62]	[-0.25, -0.23]	[-0.22, -0.20]	[-0.02, -0.01]	[0.36, 0.38]

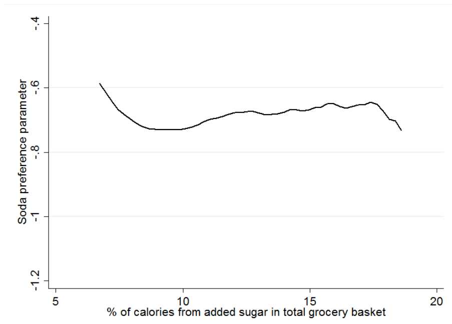
Notes: For each quartile of the distribution of share of calories from added sugar we report the mean deviation from the average value for each variable shown in the first column. 95% confidence intervals are given in brackets.

Figure 4.3: *Relationship between preference parameters and share of calories from added sugar*

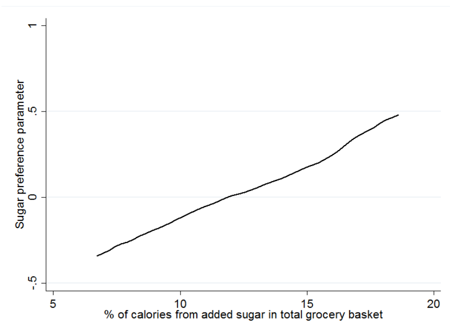
(a) Price preference parameter



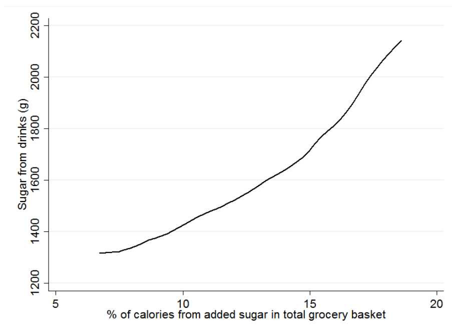
(b) Soda preference parameter



(c) Sugar preference parameter



(d) Predicted soda sugar consumption



Notes: Lines are local polynomial regression.

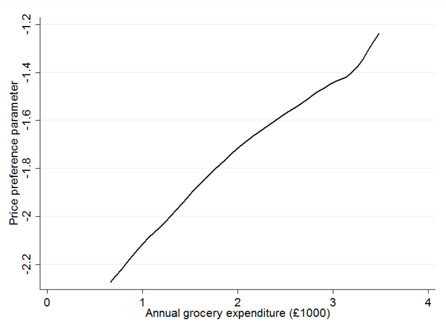
Table 4.5: *Relationship between preference parameters and equivalised annual grocery expenditure*

	Mean	Average deviation from mean preference parameter for quartile of equivalised grocery distribution: 1	2	3	4
Price preference parameter	-1.83	-0.55	-0.02	0.18	0.39
	[-1.96, -1.71]	[-0.64, -0.44]	[-0.03, 0.13]	[0.08, 0.26]	[0.23, 0.46]
Soda preference parameter	-0.65	0.26	0.24	-0.15	-0.32
	[-0.84, -0.49]	[0.15, 0.33]	[0.15, 0.26]	[-0.22, -0.09]	[-0.38, -0.21]
Sugar preference parameter	0.01	0.19	0.03	0.00	-0.20
	[-0.03, 0.05]	[0.18, 0.29]	[-0.02, 0.06]	[-0.11, -0.04]	[-0.21, -0.11]
Predicted Sugar consumption (kg)	1.62	0.11	0.19	-0.04	-0.29
	[1.59, 1.62]	[0.18, 0.20]	[0.21, 0.23]	[-0.15, -0.14]	[-0.36, -0.34]

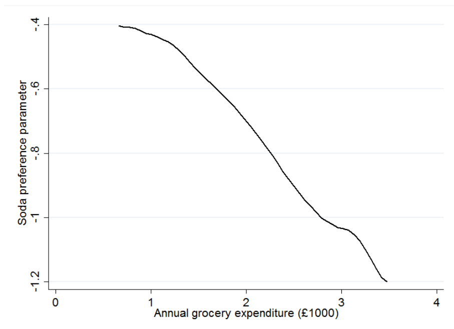
Notes: For each quartile of the equivalised annual grocery expenditure we report the mean deviation from the average value for each variable shown in the first column. 95% confidence intervals are given in brackets.

Figure 4.4: *Relationship between preference parameters and equivalised annual grocery expenditure*

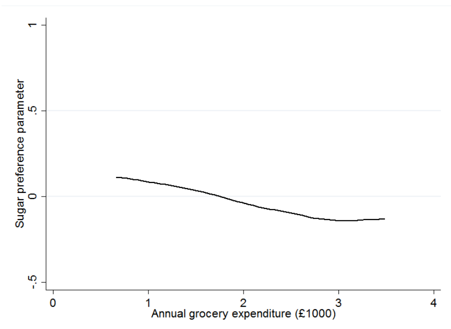
(a) Price preference parameter



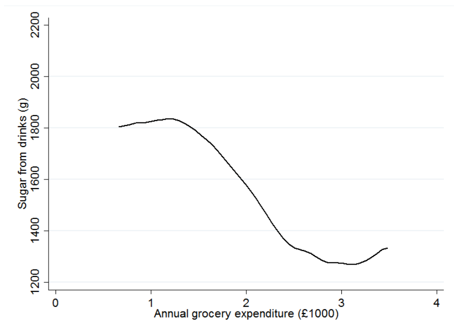
(b) Soda preference parameter



(c) Sugar preference parameter



(d) Predicted soda sugar consumption



Notes: Lines are local polynomial regressions.

4.3 Product demands

The demand curve for a product is obtained by aggregating over consumer level demands – the demand for good j at time t is $Q_t(j) = \sum_i P_{it}(j)$, where the consumer level demands, $P_{it}(j)$ are defined in Section 3.3. The own price elasticity for good j is $\frac{\partial Q_t(j)}{\partial p_{jt}} \frac{p_{jt}}{Q_t(j)}$, and the cross price elasticities are defined analogously. Rich heterogeneity in consumer preferences translates into flexibility in price elasticities, allowing us to capture well the aggregate effects of soda taxes. In Tables 4.6 and 4.7 we provide some details of price elasticities.

Table 4.6 summarises the mean own and cross price elasticities for the set of soda products, averaged across time. The first column shows own price elasticities, the next two columns show the mean cross price effects with respect to alternative sugary (including fruit juice and flavoured milk) and alternative diet products. The final column show the effect of a marginal change in price on total drinks demand. For example, a 1% increase in the price of Coca Cola in a 330ml can would result in a reduction in demand for that product of around 2.1%. Demand for alternative sugary products would rise by around 0.14% and demand for diet products would rise by 0.06%. Demand for juice drinks as a whole would fall by 0.01%. The numbers make clear that consumers are more willing to switch from sugary soda products to sugary alternatives and from diet products to diet alternatives, than they are between sugary and diet products. The table also shows that demand for the larger 500ml sizes tends to be less elastic than demand for smaller varieties.

The cross price elasticities shown in Table 4.6 mask a lot of differential substitution patterns between products. In Table 4.7 we report cross price elasticities at the product level for the two largest brands – Coca Cola and Pepsi. The table shows that substitution between the products is much stronger for products that are both sugary/diet and that either are of the same size or brand. For instance, a 1% increase in the price of Coca Cola 330ml results in an increase in demand of 0.55% for Pepsi 330ml – nearly four times as large as the increase in demand for any other product. In contrast consumers that purchase Coca Cola 500ml switch most strongly to Coca Cola 330ml.

The price effects in Table 4.6 and 4.7 govern the average response of consumers to marginal changes in price. In the next section we will consider non-marginal prices changes that would result from soda taxes and we describe heterogeneity in responses.

Table 4.6: *Price effects*

	Effect of 1% price increase on:			
	own	cross demand for:		total
	demand	sugary products	diet products	demand
Coca Cola 330	-2.182	0.141	0.057	0.013
Coca Cola 500	-1.077	0.219	0.089	-0.064
Coca Cola Diet 330	-2.033	0.047	0.153	0.015
Coca Cola Diet 500	-0.916	0.073	0.269	-0.043
Fanta 330	-2.528	0.030	0.011	0.002
Fanta 500	-1.031	0.029	0.014	-0.011
Fanta Diet 500	-0.932	0.012	0.035	-0.008
Cherry Coke 330	-2.538	0.021	0.007	0.002
Cherry Coke 500	-1.125	0.022	0.012	-0.008
Cherry Coke Diet 500	-1.029	0.010	0.028	-0.005
Oasis 500	-1.074	0.050	0.019	-0.018
Oasis Diet 500	-0.975	0.015	0.044	-0.009
Pepsi 330	-2.485	0.056	0.021	0.005
Pepsi 500	-1.670	0.133	0.058	-0.035
Pepsi Diet 330	-2.369	0.017	0.071	0.006
Pepsi Diet 500	-1.412	0.047	0.161	-0.027
Lucozade 380	-1.848	0.104	0.040	-0.001
Lucozade 500	-0.987	0.035	0.019	-0.014
Ribena 288	-2.449	0.029	0.009	0.005
Ribena 500	-1.005	0.019	0.010	-0.008
Ribena Diet 500	-0.948	0.008	0.020	-0.004

Notes: For each product we compute the change in demand for that product, for alternative sugary and diet options and for total demand resulting from a 1% price increase. Numbers are means across time.

Table 4.7: *Own and cross price elasticities for cola*

	Coca Cola				Pepsi			
	330	500	330	500	330	500	330	500
Coca Cola 330	-2.182	0.554	0.147	0.163	0.191	0.322	0.052	0.098
Coca Cola 500	0.110	-1.077	0.033	0.059	0.037	0.102	0.012	0.039
Coca Cola Diet 330	0.181	0.201	-2.033	0.635	0.064	0.122	0.248	0.425
Coca Cola Diet 500	0.036	0.066	0.116	-0.916	0.013	0.044	0.044	0.112
Pepsi 330	0.547	0.536	0.150	0.170	-2.486	0.336	0.058	0.106
Pepsi 500	0.173	0.271	0.053	0.105	0.063	-1.671	0.020	0.063
Pepsi Diet 330	0.167	0.188	0.643	0.635	0.065	0.120	-2.370	0.475
Pepsi Diet 500	0.055	0.108	0.192	0.276	0.021	0.065	0.083	-1.412

Notes: Numbers give percent changes in demand for product in first column associated with a 1% increase in price for products in the first row. Numbers are means across time.

4.4 Soda price elasticity

We also consider how demand for all soda would respond to a marginal increase in the price of all soda, and also how demand for sugary soda would respond to a marginal increase in the price of all sugary soda products. We show these "category level own price elasticities" in Table 4.8.

The own price elasticity for soda is -0.26. This is much smaller than the own price elasticity of any individual soda product. The own price elasticity for sugary soda is slightly larger (in absolute terms) at -0.56. This reflects the fact that some consumers respond to an increase in the price of sugary soda by switching to diet alternatives.

Our estimate of the own price elasticity of sugar soda is very similar to that in Wang (2015). Her estimates are based in US data on food brought into the home. As stockpiling is potentially important for soda brought into the home she builds a dynamic demand model to account for out. In contrast, our estimates are based on purchases for immediate consumption, where stockpiling is by definition ruled out. It is encouraging that both approaches generate similar estimates of overall product category price sensitivity.

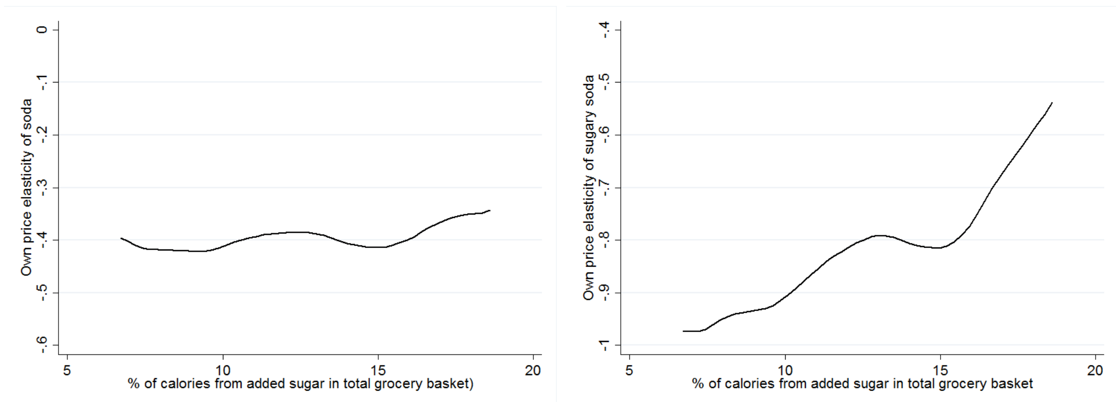
Table 4.8: *Soda own price elasticities*

	Own price elasticity
All soda	-0.41
All sugary soda	-0.84

Notes: Computed as unweighted average across consumers.

Figure 4.5: *Relationship between own price elasticities and added sugar*

(a) Soda (b) Sugary soda



Source: Kantar Worldpanel and Kantar on-the-go panel.

5 The effectiveness of soda taxes

We compare two different forms of tax – a volumetric tax on all soda (a soda tax) and a volumetric tax on sugary soda (a sugary soda tax). A number of US cities have recently legislated for the introduction of either a soda or a sugary soda tax.[8] We simulate the introduction of a tax of 25 pence per litre.[9] Letting τ denote the tax rate (and using Ω_w and Ω_{nw} to denote the set of all sodas and non sodas and Ω_s and Ω_{ns} the set of all sugary and diet sodas), post tax prices for the soda tax are given by:

$$\tilde{p}_{jrt} = \begin{cases} p_{jrt} + \tau s_j & \forall j \in \Omega_w \\ p_{jrt} & \forall j \in \Omega_{nw} \end{cases}$$

and post tax prices for the sugary soda tax are given by:

$$\tilde{p}_{jrt} = \begin{cases} p_{jrt} + \tau s_j & \forall j \in \Omega_s \\ p_{jrt} & \forall j \in \Omega_{ns} \bigcup \Omega_{nw}. \end{cases}$$

Our objective in this paper is to use a very rich demand framework to assess whether different forms of soda tax are well targeted and/or regressive. Firms may respond to such policies by, for instance, changing producer prices, adjusting advertising budgets, reformulating products and discontinuing existing products and introducing new ones. We do not model the behavioural response of firms in the paper, leaving it to future work. Bonnet and Réquillart (2013) have modelled the short term pricing response of firms to the introduction of a soda tax, keeping all other dimensions of firm response fixed, using French data. This paper finds that excise style taxes are overshifted by around 30%. A uniform pass-through rate of 1.3 would imply the tax rate we use would results in price increases of 32.5 pence, or conversely a tax rate of 19 pence would achieve the price increases that we simulate.

5.1 Variation in predictions across preference distribution

Before turning to the aggregate impact of the taxes and how well targeted they are, we demonstrate the effect that modelling rich patterns of correlations in preferences parameters has on the model's predictions. We consider the case of the soda tax and assess the effect the tax has on how much soda a consumer purchases on one given

[8] A soda tax of 1.5 cent per ounce is effective in Philadelphia as of January 2017; a soda tax of 1 cent per ounce is effective in Cook County, Illinois (which includes Chicago) as of June 2017. Berkeley, San Francisco, Oakland, Albany California and Boulder Colorado all legislated for sugary soda taxes of 1 cent per ounce (2 cents in Albany) implemented in 2017-18.

[9] At a pound-dollar exchange rate of 1.25, this corresponds to a tax of 0.93 cents per litre.

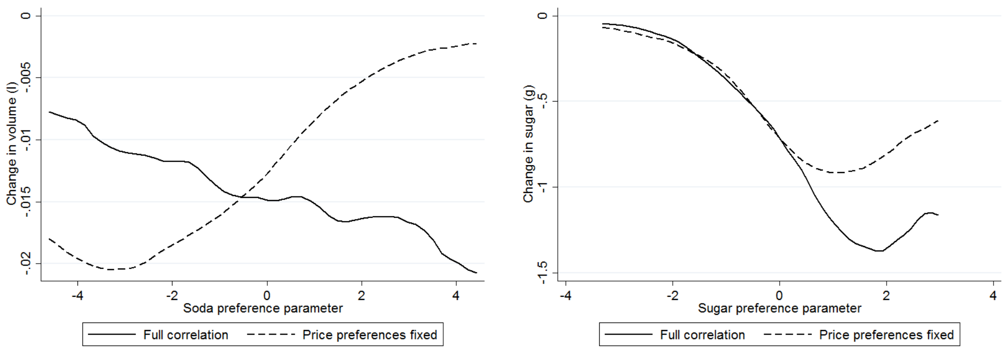
drinks purchase occasion. Panel (a) of Figure 5.1 shows the relationship between reduction in average volume of soda purchased and soda preference parameters and panel (b) shows the relationship between reduction in sugar from soda purchased and sugar preference parameters (for the finite portion of the preference distributions). The solid lines are kernel weighted local polynomial regressions based on predictions from the model. The dashed lines are local polynomial regressions based on predictions from the model when we set each consumers' price coefficient equal to the mean. This has the effect of shutting down the correlation of soda and sugar preference parameters with price preference parameters. The difference between the solid and dashed lines illustrates the importance of this correlation.

Predictions based on full preference parameter correlations show that consumers with strong soda preferences reduce their soda consumption by more (in levels) than those with weaker soda preferences. The model predicts, when we shut off the correlation of soda and price preferences, that the relationship is reversed (so those with strong soda preferences respond the least). The reason for this is that consumers with strong soda preferences tend to also be particularly responsive to price. It is important to account for this correlation in taste to accurately capture price responses across the soda preference distribution.

The relationship between sugar preference parameters and the reduction in sugar purchased as soda shows a similar pattern. The predictions based on the true estimated preference distribution show that consumers with strong sugar preferences respond by more (in levels) than consumers who dislike sugar. Shutting down the correlation in sugar and price preference parameters does not overturn this relationship, but it does make it weaker. The reason for this is that consumers that have strong sugar preferences tend to be relatively price sensitive.

Figure 5.1: *Effect of soda tax on purchase occasion:*

(a) Volume of soda purchased (b) Sugar from soda purchased



Notes: Figure shows the relationship between change in volume of soda and soda preferences (panel (a)) and between change in sugar from soda and sugar preferences (panel (b)) on a purchase occasion following the introduction of a soda tax. Lines are are local polynomial regressions. The solid lines are based on full model estimates, dashed lines are based on estimates when we set consumers' price parameters to the mean value.

5.2 Average impact of tax

Table 5.1 summarises the average impact of the soda and sugary soda tax on sugar purchases. Prior to the tax consumers, on average, obtain 1.46kg of sugar from soda purchased on-the-go annually and they obtain 1.62kg from drinks (soda plus fruit juice and flavoured milk). The soda tax causes a 3.4% fall in sugar from soda and a 2.5% fall in sugar from drinks. The sugary soda tax achieves larger reductions in sugar – inducing a 5.8% fall in sugar from soda and a 4.7% fall in sugar from drinks.

Table 5.2 provides broad details on average switching patterns for both forms of tax. It shows the average change in the volume of sugary soda, diet soda, sugary alternatives to soda and the outside option. The soda tax lowers both sugary soda (by 3.4%) and diet soda (by 2.6%) demand, while increasing demand for sugary non-soda by 5.7%. The sugary soda tax is more effective at reducing overall sugar, since it induces a larger reduction in sugary soda (5.8%) demand and smaller increase in sugar non-soda demand (by 5%) by encouraging switching towards diet products – which see a rise in demand of 3.5%.

It is not surprising that the more targeted sugary soda tax is more effective than a soda tax (with the same rate) at encouraging consumers to switch away from sugar. Reductions in sugar consumption may have associated with it benefits to society through reductions in public health care costs, as well as some benefits to consumers in the future which they may have overlooked at the point of purchase.

Of course, these must be off-set against the costs to consumers of facing higher prices. Table 5.3 summarises the welfare effects of the tax (abstracting from any saving through averted externalities or internalities). The soda tax has an average annual compensating variation per consumer of £6.00 and it also raises £5.91 per consumer in revenue. The narrower base of the sugary soda tax means it has both lower compensating variation (£3.42 per consumer) and raises less revenue (£3.32 per consumer). The fact that tax revenue is close to compensating variation reflects the low price elasticity of demand for soda and sugary soda (see Table 4.8). The flip side of this is that, while each tax raises affected price by around 10% on a average, they induce much smaller falls in sugar.

In Table 5.3 we also report the consumer cost and the net consumer cost per 100g sugar reduction. The consumer cost is based on compensating variation and the net consumer cost is based on compensating variation after tax revenue has been redistributed lump sum to consumers. The consumer cost for the soda tax is £15.07, which is over three times as large as the £4.51 cost associated with the sugary soda tax; the sugary soda tax achieves both a larger reduction in sugar for a considerably smaller consumer welfare loss. Both forms of tax raise an amount of revenue that is only marginally below compensating variation; therefore if tax revenue is lump sum redistributed to consumers the consumer cost per 100g sugar reduction is much less. The net consumer cost (including redistributed revenue) for the soda tax is £0.24, slightly less than double the cost for the sugary soda tax (£0.13). Both the consumer and net consumer costs per 100g sugar reduction indicate that the sugar-soda tax is, from the perspective of consumers, a more cost effective means of lowering sugar.

Table 5.1: *Effect of taxes on grams of sugar from drinks (per year)*

	Total	Change following:	
	pre tax (kg)	soda tax (g)	sugary soda tax (g)
Sugar from soda	1.46	-49.06	-83.88
	[1.42, 1.45]	[-51.95, -45.73]	[-88.40, -76.83]
(%)		-3.36	-5.75
		[-3.57, -3.18]	[-6.15, -5.33]
Sugar from all drinks	1.62	-39.83	-75.89
	[1.59, 1.62]	[-42.03, -38.01]	[-79.61, -69.56]
(%)		-2.46	-4.68
		[-2.60, -2.37]	[-4.96, -4.34]

Notes: Numbers are mean value across soda consumers.

Table 5.2: *% change in demand following tax*

	Soda tax	Sugary soda tax
Sugary soda	-3.42	-5.78
	[-3.62, -3.24]	[-6.16, -5.32]
Diet soda	-2.60	3.53
	[-2.77, -2.40]	[3.14, 3.79]
Sugary non-soda	5.72	4.95
	[4.36, 6.13]	[4.02, 5.35]
Outside option	15.77	9.37
	[14.94, 16.55]	[8.97, 9.82]

Notes: Numbers are mean value across soda consumers.

Table 5.3: *Welfare effects of tax*

	Soda tax	Sugary soda tax
Compensating variation (£ per consumer)	6.00	3.42
	[5.96, 6.01]	[3.38, 3.42]
Tax revenue (£ per consumer)	5.91	3.32
	[5.86, 5.91]	[3.28, 3.32]
Consumer cost per 100g sugar reduction (£)	15.07	4.51
	[14.24, 15.74]	[4.25, 4.89]
Net consumer cost per 100g sugar reduction (£)	0.24	0.13
	[0.23, 0.24]	[0.13, 0.14]

Notes: Consumers costs are based on compensating variation. Net consumer costs are based on compensating variation minus tax revenue (i.e. assuming revenue is redistributed lump sum). Numbers are mean value across soda consumers.

5.3 How well targeted is the tax?

A key factor in determining the effectiveness of tax on soda is whether it targets the people that consume the most sugar. Our demand framework captures both rich correlations in preference parameters and correlations in these preference parameters with other aspect of consumers' diet. It therefore enables us to assess the impact of different forms of soda taxes across the distribution of overall sugar consumption (which we measure as share of calories in all groceries purchases from added sugar). If the externality function from sugar consumption is convexly increasing in added sugar, a successful tax will induce larger falls in sugar consumption among consumers with relatively high added sugar in their diets.

In Table 5.4 we describe how changes in sugar – both in grams and percent terms – resulting from the soda and sugary soda tax vary across the added sugar distribution. The first column shows the mean value of each variable and the subsequent four shows the average deviation in each variable from the mean in each quartile of the added sugar distribution. In Figure 5.2 we also show the relationship graphically as kernel weighted local polynomial regressions.

Neither the soda nor the sugary soda tax are particularly effective at specifically reducing the sugar consumption of the highest sugar consumers. Consumers in the top quartile of the added sugar distribution see a marginally larger (by 3.2 gram) reduction in sugar following the soda tax compared with those in the bottom quartile. However, in percent terms the reduction in the sugar for the top quartile is 0.7 percentage points less than for consumers in the bottom quartile. The sugary soda tax does even less well in specifically targeting the high sugar consumers. Under the sugary soda tax those in the top quartile of the added sugar distribution lower their sugar intake by 10.3 grams (or 2.3 percentage points) less than those in the lowest quartile.

The reason both taxes achieve smaller percentage reductions in sugar among the top quartile of the added sugar distribution can be inferred from the relationship between preference parameters and the share of their calories consumers get from added sugar (see Table 4.4 and Figure 4.3). Consumers with a high share of added sugar in their diets both have systematically stronger preferences for sugar in drinks and are also less sensitive to price changes compared with those with less added sugar in their diets; this leads them to be less willing to switch away from sugary soda when facing higher prices due to tax.

The sugary soda tax achieves larger reductions in sugar across the added sugar distribution. This is because it encourages switching from sugary to diet sodas (unlike the soda tax which increases the price of both). However, the attractiveness of, and hence strength of switching towards, diet soda is greater for consumers in the lower part of the added sugar distribution (who tend to have weaker sugar preferences). Therefore, although the sugary-soda tax does achieve larger reductions in sugar than the soda tax across the added sugar distribution, it also exacerbates the weaker switching away from sugar of those with high added sugar in their diets *relative to* those with lower added sugar.

Table 5.4: *Relationship between change in sugar and share of calories from added sugar*

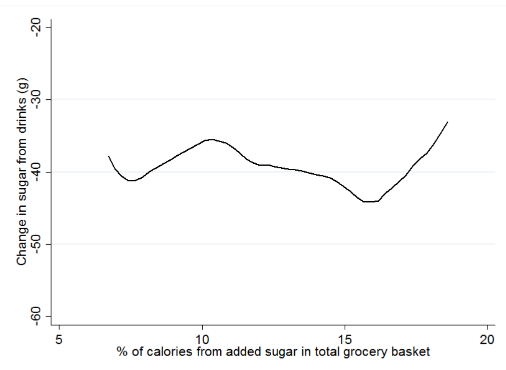
		Mean	Average deviation from mean preference parameter for quartile of added sugar distribution:			
			1	2	3	4
Soda tax	Δ sugar (100g)	-39.83	-2.26	6.69	0.18	-2.01
		[-42.03, -38.01]	[-3.48, -0.45]	[5.59, 7.68]	[-1.07, 1.98]	[-3.59, -0.86]
	(%)	-2.46	-0.60	0.10	-0.01	0.36
		[-2.60, -2.37]	[-0.70, -0.47]	[0.01, 0.17]	[-0.09, 0.10]	[0.27, 0.42]
Soda-sugar tax	Δ sugar (100g)	-75.89	-14.15	14.39	-0.63	2.59
		[-79.61, -69.56]	[-17.54, -9.41]	[11.15, 17.26]	[-3.38, 2.03]	[-0.90, 5.31]
	(%)	-4.68	-1.86	0.31	-0.08	1.01
		[-4.96, -4.34]	[-2.12, -1.52]	[0.13, 0.50]	[-0.24, 0.09]	[0.79, 1.17]

Notes: For each quartile of the distribution of share of calories from added sugar we report the mean deviation from the average value for each variable shown in the first column. 95% confidence intervals are given in brackets.

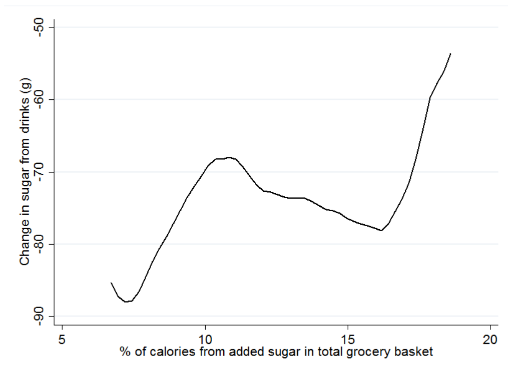
Figure 5.2: *Effect of sugar across added sugar distribution*

Change in sugar from drinks (in grams)

(a) Soda tax

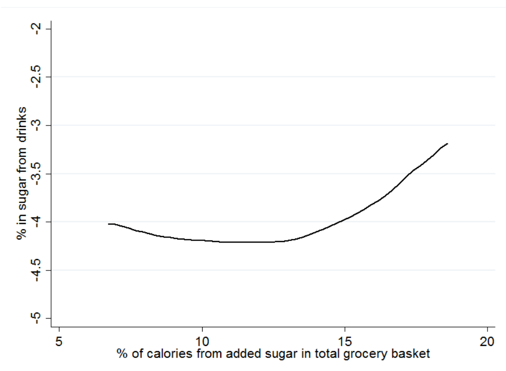


(b) Sugary soda tax

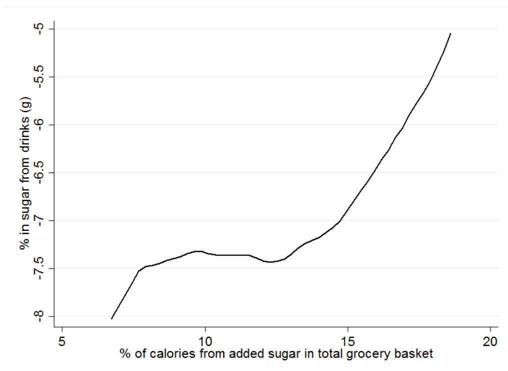


% Change in sugar from drinks

(c) Soda tax



(d) Sugary soda tax



Notes: Lines are local polynomial regressions.

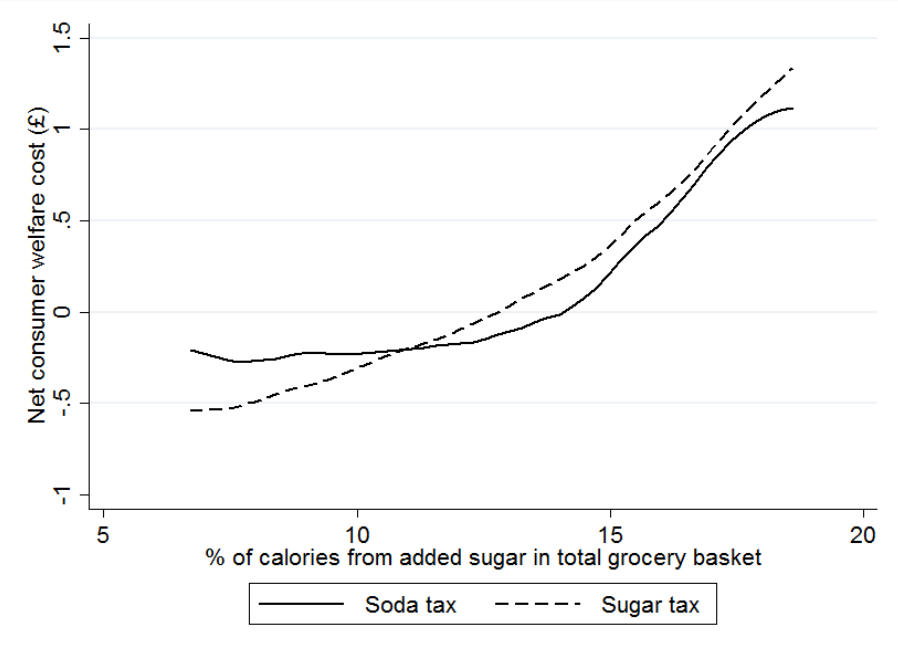
Table 5.5 and Figure 5.3 describe how the consumer welfare burden of the two taxes varies across the added sugar distribution. The consumer welfare costs are based on compensating variation minus a lump sum rebate (equal across consumers) from tax revenue. Considering the net consumer welfare costs leads the two forms of tax to have very similar mean welfare effects (and in the case of the sugary soda tax, for consumers with low added sugar in their diet to see a welfare rise), however the lump sum tax rebate does not change the shape of how welfare effects vary across the distribution. The soda tax places a larger burden on consumers in the top quartile of the added sugar distribution than those in lower quartiles (with the relationship between consumer welfare loss and added sugar declining across the these quartiles). The relationship between consumer welfare loss and added sugar is starker in the case of the sugary soda tax. In this case the welfare burden rises strongly across the added sugar distribution; consumers that get a higher fraction of their calories from added sugar tend to have stronger preferences for sugar, they tend to consume more sugary soda and consequently they see the largest fall in welfare as a consequence of a tax levied on the sugar in soda.

Table 5.5: *Relationship net compensating variation and share of calories from added sugar*

	Mean	Average deviation from mean preference parameter for quartile of added sugar distribution:			
		1	2	3	4
Soda tax	0.09	-0.12	-0.31	-0.29	0.70
	[0.09, 0.10]	[-0.13, -0.09]	[-0.33, -0.29]	[-0.31, -0.27]	[0.67, 0.72]
Soda-sugar tax	0.10	-0.56	-0.45	-0.03	0.96
	[0.09, 0.11]	[-0.57, -0.53]	[-0.46, -0.42]	[-0.05, -0.02]	[0.92, 0.97]

Notes: For each quartile of the distribution of share of calories from added sugar we report the mean deviation from the average value for each variable shown in the first column. 95% confidence intervals are given in brackets. Net compensating variation is compensating variation minus tax revenue (i.e. assuming revenue is redistributed lump sum). Numbers are mean value across soda consumers.

Figure 5.3: *Net consumer welfare loss*



Notes: Lines are local polynomial regressions. Net compensating variation is compensating variation minus tax revenue (i.e. assuming revenue is redistributed lump sum). Numbers are mean value across soda consumers.

5.4 Distributional effect

Section 4.2 shows that consumers with low equivalised household annual grocery expenditures tend to be relatively price sensitive when making their on-the-go drinks purchases and also tend to have relatively strong preferences for soda. This preference variation drives heterogeneity in responses to the soda and sugary soda tax across the grocery expenditure distribution. In Table 5.6 and Figure 5.4 we show this variation.

Consumers in the bottom quartile of the equivalised grocery expenditure respond most strongly to both forms of tax. Under the soda tax this group lower their sugar consumption by 0.8 percentage points (25.8 grams) more than those in the top quartile of the distribution. Similarly, the sugary soda tax induces a 0.9 percentage points (30.8 grams) larger reduction in sugar consumption for those in the bottom quartile of the expenditure distribution compared with the top quartile.

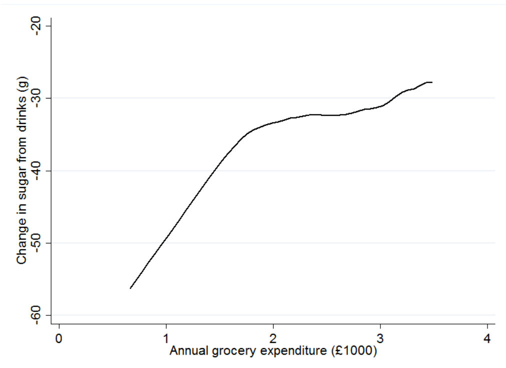
Table 5.6: *Relationship between change in sugar and annual grocery expenditure*

		Mean	Average deviation from mean preference parameter for quartile of equivalised grocery distribution: 1	2	3	4
Soda tax	Δ Sugar (100g)	-39.83	-17.12	1.54	8.02	7.90
		[-42.03, -38.01]	[-17.73, -14.36]	[0.86, 3.48]	[7.83, 10.02]	[5.17, 8.40]
	(%)	-2.46	-0.83	0.34	0.45	0.06
		[-2.60, -2.37]	[-0.72, -0.54]	[0.34, 0.49]	[0.29, 0.43]	[-0.28, -0.02]
Soda-sugar tax	Δ Sugar (100g)	-75.89	-18.00	0.59	7.89	11.33
		[-79.61, -69.56]	[-19.10, -11.24]	[-2.71, 2.43]	[5.57, 11.36]	[7.86, 14.66]
	(%)	-4.68	-0.75	0.52	0.39	-0.17
		[-4.96, -4.34]	[-0.57, -0.13]	[0.42, 0.71]	[-0.11, 0.29]	[-0.66, -0.11]

Notes: For each quartile of the equivalised annual grocery expenditure we report the mean deviation from the average value for each variable shown in the first column. 95% confidence intervals are given in brackets.

Figure 5.4: *Effect of sugar across grocery expenditure distribution*

(a) Soda tax



(b) Sugar-soda tax

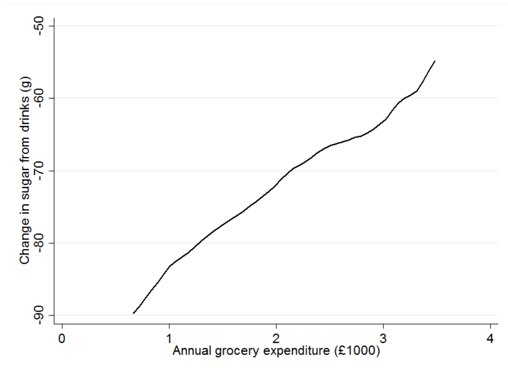
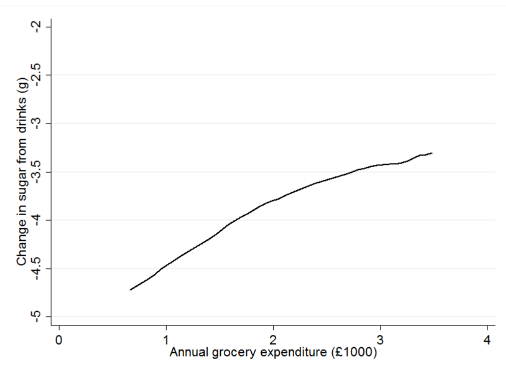
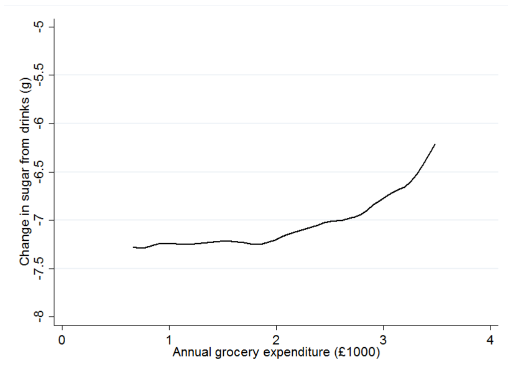


Figure 5.5: *% Change in sugar from drinks*

(a) Soda tax



(b) Sugar-soda tax



Notes: Lines are local polynomial regressions.

The relationship between net consumer welfare costs and total expenditure is similar for both forms of tax (see Table 5.7 and Figure 5.6). In each case the consumer welfare burden is concentrated among the lower spending (poorest) consumers. This is driven by the fact that this group purchase more soda and more sugary soda than richer households. If tax revenue is redistributed lump sum to consumers, consumers in the bottom half of the grocery expenditure distribution will still see a reduction in welfare (their net consumer welfare cost is positive). Higher spending consumers, on the other hand, would actually see an increase in their welfare.

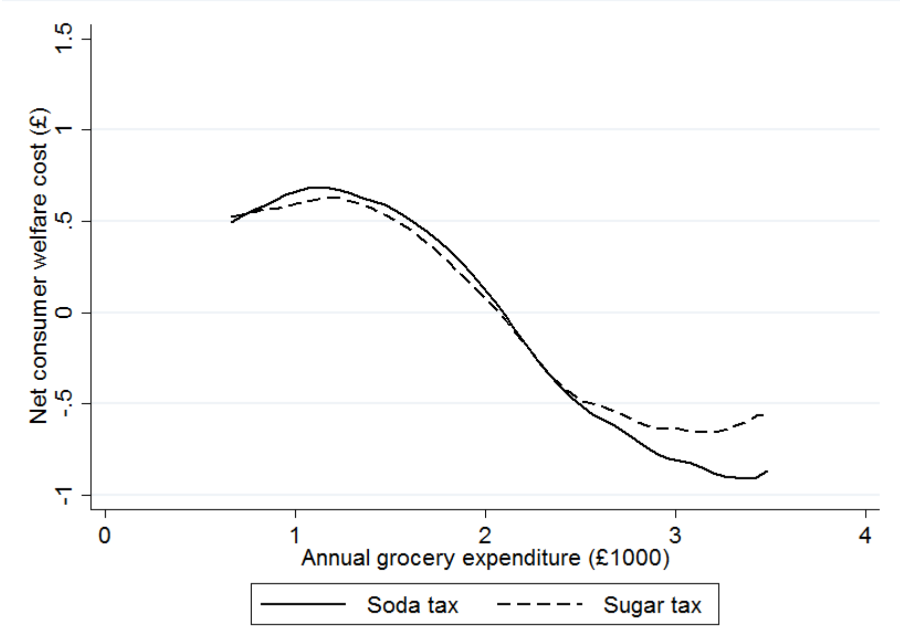
Based purely on net compensating variation, soda taxes appear regressive (a result also found by Wang (2015)). However, a complete assessment of the distributional impact of soda taxes must also account for any averted internalities due to the tax. Poor consumers lower their sugar consumption by more than better off households due to the tax. They also potentially suffer more from problems of under-weighting the future costs of their actions. In the long run therefore it is possible that soda taxes would be less regressive than an assessment based only on compensating variation suggests.

Table 5.7: *Relationship between preference parameters and share of calories from added sugar*

	Mean	Average deviation from mean preference parameter for quartile of equivalised grocery distribution:			
		1	2	3	4
Soda tax	0.09	-0.03	0.60	-0.08	-0.46
	[0.09, 0.10]	[0.24, 0.29]	[0.78, 0.82]	[-0.33, -0.30]	[-0.83, -0.78]
Soda-sugar tax	0.10	0.20	0.49	-0.06	-0.68
	[0.09, 0.11]	[0.39, 0.43]	[0.53, 0.57]	[-0.29, -0.25]	[-0.78, -0.74]

Notes: For each quartile of the equivalised annual grocery expenditure we report the mean deviation from the average value for each variable shown in the first column. 95% confidence intervals are given in brackets.

Figure 5.6: *Compensating variation*



Notes: Lines are local polynomial regressions. Net compensating variation is compensating variation minus tax revenue (i.e. assuming revenue is redistributed lump sum). Numbers are mean value across soda consumers.

6 Robustness

6.1 Bias correction for incidental parameters problem

Our maximum likelihood estimate of the parameters may suffer from an incidental parameters problem. Even if both $N \to \infty$ and $T \to \infty$, if N and T grow at the same rate ($\frac{N}{T} \to \rho$ where ρ is a non zero constant), our fixed effect estimator will be asymptotically biased (Arellano and Hahn (2007)). Bias correction methods exist that reduce the bias from being of order $1/T$ to $1/T^2$.

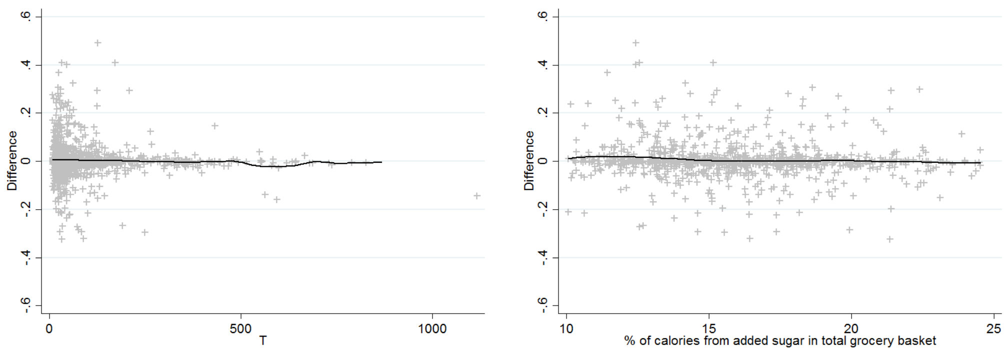
There are a set of analytical bias correction methods that involve correcting the estimator directly or correcting the moment conditions from which the estimator is derived (see survey of Arellano and Hahn (2007), Arellano and Bonhomme (2011)). An alternative approach is based on panel jackknife methods (Hahn and Newey (2004)). We use the split sample jackknife bias correction procedure suggested in Dhaene and Jochmans (2015). This entails splitting the sample in two and using the sub-sample estimates as an adjustment to correct the full sample maximum likelihood estimate (see Appendix B for more details.)[10]

[10]For some consumers, some parameters are not identified in one of the two subsamples – for instance if the sampling is such that all of a consumer's outside option purchases happen to be

In Figure 6.1 we graph the difference between the bias corrected and maximum likelihood sugar preference parameters (in Appendix B we show similar figures for other preference parameters). Panel (a) shows how this difference relates to the time a consumer is in the sample and panel (b) show the relationship with total calories from added sugar in consumers' diet. The figure shows that the difference between to the two estimates is relatively small; the standard deviation of the sugar preference parameter estimates is over 2, while the the average absolute difference between the bias corrected and maximum likelihood estimates is 0.08. Panel (a) shows that the difference is decreasing in T; those in the sample for a relatively short number of choice occasions on average have higher bias than those in the sample relatively many times. However, it also shows that, conditional on T, the average difference between the bias corrected and maximum likelihood estimates is zero – a positive difference is equally likely as negative distance. Indeed the distribution of the maximum likelihood and bias corrected estimates of the preference parameters are almost indistinguishable (see Figure 6.2). Panel (b) of Figure 6.1 shows the difference between the maximum likelihood and bias corrected estimates is completely unrelated to our measure of how much added sugar individuals have in the their overall diet.

Figure 6.1: *Bias for sugar preference parameters*

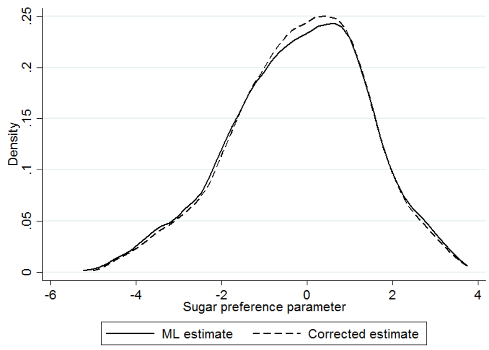
(a) Relationship with T (b) Relationship with added sugar



Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

in one subsample and all their inside option purchases are in the other. We implement the bias correction procedure on those consumers for which such an issue does not arise. In total they account for over 75% of choice occasions in our data.

Figure 6.2: *Distribution of sugar preference parameters*



Notes: Lines are kernel density estimates.

In Appendix B we show similar results for estimated price and soda preferences; the bias correction results in an almost identical preference distribution, any individual level bias is relatively small, is equally likely to be positive as negative and there is no relationship with the key variables we relate our demand effects to.

6.2 Substitution to other forms of sugar

The choice model we outline in Section 3.3 provides a rich framework for capturing consumer substitution in the drinks market. When we use this to assess the impact of soda taxes we are therefore able to capture realistic switching patterns across soda products and from soda to alternative drinks. One possible concern is that some consumers may respond to soda taxes by switching to sugar in food. To consider the possibility we extend our choice model to capture switch between drinks and chocolate, a leading alternative source of sugar.

Suppose the choice model of Section 3.3 is a second stage of a two stage decision process, which governs, conditional on choosing a drink, which drink to select. Consider a first stage in which the consumer chooses between chocolate products, choosing a non-sugary snack and choosing to select a drink. Let $k = \{\varnothing, 1, ..., K, \mathcal{D}\}$ denote first stage options. $k = \varnothing$ denotes the first stage outside option of a non-sugar snack, $k = 1, ..., K$ indexes chocolate products and $k = \mathcal{D}$ indexes choosing a drink (with the specific drinks product determined by the second stage of the decision problem). Suppose utility from these options takes the form:

$$\begin{aligned} V_{i\varnothing t} &= \varepsilon_{it} \\ V_{ikt} &= \mu_c + W_{ikt} + \varepsilon_{ikt} \quad \text{for all } k \in \{1, ..., K\} \\ V_{i\mathcal{D}t} &= \mu_d + \psi_{\mathcal{D}} W_{i\mathcal{D}t} + \varepsilon_{i\mathcal{D}t}, \end{aligned}$$

where $W_{i\mathcal{D}t}$ is the expected utility from choosing a drink product and can be computed using estimates of the second stage choice model (see Appendix C) and where $W_{ikt} = \alpha_i p_{krt} + \beta_i s_k + \zeta_{b(k)}$ is product specific utility from choosing chocolate product j. We assume that the error terms, $(\varepsilon_{i0t}, \varepsilon_{i1t}, ..., \varepsilon_{iKt}, \varepsilon_{i\mathcal{D}t})$ are distributed i.i.d. extreme value. This extends our choice model to capture switching between drinks, chocolates and non-sugar snacks and allows us to estimate the strength of switching between soda and chocolate (see Appendix C for further details).

[Results to be included]

7 Summary and conclusion

Our analysis suggests that the sugary soda tax leads to greater reductions in sugar consumption than the soda tax across the entire added sugar distribution. It does this while reducing overall consumer surplus (net of tax revenue) by less than the soda tax.

The sugary soda tax encourages stronger switching away from the sugar in soda and less switching towards the sugar in alternative drinks than the soda tax. The narrower base of the sugar soda tax leads it both to raise less revenue and to create smaller losses in consumer surplus, with these two effects roughly balancing.

The percent reduction in sugar due to the sugary soda tax declines across the added sugar distribution (the level reduction is about constant) – consumers with relatively high added sugar in their diets reduce their sugar intake in percentage terms by less. This is driven both by a particularly strong reluctance by high added sugar individuals to lower their sugary soda demands (due to strong sugar preferences and their being relatively price insensitive) and a higher willingness to switch to sugary alternatives.

References

Aguiar, M. and E. Hurst (2007). Life-Cycle Prices and Production. *American Economic Review 97*(5), 1533–1559.

Allcott, H., S. Mullainathan, and D. Taubinsky (2014, April). Energy policy with externalities and internalities. *Journal of Public Economics 112*, 72–88.

Arellano, M. and S. Bonhomme (2011). Nonlinear Panel Data Analysis. *Annual Review of Economics 3*(1), 395–424.

Arellano, M. and J. Hahn (2007). Understanding bias in nonlinear panel models: Some recent developments. In *Econometric Society Monographs*, Volume 43, pp. 381.

Bajari, P. and C. L. Benkard (2005). Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy 113*(6), 1239–1276.

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), 459–463.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from combination of micro and macro data: The new car market. *Journal of Political Economy 112*(1), 68–105.

Bonnet, C. and V. Réquillart (2013). Tax incidence with strategic firms in the soft drink market. *Journal of Public Economics 106*, 77–88.

Browning, M. and P. A. Chiappori (1998). Efficient Intra-Household Allocation : A General Characterization and Empirical Tests. *Econometrica 66*(6), 1241–1278.

Burda, M., M. Harding, and J. Hausman (2008). A Bayesian mixed logit-probit model for multinomial choice. *Journal of Econometrics 147*(2), 232–246.

Cawley, J. and D. Frisvold (2016). The Incidence of Taxes on Sugar-sweetened Beverages: The case of Berkeley, California. *Journal of Policy Analysis and Management*.

CDC (2016, September). Cut Back on Sugary Drinks. http://www.cdc.gov/nutrition/data-statistics/sugar-sweetened-beverages-intake.html.

Dhaene, G. and K. Jochmans (2015, July). Split-panel Jackknife Estimation of Fixed-effect Models. *The Review of Economic Studies 82*(3), 991–1030.

Dubois, P., R. Griffith, and A. Nevo (2014). Do Prices and Attributes Explain International Differences in Food Purchases? *American Economic Review 104*(3), 832–867.

Griffith, R., M. Lürhmann, M. O'Connell, and K. Smith (2016). Using taxation to reduce sugar consumption. IFS Briefing Note BN180, IFS, London.

Griffith, R., M. O'Connell, and K. Smith (2017, January). Design of optimal corrective taxes in the alcohol market. *IFS Working Paper W17/02*.

Gruber, J. and B. Koszegi (2004). Tax incidence when individuals are time-inconsistent: The case of cigarette excise taxes. *Journal of Public Economics 88*(9-10), 1959–1987.

Haavio, M. and K. Kotakorpi (2011, May). The political economy of sin taxes. *European Economic Review 55*(4), 575–594.

Hagenaars, A. J., K. De Vos, M. Asghar Zaidi, and others (1994). Poverty statistics in the late 1980s: Research based on micro-data.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*(4), 1295–1319.

Han, E. and L. M. Powell (2013, January). Consumption patterns of sugar sweetened beverages in the United States. *Journal of the Academy of Nutrition and Dietetics 113*(1), 43–53.

Harding, M. and M. Lovenheim (2014). The Effect of Prices on Nutrition: Comparing the Impact of Product- and Nutrient-Specific Taxes. Discussion Paper 13-023, Stanford Institute for Economic Policy Research.

Hendel, I. and A. Nevo (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica 74*(6), 1637–1673.

Lewbel, A. and K. Pendakur (2017). Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients. *Journal of Political Economy*.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Neyman, J. and E. L. Scott (1948, January). Consistent Estimates Based on Partially Consistent Observations. *Econometrica 16*(1), 1.

O'Donoghue, T. and M. Rabin (2006). Optimal sin taxes. *Journal of Public Economics 90*(10-11), 1825–1849.

Pigou, A. C. (1920). *The Economics of Welfare*. McMillan&Co., London.

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

Wang, E. Y. (2015, June). The impact of soda taxes on consumer welfare: Implications of storability and taste heterogeneity. *The RAND Journal of Economics 46*(2), 409–441.

Welsh, J., A. Sharma, L. Grellinger, and M. Vos (2011). Consumption of added sugars is decreasing in the United States. *American Journal of Clinical Nutrition 94*(3), 726–734.

WHO (2015). Sugars intake for adults and children.

Woodward-Lopez, G., J. Kao, and L. Ritchie (2010). To what extent have sweetened beverages contributed to the obesity epidemic? *Public Health Nutrition 14*(3), 499–509.

A Robustness to the large support assumption

In order to identify infinite preference parameters we rely on the "large support" assumption on the error terms and the assumption that our large T leads to asymptotic estimation. We show here that our estimation and counterfactual results are robust to the fact that T is not infinite, while keeping the parametric assumption of extreme value distribution of error terms.

To do this we use the estimated model and we simulate a series of 100 consumer choices on 100 purchase occasions for a finite soda preference γ_i and look at the probability over these 100 simulations that the consumer never bought soda. We vary γ_i to evaluate how the probability of no soda purchase increases with γ_i.

This exercise shows clearly that ... [**to be completed**]

B Incidental parameters problem

Non linear models with fixed effects give rise to an incidental parameters problem, noted by Neyman and Scott (1948). The problem is that maximum likelihood estimates of parameters are typically not consistent under asymptotics where N tends to infinity and T is fixed. The reason is only a finite number of observations are available to estimate each fixed effect, meaning the estimation error for the fixed effects remains as the sample grows. In our case, we have relatively large T, typically dozens of observations per consumer. However, even asymptotics where both N and T tend to infinity still do not necessarily solve the incidental parameters problem (see, for instance, Hahn and Newey (2004)).

We therefore implement the split panel jackknife suggested in Dhaene and Jochmans (2015). This entails obtaining estimates of the model parameters $\theta = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \eta)$ based on splitting the sample into two non overlapping random sample. Each sample contains one half of the choice occasions for each individual. We denote the maximum likelihood estimate for the full sample $\widehat{\theta}$ and the estimate for the two subsamples $\widehat{\theta}_{(1,T/2)}$ and $\widehat{\theta}_{(T/2,T)}$. The bias corrected estimator is:

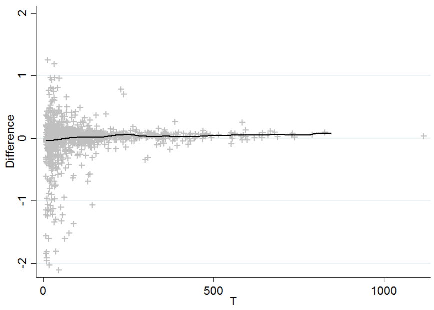
$$\widetilde{\theta}_{split} = 2\widehat{\theta} - \frac{\widehat{\theta}_{(1,T/2)} + \widehat{\theta}_{(T/2,T)}}{2} \tag{B.1}$$

Figures B.1-B.3 show, for the price, soda and sugar preference parameters, how the difference between the bias corrected estimate ($\widetilde{\theta}_{split}$) and the maximum likelihood estimate ($\widehat{\theta}$) relate to a) the time individuals are in the sample, b) the total added sugar in their diets and c) their total grocery expenditure. They show no systematic relationship in the mean of ($\widetilde{\theta}_{split} - \widehat{\theta}$) with any of these variables, with the dispersion of ($\widetilde{\theta}_{split} - \widehat{\theta}$) falling in T.

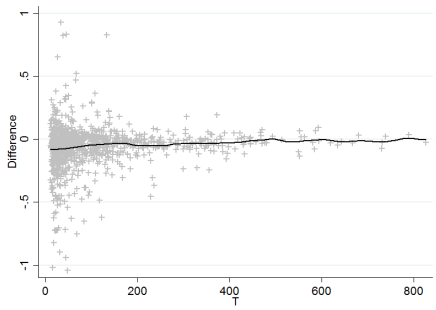
Figures B.4 plots the distributions of price, soda and sugar preference parameter estimates for both the estimators $\widehat{\theta}$ and $\widetilde{\theta}_{split}$, showing there is very little difference in the distributions.

Figure B.1: *Relationship between bias and time in sample*

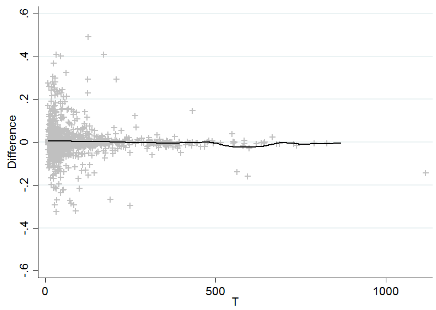
(a) Price



(b) Soda



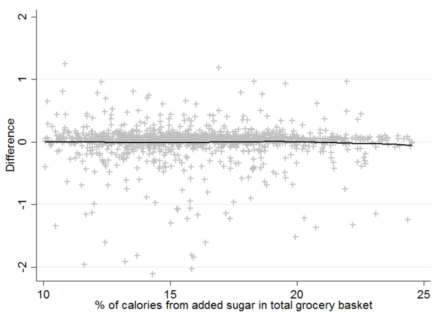
(c) Sugar



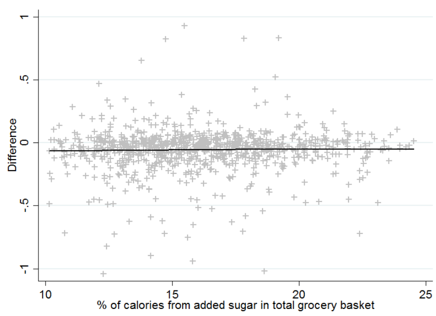
Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

Figure B.2: *Relationship between bias and added sugar*

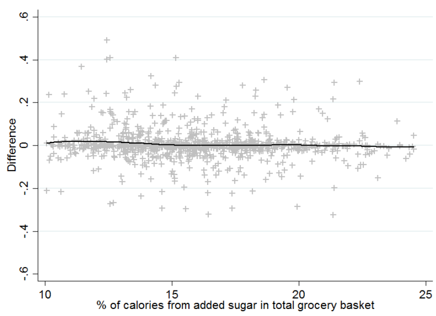
(a) Price



(b) Soda



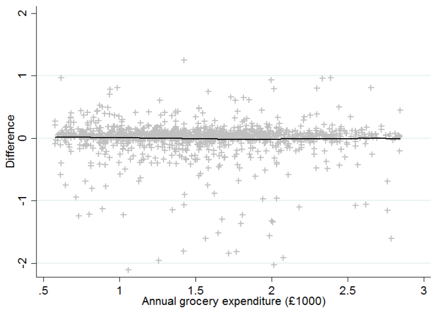
(c) Sugar



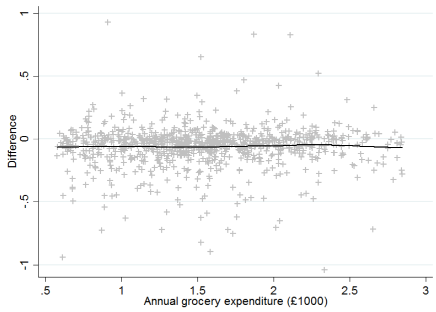
Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

Figure B.3: *Relationship between bias and grocery expenditure*

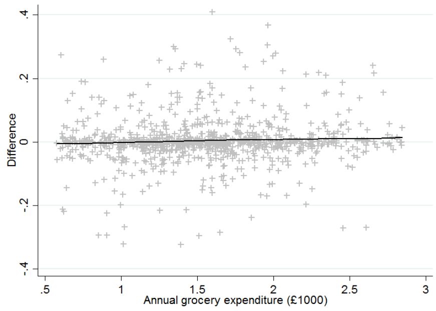
(a) Price



(b) Soda



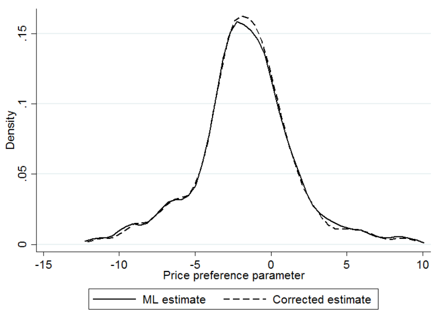
(c) Sugar



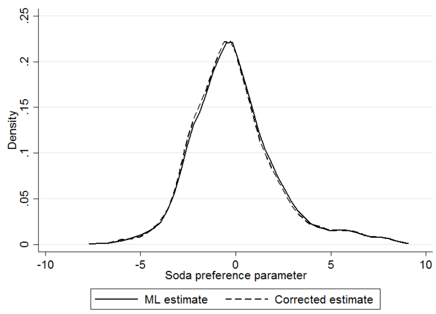
Notes: Lines are local polynomial regressions.

Figure B.4: *Preference parameter distribution*

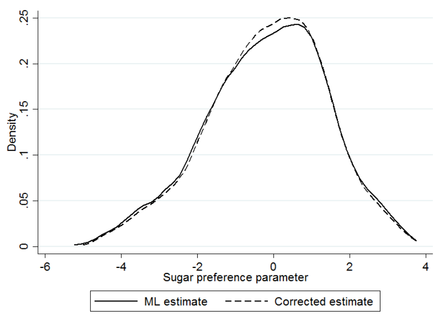
(a) Price



(b) Soda



(c) Sugar



Notes: Lines are kernel density estimates.

C Estimating switching from drinks to other forms of sugar

The choice model we outline in Section 3.3 captures consumer choice between drink products $j = \{0, 1, ..., J\} = \Omega_{\mathcal{D}}$. The drink products comprise water $j = 0$, soda, $j = \{1, ..., j'\} = \Omega_w$ and non-soda juice $j = \{j' + 1, ..., J\} = \Omega_{nw}$ The expected utility to the consumer of purchasing a drink is:

$$E_{\epsilon_{ijt}} \left[\max_{j \in \Omega_{\mathcal{D}}} U_{ijt} \right] = \ln \left(\exp \zeta_{d0rt} + \sum_{j \in \Omega_w \cup \Omega_{nw}} \exp \left(\alpha_i p_{jrt} + \beta_i s_j + \gamma_i w_j + g(\mathbf{x}_{jt}; d, \eta) \right) \right)$$
$$\equiv W_{i\mathcal{D}t}.$$

Consider a first stage decision in which the consumer chooses between options $k = \{\varnothing, 1, ..., K, \mathcal{D}\}$, where $k = \varnothing$ denotes the outside option of a non-sugar snack, $k = \{1, ..., K\} = \Omega_c$ indexes chocolate products and $k = \mathcal{D}$ indexes choosing a drink. Suppose utility from these options takes the form:

$$\begin{aligned} V_{i\varnothing t} &= \varepsilon_{it} \\ V_{ikt} &= \mu_c + W_{ikt} + \varepsilon_{ikt} \quad \text{for all } k \in \Omega_c \\ V_{i\mathcal{D}t} &= \mu_d + \psi_{\mathcal{D}} W_{i\mathcal{D}t} + \varepsilon_{i\mathcal{D}t}, \end{aligned}$$

where

$$W_{ikt} = \alpha_i p_{krt} + \beta_i s_k + \zeta_{b(k)}$$

and $(\varepsilon_{i0t}, \varepsilon_{i1t}, ..., \varepsilon_{iKt}, \varepsilon_{i\mathcal{D}t})$ are distributed i.i.d. extreme value. Note the nesting of the errors terms – consumers get a draw of first stage error terms $\boldsymbol{\varepsilon}$ and if they choose $k = \mathcal{D}$, they get a draw of second stage errors, $\boldsymbol{\epsilon}$, when selecting what drink product to choose.

This first stage choice probabilities are:

$$\begin{aligned} P_{it}(k = 0) &= \frac{1}{1 + \sum_{k' \in \Omega_c} \exp(\mu_c + W_{ik't}) + \exp(\mu_{\mathcal{D}} + \psi_d W_{i\mathcal{D}t})} \\ P_{it}(k = \mathbf{k}) &= \frac{\exp(\mu_c + W_{i\mathbf{k}t})}{1 + \sum_{k' \in \Omega_c} \exp(\mu_c + W_{ik't}) + \exp(\mu_{\mathcal{D}} + \psi_d W_{i\mathcal{D}t})} \quad \text{for all } k \in \Omega_c \\ P_{it}(k = \mathcal{D}) &= \frac{\exp(\mu_{\mathcal{D}} + \psi_{\mathcal{D}} W_{i\mathcal{D}t})}{1 + \sum_{k' \in \Omega_c} \exp(\mu_c + W_{ik't}) + \exp(\mu_{\mathcal{D}} + \psi_d W_{i\mathcal{D}t})}. \end{aligned}$$

The second stage drinks choice model allows us to identify the drinks inclusive value, $W_{i\mathcal{D}t}$, and the preference parameters (α_i, β_i). Let Ω_c^B denote the set of chocolate brands and ω_b be the set of chocolate products that belong to brand b. The second stage model also enables us to identify the chocolate brand indices:

$$z_{ibt} = \ln \sum_{k \in \omega_b} \exp \left[\alpha_i p_{krt} + \beta_i s_k \right].$$

Note that

$$\sum_{k\in\Omega_c}\exp\left(\mu_c+W_{ikt}\right)=\sum_{b\in\Omega_c^B}\sum_{k\in\omega_b}\exp\left(\mu_c+W_{ikt}\right)$$
$$=\sum_{b\in\Omega_c^B}\sum_{k\in\omega_b}\exp\left(\mu_c+\left[\alpha_i p_{krt}+\beta_i s_k+\zeta_{b(k)}\right]\right)$$
$$=\sum_{b\in\Omega_c^B}\exp\left(\tilde{\zeta}_b+z_{ibt}\right),$$

where $\tilde{\zeta}_b=\mu_c+\zeta_b$ so that the first stage purchase probabilities can be written:

$$P_{it}(k=0)=\frac{1}{1+\sum_{b'\in\omega_b}\exp\left(\tilde{\zeta}_{b'}+z_{ib't}\right)+\exp\left(\mu_{\mathcal{D}}+\psi_d W_{i\mathcal{D}t}\right)}$$
$$P_{it}(k\in\omega_b)=\frac{\exp\left(\tilde{\zeta}_b+z_{ibt}\right)}{1+\sum_{b'\in\omega_b}\exp\left(\tilde{\zeta}_{b'}+z_{ib't}\right)+\exp\left(\mu_{\mathcal{D}}+\psi_d W_{i\mathcal{D}t}\right)}\quad\text{for all } b\in\Omega_c^b$$
$$P_{it}(k=\mathcal{D})=\frac{\exp\left(\mu_{\mathcal{D}}+\psi_{\mathcal{D}} W_{i\mathcal{D}t}\right)}{1+\sum_{b'\in\omega_b}\exp\left(\tilde{\zeta}_{b'}+z_{ib't}\right)+\exp\left(\mu_{\mathcal{D}}+\psi_d W_{i\mathcal{D}t}\right)}.$$

Given identified parameters from the second stage and data on decisions consumers make over purchases of chocolate products, drinks or other snacks, the first stage choice model allows us to identify the remaining parameters $\tilde{\boldsymbol{\zeta}}=(\tilde{\zeta}_1,...,\tilde{\zeta}_B)'$, $\mu_{\mathcal{D}}$ and $\psi_{\mathcal{D}}$.

The probability of buying any chocolate is given by:

$$P_{it}(k\in\Omega_c)=\frac{\sum_{b\in\Omega_c^b}\exp\left(\tilde{\zeta}_b+z_{ibt}\right)}{1+\sum_{b\in\Omega_c^b}\exp\left(\tilde{\zeta}_{b'}+z_{ib't}\right)+\exp\left(\mu_{\mathcal{D}}+\psi_d W_{i\mathcal{D}t}\right)}.$$

The change in the probability of purchasing chocolate in response to a marginal increase in the price of all soda products is then:

$$\frac{\partial P_{it}(k\in\Omega_c)}{\partial p_{wt}}=-\psi_d\ \alpha_i\ \underbrace{P_{it}(k=\mathcal{D})}_{\substack{\text{probability of}\\ \text{buying any drink}}}\ \underbrace{P_{it}(k\in\Omega_c)}_{\substack{\text{probability of}\\ \text{buying any chocolate}}}\ \underbrace{P_{it}\left(j\in\Omega_w|k=\mathcal{D}\right)}_{\substack{\text{probability of buying}\\ \text{any soda given first}\\ \text{stage choice of drinks}}},$$

and the change in the probability of purchasing chocolate in response to a marginal increase in the price of all sugary soda products is:

$$\frac{\partial P_{it}(k \in \Omega_c)}{\partial p_{st}} = -\psi_d \, \alpha_i \underbrace{P_{it}(k = \mathcal{D})}_{\substack{\text{probability of} \\ \text{buying any drink}}} \underbrace{P_{it}(k \in \Omega_c)}_{\substack{\text{probability of} \\ \text{buying any chocolate}}} \underbrace{P_{it}\left(j \in \Omega_s | k = \mathcal{D}\right)}_{\substack{\text{probability of buying} \\ \text{any sugary soda given first} \\ \text{stage choice of drinks}}},$$

where $\Omega_s = \{j | j \in \Omega_w, s_j = 1\}$ denotes the set of sugary sodas.[11]

[11]Note, we use p_{wt} and p_{st} to denote a common component in the price of all soda and sugary soda respectively. For instance, we can generically write the price of a soda product j as $p_{jrt} = p_{wt} + \tilde{p}_{jrt}$ where p_{wt} is some common component across soda products and $\tilde{p}_{jrt}$ is the product-market specific deviation from this.