Individual preference heterogeneity, targeting and welfare effects of soda taxes

Pierre Dubois, Rachel Griffith and Martin O'Connell*

September 11, 2017

Abstract

Soda taxes aim to reduce excessive sugar consumption. Their effectiveness depends on whether they successfully target those whose consumption is associated with the highest harm. We assess the impact of soda taxes using novel longitudinal data on purchases on-the-go; we model the supply side of the market and find that pass-through is over 100%. We recover individual level responses, which we relate to markers of the likely harm to the individual from sugar consumption. We show that soda taxes are not well targeted at those with high sugar diets; such individuals do not respond any more strongly than those with low sugar diets. However, younger consumers switch relatively strongly away from sugar. We evaluate the welfare and redistributive properties of the tax.

Keywords: preference heterogeneity, internalities, 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

Sugar consumption is far in excess of recommended levels in much of the developed world, and is strongly linked with a range of diet-related diseases, including diabetes, cancers and heart disease, and is thought to be particularly detrimental to children (WHO (2015)). Soda is an important contributor to excess sugar consumption (CDC (2016)) particularly in the young (Han and Powell (2013) and Cavadini et al. (2000)). Corrective taxes have been advocated as an effective mechanism to reduce consumption associated with internalities, costs that are borne by the individual consumer themself in the future, but that are ignored at the point of consumption (Gruber and Koszegi (2004), O'Donoghue and Rabin (2006), Haavio and Kotakorpi (2011), Allcott et al. (2014)). Concerns over excess sugar intake have led to the introduction of taxes on soda in a number of jurisdictions.[1] The effectiveness of such taxes depends on how individuals' demand responses correlate with the marginal harm associated with their consumption.

Our contribution in this paper is to provide evidence on how well targeted soda taxes are; in particular, are they effective in lowering the sugar consumption of individuals for whom the marginal harm of consumption (their internality) is largest? Relative to the existing literature we make two main advances.

First, and most importantly, we depart from the standard approach to modelling consumer preference heterogeneity in choice models by treating the preferences as parameters to be estimated (rather than random effects drawn from a mixing distribution). This means we can directly relate predictions of the impact of tax on individuals to their characteristics in a very flexible way; this is crucial to be able to assess how well targeted the tax is.

Second, we study individual purchase decisions made for immediate consumption "on-the-go" using novel longitudinal data on a representative sample of UK individuals (including children). Most of the literature studies choice behaviour in supermarkets, where purchases are taken home so decisions are made for future consumption. Around half of soda purchases are made "on-the-go", and decisions made for immediate consumption are where consumers are most likely to suffer from problems of self-control and temptation, and where internalities (a divergence between short-run and long-run preferences) are likely to be largest.

We use a discrete choice demand model in which consumer preferences are defined over product attributes. Like much of the literature on choice models (Berry

[1]A number of US cities, including Philadelphia, San Francisco and Chicago, in addition to France, Mexico and the UK, either have introduced or are planning to introduce taxes levied on soda.

et al. (1995), Nevo (2001), Train (2003)), we allow for consumer specific preference parameters. However, we depart from the standard approach by treating these preferences as parameters to be estimated (rather than random draws from a mixing – or random coefficient – distribution). This means we can allow for arbitrary dependence between both preference estimates and consumer specific predictions from tax simulations and any set of attributes of individual consumers. In standard random coefficient models the interactions between consumer attributes and the preference distribution need to be specified ex ante and a specific functional form imposed. Our approach enables us to estimate how the entire distribution of effects from the tax vary across any consumer attributes and, crucially, whether the tax targets sugar consumption among those individuals most likely to be consuming excessively.

There is a growing theoretical literature that suggests that some individuals fail to accurately weight future costs when making consumption decisions (for a survey see Rabin (1998)). In the context of sugar consumption such failures can have serious consequences, with, for instance, excess intake being linked with increasing incidence of type II diabetes. There is evidence of people exhibiting behavioural bias with respect to food consumption, both experimental (for instance Read and Van Leeuwen (1998) and Gilbert et al. (2002)) and through the existence of a multi-billion pound dieting industry (Cutler et al. (2003)). Internalities are difficult to observe directly, however, it is clear that the costs of excess sugar consumption are higher for those for whom added sugar makes up a large share of their total diet (relative to those for whom it makes up a small share), and are higher for younger individuals, and as a result these groups have been the focus of public health initiatives. We focus on these two groups and ask how well targeted taxes on soda are at reducing their sugar purchases.

We find that both individual from households with a high share of calories from sugar and younger individuals have stronger preferences for sugar. Consumers with high sugar diets are less price sensitive – the own price elasticity of demand for sugary soda of those in the top decile of the added sugar distribution is less than half the magnitude of those in the bottom decile. The relationship between price sensitivity and age is less strong, though the young tend to be more price sensitive than older individuals. These patterns have important implications for the effect of soda taxes.

We consider the introduction of a tax on sugary soda products. We simulate pass-through of the tax to consumer prices using our demand estimates and first-order conditions from a Nash-Bertrand pricing game played by soda firms, finding that for the non-diet products consumer prices rise by more than the tax and the prices of some diet products falls. Firms' equilibrium pricing responses act to magnify the price differential between sugary and diet varieties (Bonnet and Réquillart (2013) find similar results for the French soda market). The tax is poorly targeted in the sense that the distribution of individual level reductions in sugar in the new equilibrium does not systematically vary with the total amount of sugar in individuals' diets. However, younger individuals do reduce sugar consumption by more than older individuals, especially those aged 21 or less.

Whether a consumer is made better off by the tax depends on whether the direct loss in consumer surplus they suffer due to higher prices is offset by any averted future costs (internalities) due to lower sugar consumption. We do not directly observe the size of these internalities; however, for each individual we can compute what would be the break-even level of internality per 100g sugar reduction that would be necessary to leave them indifferent to the introduction of the tax.

The break-even internality is nearly 2.5 times higher for individuals in the top decile of the added sugar in total diet distribution. Individuals at the top of the added sugar distribution experience the largest direct reductions in consumer surplus, which are not made up for by large reductions in sugar. The more convex are internalities in total dietary sugar, the more likely it is that individuals with high sugar diets will be made better off from the tax.

The median break-even internality is around 25-30% higher for those aged below 21 than for middle-aged individuals. This is because the young have relatively large consumer surplus losses from the tax. However, the young are likely to benefit significantly in terms of reduced future costs, so this policy might leave them better off. Older consumers (over 60) also have a high break-even internality, because they are not very responsive to the tax in terms of reducing their sugar intake. It is less likely that people over the age of 60 will benefit from reduced future costs to such an extent, so it seems likely that they will be a group who will loose out from the tax.

We assess the redistributive properties of the tax. While those consumers with low annual total equivalised household grocery expenditures tend to consume more sugary soda, and therefore suffer more consumer welfare loss due to a higher price, they also reduce their sugar consumption relatively strongly. The result is that the break-even internality is broadly constant across deciles of the total expenditure distribution. Therefore, to the extent poorer individuals suffer higher internalities for a given amount of sugar consumption, the progressive nature of internalities saving may outweigh the regressivity of the tax's traditional economic burden.

We conduct our analysis using data on food and drink a sample of UK individuals purchase on-the-go for immediate consumption. These data are both novel and have a number of very useful features. Most importantly, consumption of soda outside the home is very 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 very few studies of this important part of the market. In addition, the data are individual rather than household level (meaning we avoid the need to make strong assumptions about intra-household preferences; see Browning and Chiappori (1998)), they contain observations on teenagers (a group where internality concerns may be particularly strong) and each individual is observed making purchases many times (enabling us to estimate consumer specific preference parameters). As, by definition, observed purchases are not taken into the home, we also sidestep complications that arise when people use sale periods to stockpile (see, Hendel and Nevo (2006) and Wang (2015)).

To estimate consumer specific preferences we exploit the time dimension of our data (i.e. the relatively large number of observations we have for each person). However, our estimates may suffer from an incidental parameters problem that is common in non-linear panel data estimation. Even if both the number of individuals in the sample and the number of observations for which we observe them both tend to infinity, an asymptotic bias (which nevertheless shrinks with the sample size) may remain (Arellano and Hahn (2007)). To assess whether this bias is important in our setting we employ the split sample jackknife bias correction procedure suggested in Dhaene and Jochmans (2015), showing our conclusions are robust to this correction.

The rest of this paper is structured as follows. In Section 2 we discuss what factors determine the impact of a soda tax on consumer welfare. In Section 3 we discuss our empirical approach for shedding light on these determinants. Section 4 presents the novel individual level data we use along with model estimates. In Section 5 we present the results of our counterfactual analysis of a soda tax, describing the impact on equilibrium prices and the distribution of sugar purchases and discussing the likely impact of the tax on welfare and its redistributive properties. In Section 6 we extend our demand model to incorporate broader patterns of consumer switching, including towards food, and show our results based on drinks demand are robust to inclusion of these additional margins of consumer response. A final section concludes.

2 The consumer welfare effects of a tax on soda

Before turning to our model we describe the effects of a soda tax using a simple welfare criterion in order to highlight the key forces at play, which helps structure our empirical analysis. The rationale for such a tax is that excess sugar consumption is associated with internalities, in the form of unanticipated future health and well-being costs to individuals themselves (and possibly also externalities, in the form of public costs of funding healthcare systems). A soda tax will be most effective when it leads to larger reductions in sugar amongst those with large internalities, while leaving the behaviour of consumers with smaller internalities relatively unchanged. Additionally, if poorer consumers continue to consume the taxed good, the tax might also have undesirable redistributive properties.[2]

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

[2] Of course, what matters for equity considerations is the redistributive effect of the tax and benefit system as a whole. The introduction of a new corrective tax could in be matched by adjustments in other parts of the system (e.g. income tax and benefits) to offset any negative redistributive consequences (see Kaplow (2012)). However, in practice, such off-setting adjustments are often not made.

internalise these costs. However, whether tax policy can indeed improve welfare will depend on how successfully it averts internalities by lowering the 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. This may be zero, or it may capture the value the consumer obtains from spending on public goods.

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 given by

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

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

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

where λ_i denotes the marginal (decision) utility of income of consumer i , $\bar{\lambda}$ is the weighted average marginal utility of income, $\bar{\lambda} = \sum_i \beta_i \lambda_i$, $\phi_i' \equiv \phi_i' \left(\mathcal{S}_i(\mathbf{p}, y + r_i) \right)$ 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, which we aim to measure empirically.

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 (as long as the tax on the sugar in soda lowers demand for sugar in soda $S_i'^w < 0$). For these consumers the reduction in sugar from soda that results from an increased tax rate leads,

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

through this channel, to a welfare gain. The size of this gain is proportional to how responsive 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 $S_i'^{nw} > 0$, 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. Notice though that the redistributive term captures the traditional economic incidence of the tax. Even if the soda tax is regressive when assessed on the traditional basis of who pays it, in a broader sense, it may still be progressive if those with high marginal utilities of income benefit most in the long run from the averted internalities.

3 Empirical framework

To assess empirically the likelihood of taxes on soda being effective, we require estimates of the demand shape. This will tell us how strongly consumers will switch away from the sugar in soda ($S_i'^{w}$) and how strongly they will switch to alternative sources of sugar ($S_i'^{nw}$) in response to price changes. Estimates of demand shape and our supply side model of Bertrand-Nash price competition means we can evaluate pass-through of tax and therefore determine the new price equilibrium. We also require measures of marginal internalities (ϕ_i') and the marginal utility of income (λ_i) in order to evaluate the impact on consumer welfare. Crucially, we need to know the correlation between all these variables. In this section we develop a demand model and a supply side oligopoly model that allows us to estimate these correlations as well as pass-through of tax to consumer prices.

We estimate our model using novel longitudinal data on food and drink purchases consumers make whilst on-the-go (i.e. food and drinks bought from retailers for immediate consumption). This is both an important and understudied segment of the market and it also means we avoid the usual problem of observing household instead of individual demand realisations and purchase that differ from consumption due to stockpiling. For a sample of several thousand consumers we observe a long history of on-the-go purchases (81 on average). We describe these data more fully in Section 4.1.

3.1 Internalities

Excess sugar consumption is strongly related to obesity and diet related disease (WHO (2015)). The World Health Organization recommends no more than 5% of calories should come from added sugar, yet across the developed world people exceed this recommendation (see Azaïs-Braesco et al. (2017)). In Figure 1(a), using data from the National Diet and Nutrition Survey over 2008-2011 for a representative sample of 3,073 UK adults and children, we plot the distribution of share of calories from added sugar – 94% of individuals exceed the WHO recommendation.[4] The propensity for people to over-consume sugar and the profound effects excessive intake on health are clear.

Internalities are the portion of the harm that people incur in the future due to their current behaviour that they do not take account of at the point of consumption. While measuring total harm from excess sugar intake is possible (if a challenging exercise), measuring the precise size of internalities is very difficult. We do not attempt this in this paper. Rather we consider two important dimensions across which marginal internalities are increasing and we consider how the effects of a soda tax differ across these dimensions.

First, the costs of an additional unit of sugar is likely to be increasing in the level of sugar in one's diet – which corresponds to the internality function, ϕ_i, being convex. This means, all else equal, those with more sugar in their diet will tend to impose larger internalities on themselves. Evidence of this type of effect is provided in the medical literature – for instance Hall et al. (2011) show adults with greater adiposity experience larger health gains from a given reduction in energy intake. For each individual in our sample we observe the total grocery basket bought by the household that person belongs to. From this we can compute the share of household calories from added sugar.

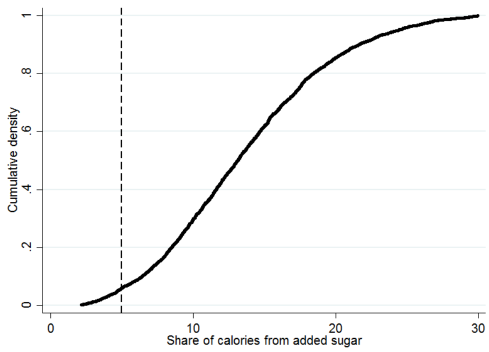
Second, for a given level of sugar consumption, internalities will vary across individuals. Some individuals are more likely to suffer from imperfect information about the future health consequence of current sugar consumption, or be more likely to underweight these future costs than others. One group for which internalities are likely to be particularly important are young people. There is evidence that excess sugar consumption is associated with poor mental health and poor school

[4] In the Online Appendix we show a pattern for the US.

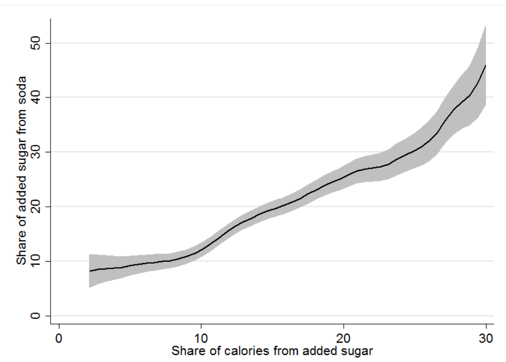
performance in children, and poor childhood nutrition is thought to be an important determinant of later life health, social and economic outcomes and of persistent inequality.[5]

Figure 3.1: *Added sugar and soda*

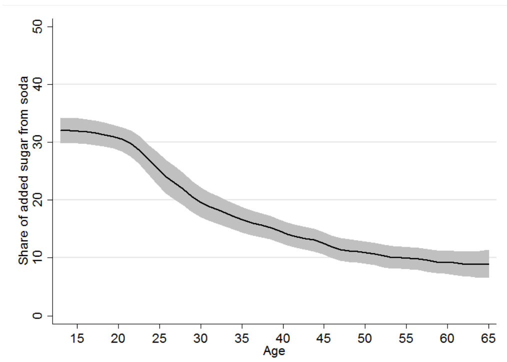
(a) Share of calories from added sugar



(b) Soda and added sugar



(c) Soda and age



Notes: Numbers computed using National Diet and Nutrition Survey 2008-2011. Vertical line in panel (a) denotes the WHO target of no more than 5% of calories from added sugar. Shaded areas in panels (b) and (c) denote 95% confidence intervals.

Panels (b) and (c) of Figure 3.1 show that the share of added sugar calories that people obtain from soda is increasing in the share of total calories from added sugar and is decreasing in age. Therefore, a soda tax does impact a disproportionately high fraction of the added sugar intake of two of the groups most likely to suffer from sugar related internalities. However, whether a soda tax will succeed in improving welfare of these groups will depend not only on the extent to which these groups consume soda, but crucially on how strongly they switch away from the sugar in soda and how strongly they switch to alternative sources of sugar. Therefore, to

[5]Cavadini et al. (2000) document an increase in soft drink consumption in the US for 11-18 year olds of almost 300% for boys, and over 200% for girls between 1965 and 1996. Nielsen and Popkin (2004) document a contemporaneous fall in the share of calories children get from milk. Medical evidence suggests that exposure to sweetened beverages early in life can establish strong lifelong preferences for these products (Mennella et al. (2016)).

assess whether a soda tax is likely to affect the groups most likely to suffer from internalities we need demand estimates that incorporate heterogeneity across these dimensions. We develop a framework that enables us to estimate individual specific preference parameters and therefore relate the effects of taxes flexibly to these internality proxies, as well as to measures of incomes.

3.2 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 each consumer on many choice occasions, indexed by $t = \{1, ..., T\}$. A choice occasion refers to a consumer visiting a store and purchasing a drink. We therefore take the decision to purchase a drink as exogenous. In Section 6 we explore the robustness of this assumption by incorporating switching to non-drink sources of sugar as well as to non-sugary snacks.

The "inside" products include sodas, $j \in \{1, ..., j'\} = \Omega_w$, and alternative drinks (fruit juice and flavoured milk), $j \in \{j' + 1, ..., J\} = \Omega_{nw}$. We will distinguish the set of sugary sodas, Ω_s, from non sugary sodas, Ω_{ns}; $\Omega_w = \Omega_s \cup \Omega_{ns}$. Each product belongs to a brand – we denote the brand product j belongs to as $b(j)$. There are fewer brands than products; soda brands typically comprise at least two different sizes and a sugary and diet variety.

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

$$U_{ijt} = \alpha_i p_{jrt} + \beta_i s_j + \gamma_i w_j + \delta_{d(i)} z_j + \xi_{d(i)b(j)t} + \zeta_{d(i)b(j)r} + \epsilon_{ijt}, \tag{3.1}$$

where ϵ_{ijt} is an idiosyncratic shock distributed type I extreme value. p_{jrt} denotes the price of product j – it varies over time (t) and cross-sectionally across retail outlets (indexed by r). s_j is a dummy variable for whether the product is a sugary variety (rather than a diet variety) and w_j is a dummy variable for whether the product is a soda.

Importantly, we allow the preference parameters on these product attributes (α_i, β_i and γ_i) to be consumer specific. We also include size-carton type effects (z_j), time-varying brand effects ($\xi_{d(i)b(j)t}$) and retailer-brand effects ($\zeta_{d(i)b(j)r}$). In each case we allow the influence of these attributes on the pay-off to vary by gender and age (whether the consumer is younger than 40) – we denote the gender-age groups by $(1, ..., D)$ and let $d(i)$ denote the group consumer i belongs to.

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

$$U_{i0t} = \xi_{d(i)0t} + \zeta_{d(i)0r} + \epsilon_{i0t}, \tag{3.2}$$

where $\xi_{d(i)0t}$ and $\zeta_{d(i)0r}$ are gender-age-time and gender-age-retail outlet specific deviations in the mean outside option pay-off.

$\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_N)'$ 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 nonparameteric 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.

A number of papers (see, for instance, Berry et al. (1995), Nevo (2001) and Berry et al. (2004)) show that incorporating consumer level preference heterogeneity, through parametric random coefficients, is important for enabling choice models to capture switching patterns across products,[6] while a few papers have used non-parametric methods to relax parametric restrictions on random coefficients.[7] Like these papers we model consumer specific preferences, however, in contrast to them, we treat the preferences as parameters to be estimated rather than draws from a random coefficient distribution. This means we can flexibly relate the preference parameters to any observable attributes of consumers.

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. We implement the split sample jackknife procedure suggested in Dhaene and Jochmans (2015) and in Section 4.3 show that our maximum likelihood and jackknife estimates are similar and that the bias correction does affect our main results.

A convenient feature of considering soda purchased on-the-go for immediate consumption is that it minimises concerns about dynamics in demand arising due to consumer stockpiling (a case considered in Wang (2015)); by definition the consumption occasions that we are modelling do not involve storage.

[6]Lewbel and Pendakur (2017) show similar results apply in non-linear continuous choice models, with the incorporation of random coefficients resulting in their model much more effectively capturing the distributional impacts of taxation.

[7]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.

Contrary to standard logit discrete choice models, we allow consumers to dislike some products and so have a zero probability of purchasing them. We use the long time dimension of our data to identify those consumers that have infinite negative preferences for some characteristics and allow them to never purchase products with these characteristics. Assuming that the unobservable error term has "large" support (we assume infinite support with an extreme value distribution), a consumer that never chooses one of the soda options, but does choose one of the other products such as 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 that imply $P_{it}(j) = 0$ 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 $\gamma_i = \infty$ (or equivalently to have negative infinite preference for non-soda) and those that sometimes purchase soda have finite soda preferences $\gamma_i \in (-\infty, \infty)$. 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 ($\beta_i \in (-\infty, \infty)$) (or negative infinite taste for diet drinks).

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.

Since the consumer will never buy some products with characteristics he dislikes, the logit choice probability will be zero for these products and for the other products, it will depend only on the set of products for which the consumer does not have an infinite distaste. It is convenient to define the consumer i specific set of products with non-zero purchase probabilities, denoted by Ω_i, as

$$\Omega_i = \begin{cases} \Omega_s \cup \Omega_{ns} \cup \Omega_{nw} \text{ if } \beta_i \in (-\infty, \infty) \text{ and } \gamma_i \in (-\infty, \infty) \\ \Omega_{ns} \cup \Omega_{nw} \text{ if } \beta_i = -\infty \text{ and } \gamma_i \in (-\infty, \infty) \\ \Omega_s \cup \Omega_{nw} \text{ if } \beta_i = +\infty \text{ and } \gamma_i \in (-\infty, \infty) \\ \Omega_s \cup \Omega_{ns} \text{ if } \beta_i \in (-\infty, \infty) \text{ and } \gamma_i = \infty \\ \Omega_{ns} \text{ if } \beta_i = -\infty \text{ and } \gamma_i = \infty \\ \Omega_s \text{ if } \beta_i = +\infty \text{ and } \gamma_i = \infty. \end{cases}$$

It is also convenient to rewrite the pay-off from choosing option $j > 0$ as:

$$U_{ijt} = \alpha_i p_{jrt} + \beta_i s_j + \gamma_i w_j + \eta_{ijrt} + \epsilon_{ijt}$$

where $\eta_{ijrt} = \delta_{d(i)} z_j + \xi_{d(i)b(j)t} + \zeta_{d(i)b(j)r}$ and the utility from the outside option as

$$U_{i0t} = \eta_{i0rt} + \epsilon_{i0t},$$

where $\eta_{i0rt} = \xi_{d(i)0t} + \zeta_{d(i)0r}$.

The choice probability of purchasing any good j by consumer i can then be written as:

$$P_{it}(j) = \frac{1_{\{\gamma_i \in (-\infty,\infty), j=0\}} \exp \eta_{i0rt} + 1_{\{j \in \Omega\}} \exp(\alpha_i p_{jrt} + \beta_i s_j 1_{\{\beta_i \in (-\infty,\infty)\}} + \eta_{ijrt})}{1_{\{\gamma_i \in (-\infty,\infty)\}} \exp \eta_{i0rt} + \sum_{k \in \Omega_i} \exp(\alpha_i p_{krt} + \beta_i s_k 1_{\{\beta_i \in (-\infty,\infty)\}} + \eta_{ijrt})}.$$

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:

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

and, denoting the gender-age specific preference parameters, $\boldsymbol{\eta}$, the associated log-likelihood function is:

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}) = \sum_i \ln \mathcal{P}_i(y_i). \tag{3.3}$$

3.3 Identification

Our main identification challenge is pinning down the causal impact of price on demand. Our strategy for doing this relies on two sources of price variation. Firstly, we exploit cross-retailer price variation. We observe individuals making purchases in different retailers (and thereby facing different price vectors). We assume the retailer choice is not driven by shocks to demand for specific drinks products. Secondly, we exploit variation within brand of prices across different containers and sizes. While there may be some aggregate shock to demand for a specific brand (that manufacturers observe and changes prices as a consequence of), we assume that there are not aggregate shocks within brand for different container types. We discuss each source of variation in turn.[8]

The price vector an individual faces in our model depends on which retailer they visited. These retailers include a large retailer that prices nationally, smaller retailers with regionally varying prices and vending machines. We include common (by gender-age group) time varying brand effects $\xi_{d(i)b(j)t}$ and retailer effects, interacted with soda, the non-soda drinks and the outside option, $\xi_{d(i)b(j)r}$. The former capture aggregate (demographic specific) fluctuations in brand demand over time and the latter capture any differential propensity of consumers to choose different

[8]In the Online Appendix we describe some of the variation in product prices.

drink types across retailers. Conditional on these, the cross-retailer difference in prices provides a useful source of price variation.

There are two main concerns with exploiting this type of price variation. First, an issue would arise if individual level demand shocks to specific soda products drive store choice for the on-the-go market; for instance, 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 only a very small fraction of total grocery spending.

Second, an issue would arise if differential changes in the prices of different sodas across retailers are driven by retailer-time varying demand shocks for soda products. In the UK the vast majority 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.

The second source of price variation we exploit is non-linear pricing across container sizes that is common in the UK (prices are linear for a fixed container size but non-linear across different container sizes of the same brand). This price variation is not collinear with the size fixed effects and the extent of non-linear 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. It seems 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 brands.

3.4 Pass-through of a tax on sugary soda

We consider the impact that a tax levied on sugary soda would have on sugar purchases. We focus on a volumetric tax applied only to sugary soda.[9] A number of US cities have recently legislated for the introduction of such a tax[10], the UK is set to introduce a tax on sugary soda in 2018 and France and Mexico have had soda taxes in place since 2012 and 2014. We simulate the introduction of a tax of 25 pence per litre.[11]

The degree of pass-through of the tax to consumer prices will depend on the nature of competition in the market. We model tax pass-through by assuming that drinks manufacturers compete by simultaneously setting prices in a Nash-Bertrand game. We consider a mature market with a stable set of products, and we therefore abstract from entry and exit of firms and products from the market. We use our demand estimates and an equilibrium pricing condition to infer firms' marginal costs (see Berry (1994) or Nevo (2001)) in order to then simulate the effect of a tax on consumer prices.

Let $f = \{1, ..., F\}$ index firms and F_f denote the set of products owned by firm f. Normalising the size of the market to one and aggregating across consumer level purchase probabilities we obtain the demand function in market t, $q_{jt}(\mathbf{p}_t) = \sum_i P_{it}(j)$ for each product j. Denoting the pre-tax price by $\tilde{p}_{jt}$, firm f's (variable) profits in market t are given by:

$$\Pi_{ft} = \sum_{j \in F_f} (\tilde{p}_{jt} - c_{jt}) q_{jt}(\mathbf{p}_t) \tag{3.4}$$

where the demand depends on post-tax prices. As the relationship between the prices chosen by manufacturers and prices charged to consumers depends on the sugary soda tax in the simple following way:

$$p_{jt} = \begin{cases} \tilde{p}_{jt} + \tau z_j & \forall j \in \Omega_s \\ \tilde{p}_{jt} & \forall j \in \Omega_{ns} \bigcup \Omega_{nw}, \end{cases}$$

[9] In the Online Appendix we report results for a tax levied on all soda, computing the pass-through and demand changes.

[10] A tax of 1.5 cent per ounce on regular and diet soda 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.

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

it implies that $\frac{\partial p_{jt}}{\partial \tilde{p}_{jt}} = 1$ and thus the first order conditions satisfied by the Nash-Bertrand equilibrium are

$$q_{jt}(\mathbf{p}_t) + \sum_{k \in F_f} (\tilde{p}_{kt} - c_{kt}) \frac{\partial q_{kt}(\mathbf{p}_t)}{\partial p_{jt}} = 0 \quad \forall j \in F_f. \tag{3.5}$$

for all firms.

We observe prices in the absence of a soda tax. In this case consumer and producer prices coincide; $p_{jt} = \tilde{p}_{jt}$ for all j. Under the assumption that these prices are an equilibrium outcome of the Nash-Bertrand game played by firms, given our estimates of the demand function, we can invert firms' first order conditions to infer marginal costs c_{jt}.

In the counterfactual we solve for the new price equilibrium $\tilde{p}_{jt}$ satisfied by first order conditions 3.5 when τ is non-zero. This new price equilibrium depends on the tax rate τ and the change from the current prices to the counterfactual prices (relative to amount of tax levied) is the pass-through of the tax [12].

4 Soft drink demand in the on-the-go market

4.1 Data

A substantial portion of soda is consumed on-the-go; in the US around half of sugar-sweetened beverages are consumed outside the home (Han and Powell (2013)). These purchases are for immediate consumption, in contrast to purchases made in the grocery store, which are for planned future consumption. It is more likely that consumers will be more influenced by self control problems when purchasing for future consumption. Despite the importance of this market segment there are few studies modelling purchase behaviour on-the-go in a discrete choice framework, largely due to the lack of high quality 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), providing the opportunity to study in detail decision-making in this part of the market. Participants record all purchases of snacks and non-alcoholic drinks at the barcode

[12] In principle we could solve for a separate price equilibrium in each time period and retailer market (246 markets). Instead we average all markets at the year level and solve the supply side model for this sort of representative market, and simulate the the counterfactual price changes for this aggregate market. We then simulate the counterfactual demands on all markets using the the tax pass-through obtained for each product on this average representative market, abstracting from seasonal and regional effects for the sake of time.

(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.

We have information on 5,373 individuals over the period June 2009-October 2012. We observe each person making purchases on a minimum of 25 days and on 81 days on average. We observe 2,563 individuals purchasing soda on at least three occasions – together theses individuals account for 99% of all soda purchases. We focus on this sample (i.e. the soda purchasers) when estimating soda demand. We drop a small number of observations that due to measurement error in the data. Our results are robust to keeping this group in the sample, however it dampens the overall effect of the tax on aggregate sugar purchases. See Appendix A.1 for further details on the data.

We model consumer choice among soda products, as well as fruit juice, flavoured milk and mineral water (the outside option). Table 4.1 shows the main products in the market, along with the firm that produces them, the brand to which they belong, the size and container type and their market share. 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%. The fruit juice and flavoured milk products are composite products; their inclusion allows us to capture the possibility that consumers might respond to a soda tax by switching to alternative non-soda (but high sugar) drinks. These products are not subject to the counterfactual tax (which applies only to regular sodas); we assume their price remains fixed.

Table 4.1: *Drinks products*

	Product						
	Firm	Brand	Variety	Size	Market share	Price (£)	g sugar per 100ml
Sodas							
	Coca Cola Company				**54.7%**		
		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
	Pepsico				**15.1%**		
		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
	GSK				**11.7%**		
		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-sodas		Fruit juice		330ml	4.0%	1.39	10.6
		Flavoured milk		500ml	2.2%	0.96	10.6
Outside option		Water		12.3%			

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

4.2 Demand Estimates

In this section we summarise the main features of the demand estimates that drive the impact of a soda tax on sugar purchases; we report the parameter estimates in Appendix A.3.

The estimates of the consumer specific preference parameters (on price, sugar and soda) reveal there is a large degree of heterogeneity in preferences across individuals – the standard deviation for price preferences is 3.0 (with a coefficient of variation of 1), while the standard deviation for sugar and soda is 1.8 and 2.4. The marginal distributions of preferences over each product attributes depart significantly from normality; a major reason for this is that we allow for infinite portions of the parameter space, capturing, for instance, people that exclusively purchase either regular or diet sodas. The preferences also exhibit a considerable degree of correlation – price sensitive consumers tend, to some degree, to also have relatively strong sugar preferences (the correlation coefficient between price and sugar preferences is -0.2), but as well relatively strong preferences for the soda product attribute (the correlation coefficient between price and soda preferences is -0.8).

It is not possible to infer how sensitive consumers' soda demands are to price changes directly from the preference parameter estimates. To do this we report price elasticities in Table 4.2. The top panel of the table reports elasticities for four of the most popular products in the market – the regular and diet versions of a 500ml bottle of Coca Cola and Pepsi. In column 1 we report the percent change in demand for the product when its price increases by 1%. Columns 2-4 report how demand for alternative products (sugary sodas, diet sodas and alternative sugary drinks) would change and a final column reports what would be the overall change in demand for juice drinks (soda, fruit juice plus flavoured milk). For example, a 1% increase in the price of (regular) Coca Cola would result in a reduction in demand for that product of around 1.7%. Demand for alternative sugary sodas would rise by around 0.4%, demand for diet sodas would rise by 0.1% and demand for non-soda sugary drinks would rise by 0.2%. Demand for juice drinks as a whole would fall by 0.07%. 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.

In the bottom panel of the table we report the effect on demand of a marginal increase in the price of all sugary soda and in the price of all soda (i.e. both sugary and diet). The own price elasticity for soda is -0.3. This is much smaller than the own price elasticity of any individual soda product. The own price elasticity for sugary soda is larger (in absolute terms) at -0.7. This reflects the fact that some consumers respond to an increase in the price of sugary soda by switching to diet alternatives.

Table 4.2: *Price effects*

	Effect of 1% price increase on:				
	Own	cross demand for:			Total
	demand	sugary soda	diet soda	sugary alternatives	demand
Coca Cola 500	-1.747	0.372	0.117	0.184	-0.071
Coca Cola Diet 500	-1.467	0.111	0.367	0.057	-0.052
Pepsi 500	-2.135	0.202	0.067	0.090	-0.043
Pepsi Diet 500	-1.891	0.064	0.232	0.027	-0.037
Soda	-0.336			0.768	-0.267
Sugary soda	-0.723		0.496	0.768	-0.153

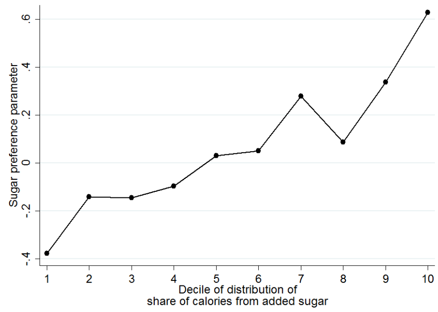
Notes: For each of the four products listed we compute the change in demand for that product, for alternative sugary and diet options and for total demand resulting from a 1% price increase. We also compute demand response for a 1% increase in the price of all soda products and all sugary soda products. Numbers are means across time.

A key feature of our model is it allows us to flexibly relate preference parameters (and elasticities), as well as estimates of the impact of a soda tax, to characteristics of consumers, including, most importantly, proxies for the internalities associated with their sugar consumption. For each individual in our sample we compute the share of their total households calories from added sugar, based on their grocery purchases. In Figure 4.1 we show how preferences vary across deciles of the distribution of total household calories from added sugar. Panel (a) shows that the mean estimated preference for sugar increases as we move from low to higher deciles of the added sugar distribution. This relationship is intuitive. Note though that 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. Panel (b) shows that consumers with a relatively high share of added sugar in their total grocery baskets tend to be relatively insensitive to price changes – they respond to a marginal increase in the price of all sugary sodas, in percentage terms, less than those with relatively low sugar diets. Panel (c) shows how consumers' propensity to switch from sugary sodas to sugary alternatives varies with the share of their overall household calories from added sugar. The percent increase in demand for sugary alternatives, in response to a 1% increase in the price of sugary sodas, is broadly constant across the added sugar distribution.

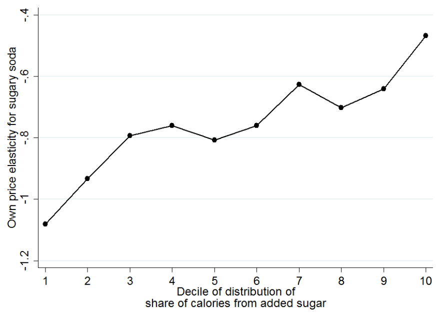
Figure 4.2 shows how sugar preferences and sensitivity to price changes varies across the age distribution. Young individuals tend to have relatively strong preferences for sugar. However, the own price elasticity for sugary soda is broadly constant across the age distribution. This does not necessarily mean the young will not respond more strongly in level terms to a soda tax – as highlighted in Figure 3.1, young consumers tend to have a much higher level of soda demand.

Figure 4.1: *Demand estimates by added sugar*

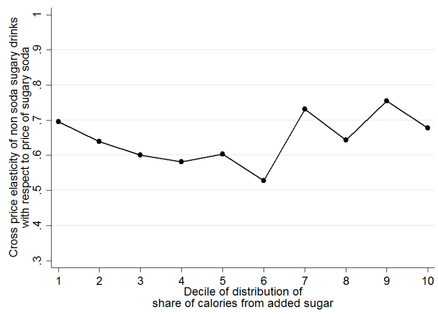
(a) Sugar preference parameter



(b) Own price effect for sugary soda



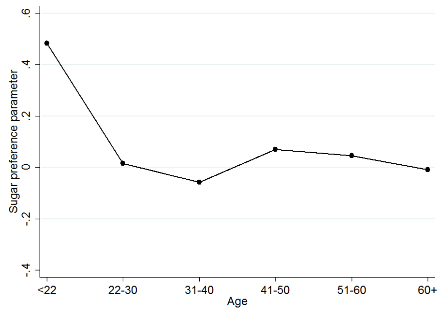
(c) Cross price effect to other sugar



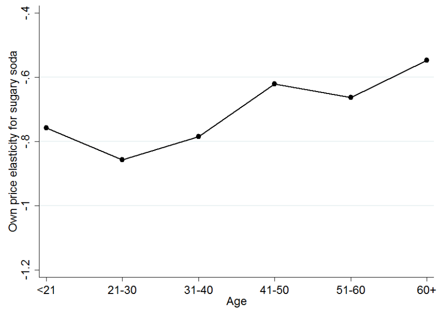
Notes: Figure shows how the mean sugar preference, own price elasticity of sugary sodas and cross price elasticity towards sugary alternatives vary across deciles of the distribution of total calories from added sugar in annual household grocery baskets.

Figure 4.2: *Demand estimates by age*

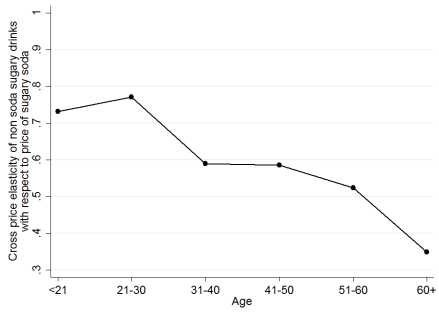
(a) Sugar preference parameter



(b) Own price effect for sugary soda



(c) Cross price effect to other sugar



Notes: Figure shows how the mean sugar preference, own price elasticity of sugary sodas and cross price elasticity towards sugary alternatives vary across age groups.

4.3 Bias correction for incidental parameters problem

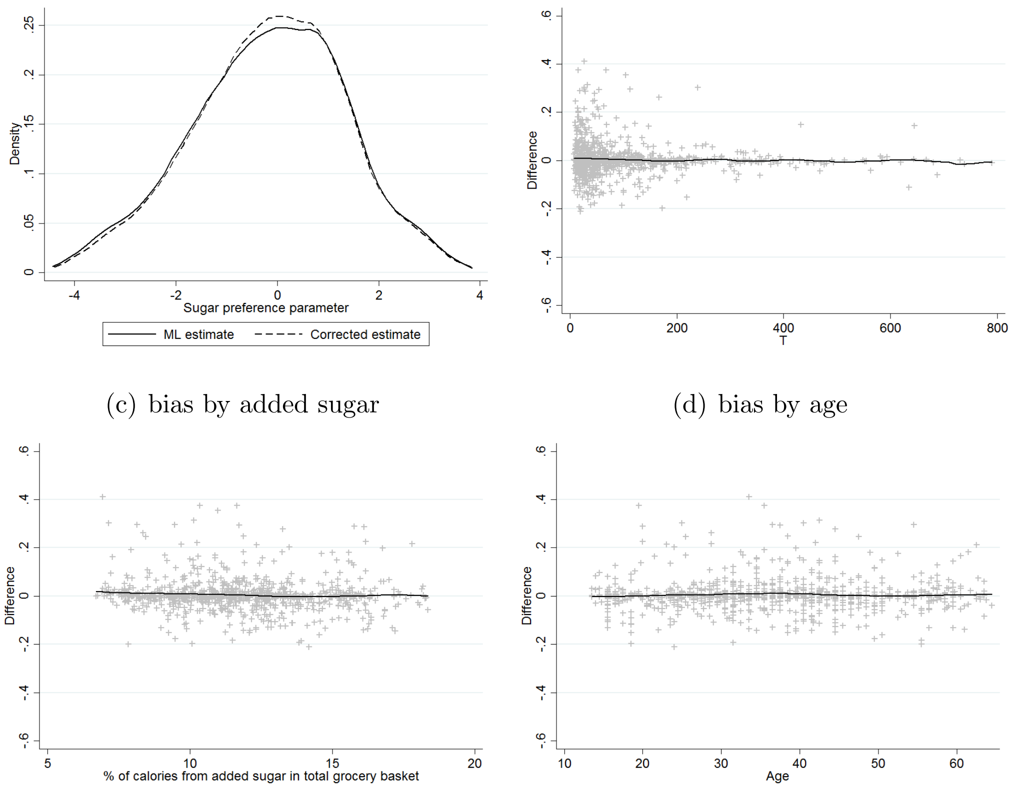
Our maximum likelihood estimate of the parameters may suffer from an incidental parameters problem. Even if both $N \to \infty$ and $T \to \infty$, if N and T grow at the same rate ($\frac{N}{T} \to \rho$ where ρ is a non zero constant), our fixed effect estimator will be asymptotically biased (Arellano 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 Online Appendix for more details.) A drawback of the split sample jackknife procedure is that, for some consumers, some parameters may not be 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 one subsample and all their inside option purchases are in the other. Therefore, in this section, when comparing the maximum likelihood and jackknife estimates we use only those consumers for which such an issue does not arise. In total they account for over 75% of choice occasions in our data.

In Figure 4.3 we graph the difference between the jackknife (bias corrected) and maximum likelihood sugar preference parameters. Panel (b) shows how this difference relates to the time a consumer is in the sample and panel (c) show the relationship with total calories from added sugar in consumers' diet. The figure shows that the difference between the two estimates is relatively small; the standard deviation of the sugar preference parameter estimates is 1.8, while the average absolute difference between the jackknife and maximum likelihood estimates is 0.08. Panel (b) shows that the difference is decreasing in T; those in the sample for a relatively short number of choice occasions on average have higher difference than those in the sample relatively many times. However, it also shows that, conditional on T, the average difference between the jackknife and maximum likelihood estimates is zero – a positive difference is equally likely as a negative difference. Indeed the distribution of the maximum likelihood and jackknife estimates of the preference parameters are almost indistinguishable (see panel (a)). Panels (c) and (d) of Figure 4.3 shows the difference between the jackknife and maximum likelihood estimates is completely unrelated to either the share of their total household calories individuals get from added sugar or their age.

In the Online Appendix we show similar conclusions to those for sugar hold for estimated price and soda preferences; the maximum likelihood and jackknife procedures yield almost identical preference distributions, any individual level differences are relatively small and are equally likely to be positive as negative and there is no relationship whatsoever with the key variables we relate our demand effects to. As a consequence our results regarding the effectiveness of soda taxes are completely robust to the bias correction procedure. Given this and given the maximum likelihood estimates do not involve a reduction in our sample we proceed in the rest of the paper with these.

Figure 4.3: *Sugar preference parameters*



(a) kernel density (b) bias by T

(c) bias by added sugar (d) bias by age

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

5 The effects of a soda tax

5.1 Prices

We use our demand estimates along with the supply side model outlined in Section 3.4 to simulate the introduction of a tax levied on sugary soda. We consider the introduction of a tax of 25 pence per litre. In solving for the post tax equilibrium we hold fixed the prices of the non-soda composite products, fruit juice and flavoured milk, as well as the outside option. We model the pricing response of soda manufacturers, including changes in prices for products not directly subject to the tax (i.e. diet sodas).

In Table 5.1 we report the mean tax per product levied, price change and change in share of the drink market due to the tax. We report these for the set of sugary soda, diet soda and sugary alternatives products and for the outside option. The average tax liable on sugary soda is 11 pence – for products with 500ml the tax liable is 12.5 pence, while for those with 330ml it is 8.25 pence. On average, the price of sugary sodas rises by 13 pence – average pass-through of the tax is therefore 130%. Pass-through rates vary across product, with the larger 500ml bottle products typically having rates of around 150% and smaller 330ml canned products having rates of around 100%.

Soda manufacturers also optimally respond to the tax by, on average, lowering the price of diet products. The average reduction in price is 2 pence, however, the 500ml bottle products see larger price reductions of around 5 pence, with little change in the equilibrium price of the smaller 330ml canned products.

The pricing response of soda manufacturers therefore acts to magnify the price differential that the tax creates between sugary and diet products. Relative to the case in which producer prices do not respond to the tax (so pass-through of tax is 100%), firms' equilibrium pricing response to the tax induces more switching away from sugary soda and more towards diet soda. Pass-through in excess of 100% suggests there are enough consumers that have sufficiently strong tastes for sugary sodas to make it profitable for manufacturers to respond to the tax by increasing margins. The flexible preference distributions we allow for price, sugar and soda product attributes is crucial in enabling us to capture realistic supply side responses.

Our finding of pass-through in excess of 100% for a sugary soda tax and reductions in the prices of diet products conforms with evidence found in other studies. For instance, using historic changes in US sales taxes, Besley and Rosen (1999) find the US soda industry over-shifts tax, while in an ex ante study of the effects of a sugary soft drinks tax in France, Bonnet and Réquillart (2013) find pass-through exceeding 100% and reductions in the prices of diet products.

The final row of Table 5.1 shows the change in market shares for the four sets of products. The market share of sugary sodas falls by 5.5 percentage points. Most of this demand switches to diet sodas (which see an increase in share of 3.4 percentage points), though alternative sugary drinks and the outside option also see increases in market share of 0.6 and 1.5 percentage points.

Table 5.1: *Impact of tax on market equilibrium*

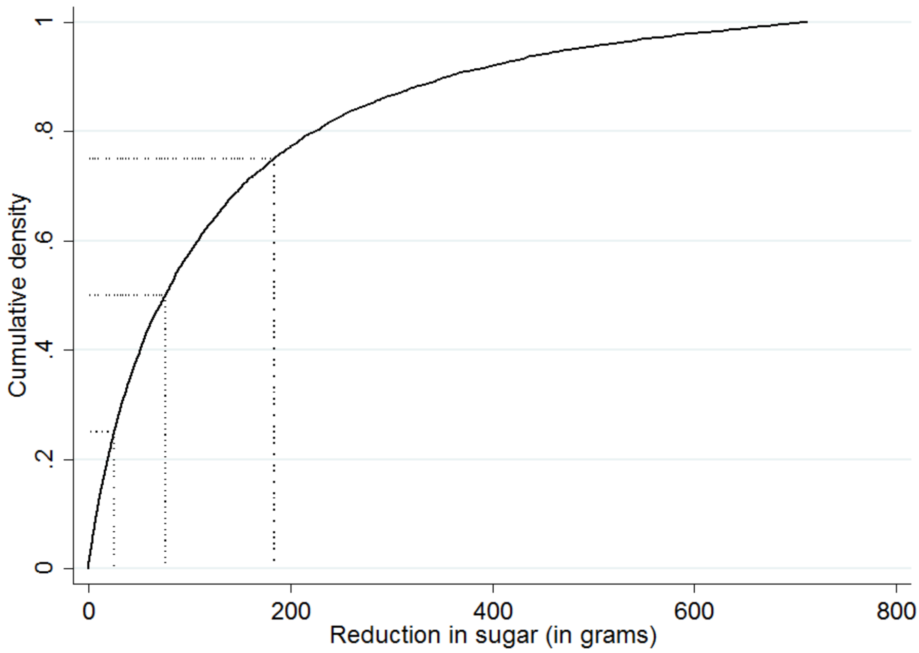
	Sugary soda	Diet soda	Sugary alternatives	Outside option
Tax (£)	0.10	0.00	0.00	0.00
Δ price (£)	0.13	-0.02	0.00	0.00
Δ share (p.p)	-5.50	3.41	0.59	1.51

Notes: Numbers are mean across products. Tax and price change are weighted by market share. Zero tax and Δ price for Sugary alternaties and outside option are by assumption.

5.2 Reduction in sugar as a result of the tax

Figure 5.1 shows the cumulative distribution of the reduction in total sugar resulting from the tax, measured in grams per year. The dotted lines show the 25th percentile (where the reduction is 10g), the median (59g) and the 75th percentile (183g). The reduction in sugar from soda is slightly higher (29g at the 25th percentile, 86g at the median and 207g at the 75th percentile); some consumers respond to the tax by switching to alternative (non-soda) sugary drinks. Reductions in sugar due to the tax are positively correlated with pre-tax sugar (with a correlation coefficient of 0.54).

Figure 5.1: *Distribution of reduction in sugar as a result of the tax*

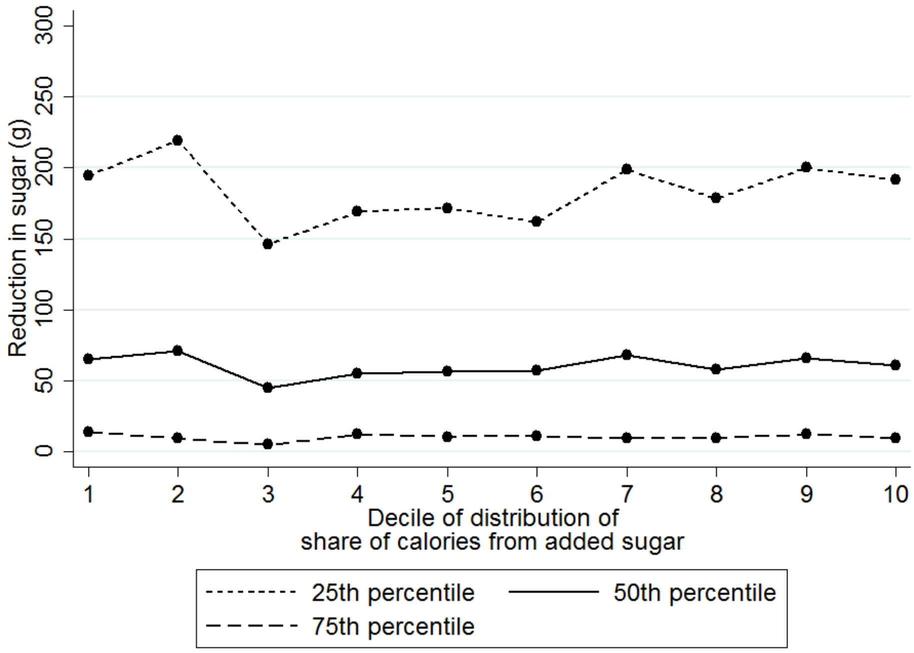


Notes: .

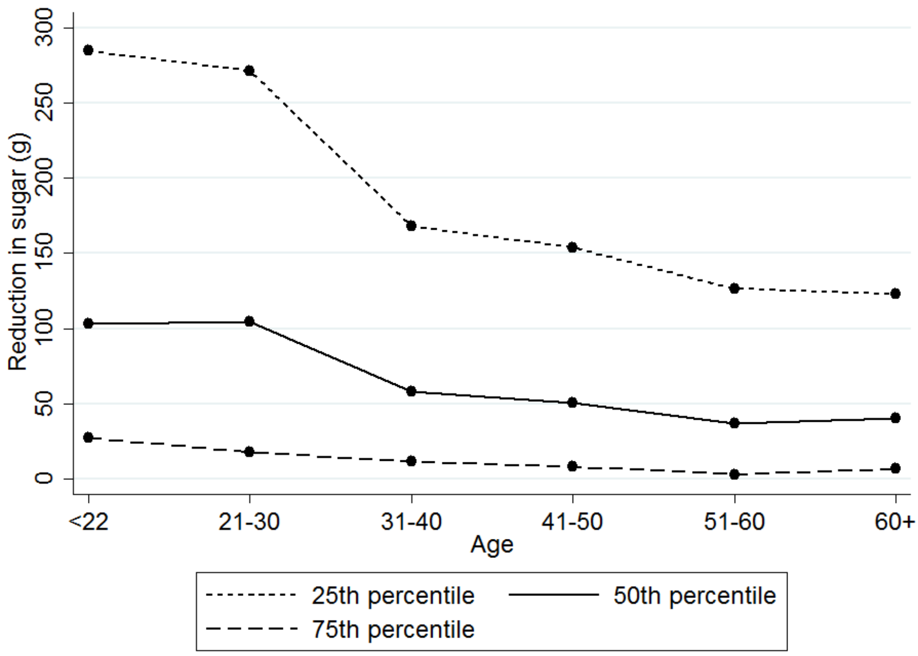
Section 2 makes clear that a well-targeted soda tax will lead to a greater reduction in sugar purchases among individuals who are most likely to suffer from sugar related internalities (or to create sugar induced externalities). Here we use our estimates of the impact of the tax, which we have at the individual level, and compare them to individual characteristics. In Section 3.1 we argued that internalities are increasing in the total amount of sugar in an individual's diet and declining in age.

Figure 5.2: *Reduction in sugar purchases as a result of the tax*

(a) Change across deciles



(b) Change across ages



Notes: The figures show how the 25th, 50th and 75th percentiles of the distribution of sugar reduction from the tax varies, in panel (a) across deciles of the distribution of total calories from added sugar in annual household grocery baskets and in panel (b) by age.

In Figure 5.2 panel (a) we show how the 25th, 50th and 75th percentile of the distribution of reduction in sugar that results from the tax varies across deciles of the distribution of total added sugar in the individual's grocery basket. The figure shows that the tax is poorly targeted, it does not achieve larger reductions for those in higher deciles. If internalities were driven purely by the share of added sugar in people's diets, this would suggest the tax fails to achieve the largest behavioural changes among those most afflicted by internalities.

In panel (b) we show the 25th, 50th and 75th percentile of the distribution of reduction in sugar that results from the tax varies across by age. The distribution of sugar reductions is higher for those aged below 31, than those aged over 31, and is highest for those below 22 years old.

5.3 Consumer welfare

The soda tax reduces annual sugar consumption of the mean consumer by 160g. We measure the loss in consumer surplus from higher prices using compensating variation – the payment a consumer would require to be indifferent to the introduction of the tax based on their decision utilities (i.e. excluding any gains from averted internalities). The mean compensating variation is £3.81 per consumer. The tax raises revenue of £2.85 per consumer. If all consumers are identical, whether the tax improves consumer welfare (net of tax revenue) would boil down to whether the 160g reduction in sugar averts more internalities (or externalities) than £0.96 (the difference between the per consumer compensating variation and revenue). However, as we demonstrate, consumers are heterogeneous in their response to the soda tax (and therefore in their sugar reduction and compensating variation) and in how this relates to how much internality their sugar consumption is liable to generate.

In Figure 3(a) we show how compensating variation varies across deciles of the distribution of total household calories from added sugar. We show the 25th, 50th and 75th percentiles of the compensating variation distribution. In each case compensating variation is higher for the top half of the added sugar distribution, and highest for the top two deciles – in the case of the median, compensating variation for the top decile is almost three times as large as for the bottom decile. The reason for this is individuals with a high share of sugar in their overall diet tend to consume a relatively high amount of sugary soda (see Figure 3.1). Therefore, the economic burden of the tax falls disproportionately on this group.

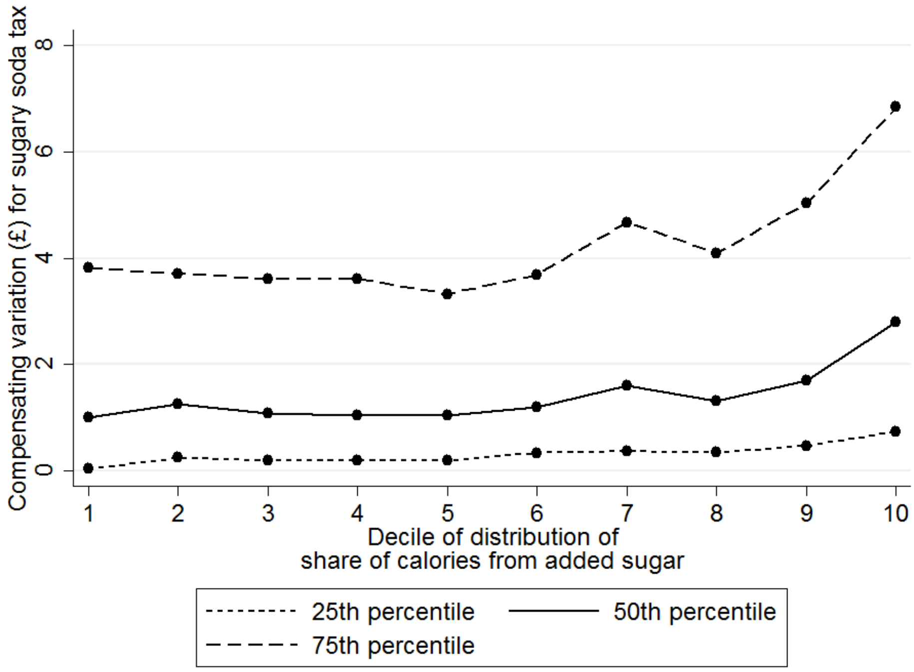
However, this does not necessarily mean the group of consumers with high sugar diets will loose out more as a consequence of a soda tax. The most compelling motivation for taxing sugar is that excess consumption generates costs not taken account of by individuals at the point of consumption (many of which are likely to be internalities). If the tax leads to averted internalities that exceed consumers' compensating variation, then it may actually improve their long run welfare.

To measure how large internality savings would have to be for the tax to improve consumer welfare, for each individual in our sample, we compute the average value of internality per 100g reduction in sugar that would have to be averted as a result of the tax for the consumer to be indifferent to the introduction of the tax. In this calculation we assume that tax revenue is not redistributed back to soda consumers. This differs from compensating variation, which only takes account of the loss in direct decision utility the consumer incurs from the tax.

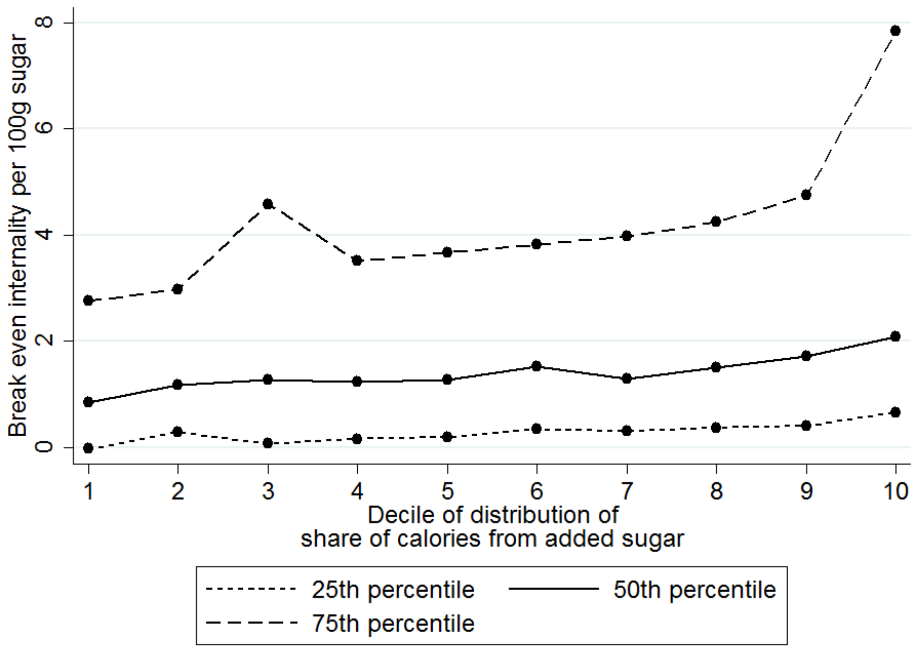
Figure 3(b) shows how this break even internality per 100g of sugar varies across the deciles of the total added sugar distribution. The broad pattern is that as we move from low to high deciles the break even internality increases – the median value for the top decile is £2.08 per 100g, while for the bottom decile it is £0.84. This pattern can be explained by the reduction in sugar due to the tax being broadly constant across the added sugar deciles, but compensating variation rising across them. Those individuals with a high share of their calories from added sugar tend to consume a lot of soda (and therefore are particularly adversely affected by the higher prices due to the tax) and they do not systematically lower their sugar intake by more due to the tax (they in fact have sugary soda demand that is less price elastic). Therefore the value of internality averted for a given reduction in sugar must be higher for this group than for those with less sugar in their diets in order for their welfare to be improved by the tax. If the internalities from sugar consumption are sufficiently convex (so for instance an extra 100g of sugar consumption imposes around 2.5 times the internality costs on those in the top decile of the distribution than those in the bottom), then this condition will be met.

Figure 5.3: *Consumer welfare effects by added sugar*

(a) Compensating variation



(b) Break even internality/100g sugar

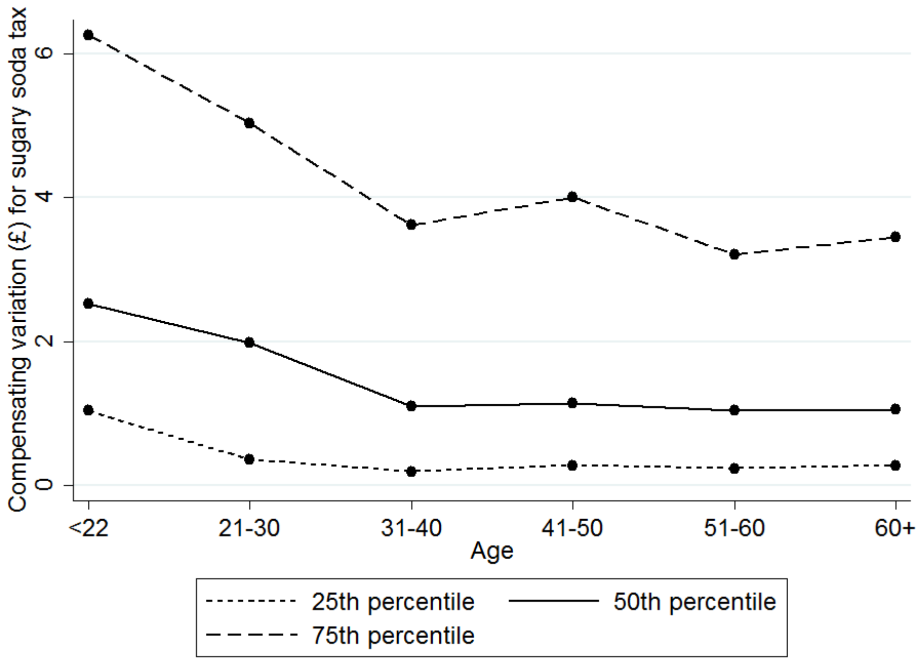


Notes: Panel (a) shows annual compensating variation and panel (b) shows the value of internalities required to be averted per 100g reduction in sugar to make the consumer indifferent in the long run to the tax. In each case we show how the 25th, 50th and 75th percentiles of the distribution varies across deciles of the distribution of total calories from added sugar in annual household grocery baskets.

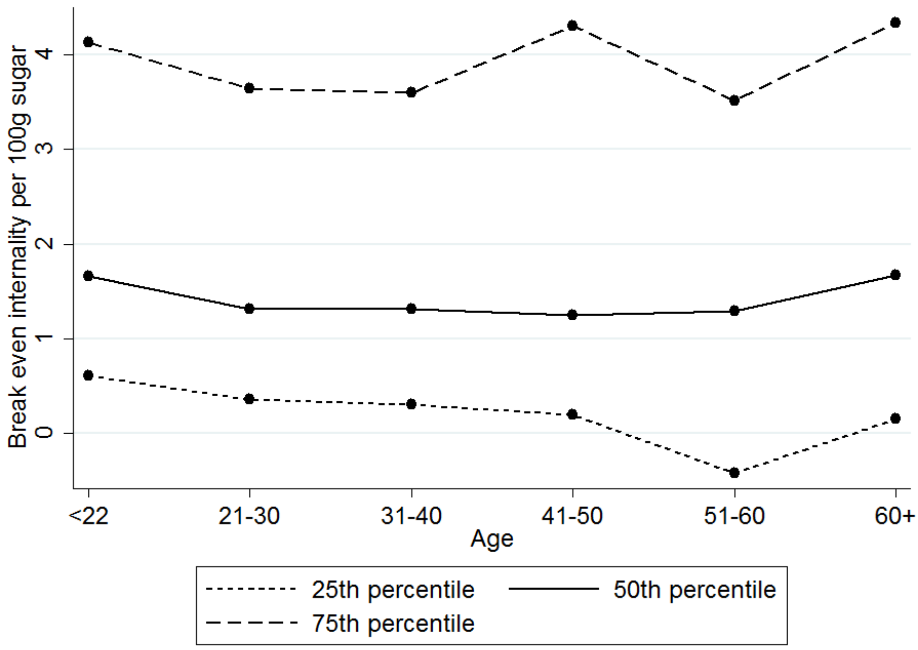
In Figure 5.4 we repeat the analysis but this time look at how welfare effects vary across the age distribution. Panel (a) shows how compensating variation varies across age groups and panel (b) shows how the break even internality per 100g of sugar varies. Compensating variation declines across the first three age groups before levelling off; median compensating variation for those aged below 22 is 1.3 times higher than for those aged 22-30 and about 2.3 times higher than those in older age groups. The break even level of internality exhibits less variation across age groups than across the added sugar distribution. The median value is highest for those aged below 22 and those over 60 and between 25-30% higher than for those in intermediate age groups. The relatively high value for the youngest age group is because, even though the tax leads to a relatively large reduction in sugar for this group, it also, to a greater extent, leads to relatively large compensating variation. However, given the considerable evidence that the young are most likely to suffer from large internalities from excess sugar consumption, it is also likely they as a group will benefit most from the tax in terms of internalities averted by a unit reduction in sugar. The relatively high level of break even internalities for the older group is due to their consumption response to the tax being relatively small – they have particularly inelastic demand. It is much less clear than for the young, that the oldest group of consumers can expect to benefit highly from averted internalities due to a given amount of sugar reduction.

Figure 5.4: *Consumer welfare effects by age*

(a) Compensating variation



(b) Break even internality/100g sugar



Notes: Panel (a) shows annual compensating variation and panel (b) shows the value of internalities required to be averted per 100g reduction in sugar to make the consumer indifferent in the long run to the tax. In each case we show how the 25th, 50th and 75th percentiles of the distribution varies across age groups.

5.4 Redistributive concerns

Our discussion of consumer welfare in the preceding section asks, for each consumer, what level of internality must her sugar consumption generate for the soda tax to improve her welfare. However, a soda tax will also have redistributive implications that policymakers may care about. For instance, if low income consumers continue to purchase a lot of soda when it is subject to tax they may face a disproportionate share of the economic burden of the tax. On the other hand, if the tax strongly leads low income consumers to avoid internalities, either because they reduce their sugar consumption by a large amount or incur large internalities from a given amount of sugar consumption, then this may overcome any regressivity associated with the traditional economic burden of tax. Indeed, it is possible a soda tax is regressive when one considers only who pays the tax, but progressive when one also takes account of who, in the long run, benefits from averted internalities.

To assess the redistributive implications of soda taxes we use a proxy for consumers' income based on their total grocery expenditure. For each individual in our sample we observe all expenditure their household makes on groceries (including food and drink, alcohol and cleaning products) over the course of the year. We compute the total annual equivalised household grocery expenditure and use this as a proxy for individual income – we expect wealthier individuals to systematically spend more per equivalised adult on groceries than poorer individuals.[13]

In Figure 5.5 we show the effects of the sugary soda tax across deciles of the distribution of total equivalised grocery expenditure. Panel (a) shows that compensating variation is declining moving from low to higher deciles of the expenditure distribution. Based on their decision utilities (i.e. ignoring internalities) low spending individuals would require to be paid more to be indifferent to the introduction of the tax. Ignoring the internality rationale for the tax would lead one to conclude its introduction is regressive. However, panel (b) shows that the tax achieves considerably larger reductions in sugar for consumers in the bottom two deciles of the expenditure distribution than for higher spending individuals (nearly twice as large at the median). Therefore, low expenditure consumers may benefit more from averted internalities.

Panel (c) shows that the level of internality per 100g of sugar required for long run welfare of the consumer to be made no worse off from the introduction of the tax is broadly constant across deciles of the grocery expenditure distribution. Thus if low income individuals suffer more internality per 100g of sugar, a soda tax is likely

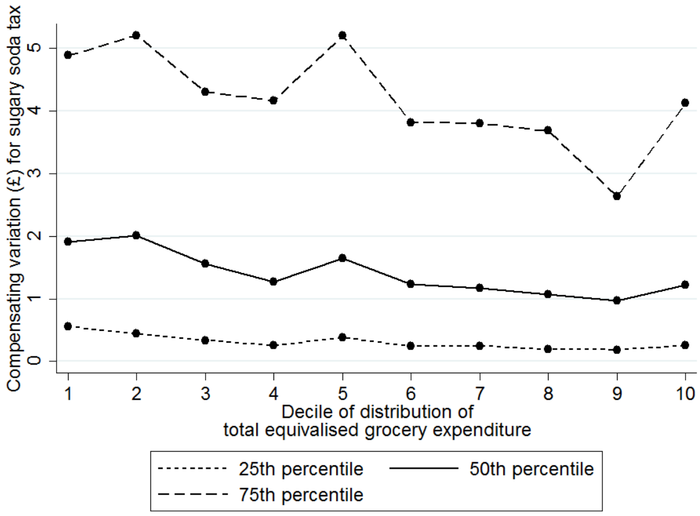
[13]We use the OECD modified equivalence scale, see Hagenaars et al. (1994).

to be progressive (once the full incidence – including its impact on internalities) is taken into account.

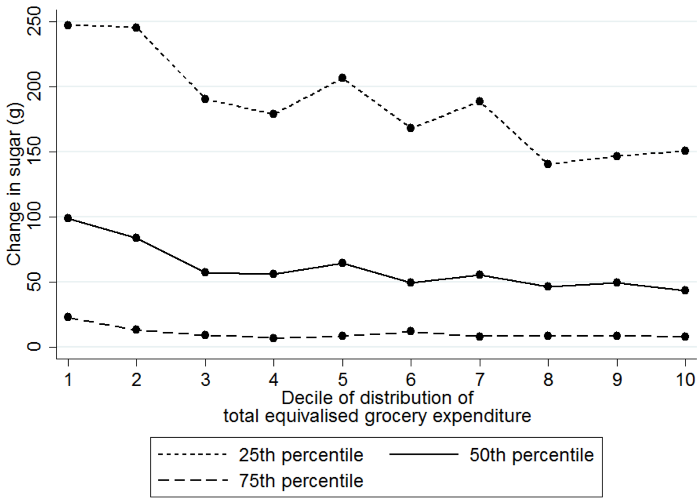
There is evidence to suggest that this is indeed the case. For example, Haushofer and Fehr (2014) and Mani et al. (2013) suggest that the stress and cognitive loads of being in poverty means people are more likely to make unwise decisions and underweight the future. Focusing on asset accumulation, in a recent paper, Bernheim et al. (2015) argue that poverty can perpetuate itself by undermining the capacity for self control: low initial wealth precludes self control, and hence asset accumulation, creating a poverty trap. Banerjee and Mullainathan (2010) take an alternative approach by assuming that "temptation goods" are inferior goods, which leads to a similar conclusion that self control problems give rise to asset traps.

Figure 5.5: *Consumer welfare effects by expenditure*

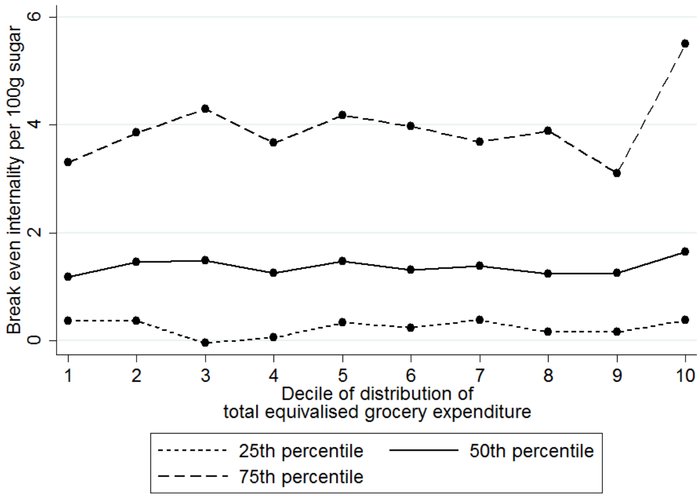
(a) Compensating variation



(b) Sugar reduction



(c) Break even internality/100g sugar



Notes: Panel (a) shows annual compensating variation, panel (b) shows reduction in sugar and panel (c) shows the value of internalities required to be averted per 100g reduction in sugar to make the consumer indifferent in the long run to the tax. In each case we show how the 25th, 50th and 75th percentiles of the distribution varies across deciles of the distribution of total annual equivalised household grocery expenditure.

6 Substitution to sugar in food

Our analysis so far has considered the impact of a soda tax, incorporating rich patterns of consumer switching across drinks (including both sodas and alternatives). We have thus far not modelled the possibility that consumers respond to the tax by switching from soda to foods. Ex ante such switching seems likely to be of much smaller magnitude than substitution towards alternative drinks and there is some limited medical evidence that calories from liquids do not displace those from solid (see, for instance, DiMeglio and Mattes (2000)). In this section we explore how important consumer switching from sugar in soda to sugar in food in response to a soda tax is likely to be. As it would be numerically very hard to estimate a model with all food on the go items simultaneous choices, we embed our drinks model into a two stage food on-the-go choice model where the idiosyncratic unobserved shocks on affecting the choice of which drink to consume are unknown on the first stage so that we can simplify the choice model between drinks and non drinks taking into account all the heterogeneity of tastes and preferences of consumers for drinks but not the idiosyncratic iid extreme value shock in the first stage. These shocks are then taken in expectation in the first stage.

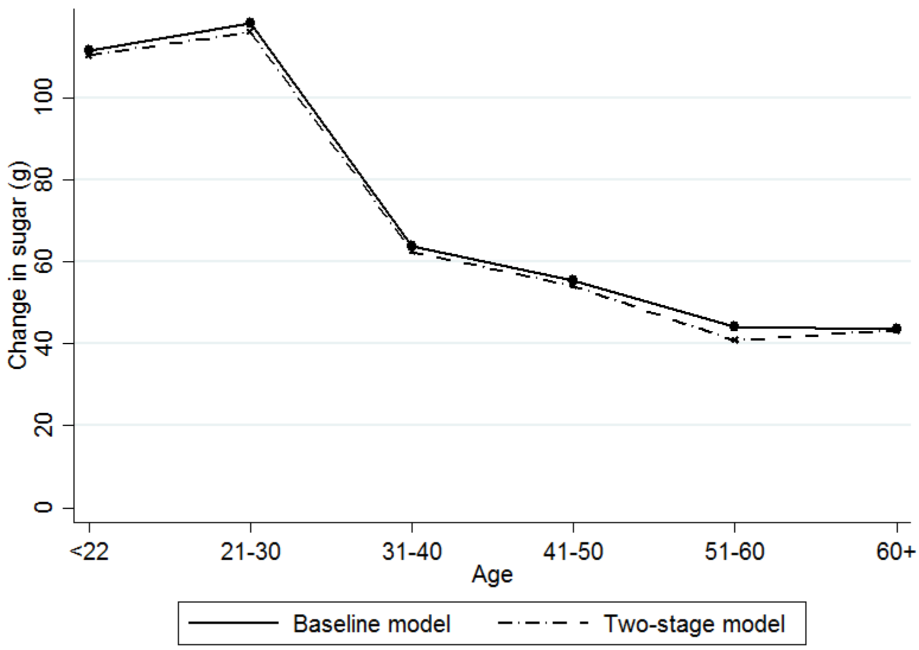
Thus, we suppose the choice model of Section 3.2 is a second stage of a two-stage decision process, which governs, conditional on choosing a drink, which drink to select. Consider a first stage in which the consumer chooses between chocolate products, choosing a non-sugary snack and choosing to select a drink. Let $k = \{\varnothing, 1, ..., K, \mathcal{D}\}$ denote first stage options. $k = \varnothing$ denotes the first stage outside option of a non-sugary 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:

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

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 and where $W_{ikt} = \alpha_i p_{krt} + \beta_i s_k + \vartheta_{b(k)}$ is the product specific utility from choosing chocolate product k. 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 the Online Appendix for further details).

We estimate the extended choice model allowing both constants in the drinks pay-off, $\mu_{i\mathcal{D}}$, and the parameter on the expected second stage utility from drinks, $\psi_{i\mathcal{D}}$, to vary across the 6 age groupings shown in the horizontal axis in Figure 6.1. For each age group the coefficient estimate is positive and statistically significant indicating that an increase in the price of soda (and thus a fall in the expected utility from choosing a drink) does induce some switching away from drinks and towards foods. However, Figure 6.1 illustrates that the strength of this switching to food is very small: it shows the median reduction in sugar due to the soda tax for the baseline model (results presented in Section 5) and the extended two-stage model. Taking account of switching to food sources of sugar dampens the median overall reduction in sugar by between 0.5% (for those aged over 60) to 7.4% (for those aged 51-60) and has no bearing on the qualitative relationship that sugar reductions are considerably larger for younger individuals. More broadly, none of our conclusions about the impact of a soda tax are materially affected by accounting for the (very limited) switching to sugar in food.

Figure 6.1: *Change in sugar by age; switching to food*



Notes: Figure shows how the median sugar reduction varies across age groups for both the baseline model and the extended two-stage model (which incorporates the possibility of switching to food).

7 Summary and conclusion

Corrective taxes have traditionally been applied to alcohol, tobacco and gambling. Recently there has been a drive to extend them to cover some types of foods, with soda taxes being at the vanguard of this move. The principal economic rationale for such taxes is that they discourage consumption that generates costs not taken account by individuals at the point of consumption. In the case of sugar, there is clear medical evidence that excess consumption can lead to very large future health costs, while almost all individuals exceed official recommendation on how much to consume. It is plausible that, at least for some consumers, these health costs are not factored in at the point of consumption. This is most obviously true for children, but is also likely to be the case for some individuals with high sugar diets and who therefore are at elevated risk of suffering health problems. The efficacy of a soda tax relies on to what extent it can encourage these groups to avoid internalities and at what cost to consumers in terms of welfare loss associated with higher prices.

Our results show that young consumers would lower their sugar consumption by more than older individuals in response to a soda tax. The tax does therefore succeed in achieving relatively large reductions in sugar among one group most likely to suffer from internalities. However, the young also loose out most in terms of direct consumer surplus loss due to higher prices. The relatively large internalities some young people impose on themselves makes it likely that the gain from averted internalities will outweigh this. The performance of the tax in terms of reducing the sugar intake of those with the most sugary diets is less good – those with high sugar diets are relatively price inelastic and therefore fail to lower their sugar consumption in response to the tax by more than more moderate sugar consumers. Nevertheless, if internalities are sufficiently convex in total sugar, this group may still benefit from the tax. The redistributive properties of the tax are more attractive than one based purely on traditional economic tax incidence. While the traditional economic burden of the tax falls disproportionately on the poor, the poor also lower their sugar consumption by a relatively large amount and therefore are likely to benefit by more than better off consumers due to averted internalities.

In our analysis we have taken account both of consumer demand responses and the equilibrium pricing response of soda manufacturers. In the longer run we would expect firms to also respond to the tax by changing their product portfolios and changing the sugar content of existing products. Our results therefore provide a picture to the short-medium run impact of soda taxes. An important direction for future work will be to incorporate how firm portfolio choice will be affected by such policies.

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.

Azaïs-Braesco, V., D. Sluik, M. Maillot, F. Kok, and L. A. Moreno (2017). A review of total & added sugar intakes and dietary sources in Europe. *Nutrition Journal 16*, 6.

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.

Banerjee, A. and S. Mullainathan (2010, May). The Shape of Temptation: Implications for the Economic Lives of the Poor. Working Paper 15973, National Bureau of Economic Research.

Bernheim, B. D., D. Ray, and Ş. Yeltekin (2015). Poverty and Self-Control. *Econometrica 83*(5), 1877–1911.

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.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics 25*, 242–262.

Besley, T. J. and H. S. Rosen (1999). Sales Taxes and Prices: An Empirical Analysis. *National Tax Journal 52*(2), 157–78.

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.

Cavadini, C., A. M. Siega-Riz, and B. M. Popkin (2000, July). US adolescent food intake trends from 1965 to 1996. *Archives of Disease in Childhood 83*(1), 18–24.

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

Cutler, D., E. Glaeser, and J. Shapiro (2003). Why Have Americans Become More Obese? *The Journal of Economic Perspectives 17*(3), 93–118.

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

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

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.

Gilbert, D. T., M. J. Gill, and T. D. Wilson (2002, May). The Future Is Now: Temporal Correction in Affective Forecasting. *Organizational Behavior and Human Decision Processes 88*(1), 430–444.

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.

Hall, K. D., G. Sacks, D. Chandramohan, C. C. Chow, Y. C. Wang, S. L. Gortmaker, and B. A. Swinburn (2011, August). Quantification of the effect of energy imbalance on bodyweight. *The Lancet 378*(9793), 826–837.

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.

Haushofer, J. and E. Fehr (2014, May). On the psychology of poverty. *Science 344*(6186), 862–867.

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

Kaplow, L. (2012). Optimal Control of Externalities in the Presence of Externalities. *International Economic Review 53*(2), 487–509.

Lewbel, A. and K. Pendakur (2017). Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients. *Journal of Political Economy 125*(4).

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013, August). Poverty Impedes Cognitive Function. *Science 341*(6149), 976–980.

Mennella, J. A., N. K. Bobowski, and D. R. Reed (2016, June). The development of sweet taste: From biology to hedonics. *Reviews in Endocrine and Metabolic Disorders 17*(2), 171–178.

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.

Nielsen, S. J. and B. M. Popkin (2004, October). Changes in beverage intake between 1977 and 2001. *American Journal of Preventive Medicine 27*(3), 205–210.

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

Rabin, M. (1998). Psychology and Economics. *Journal of Economic Literature 36*(1), 11–46.

Read, D. and B. Van Leeuwen (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational behavior and human decision processes 76*(2), 189–205.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

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.

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

A Appendix

A.1 Data description

In Table A.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 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 A.1 also shows that males and females under the age of 40 are more likely to purchase soda than older people.

Table A.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.

Of the 2,563 consumers with positive soda demands, we 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 also distinguish between consumers who, when buying an inside option, always, sometimes or never choose a sugary drink. Table A.2 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 A.2: *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.

In the main paper we show results by age groups < 22, $21-30$, $31-40$, $41-50$, $51-60$ and 60+. Table A.3 shows the fraction of consumers in our sample that fall into each group.

Table A.3: *Age groups*

Group	% of sample
< 22	10.9
$21-30$	16.8
$31-40$	28.0
$41-50$	22.8
$51-60$	13.9
60+	7.6

Notes: Numbers are fraction of individual-years that fall into each age category.

In the main paper we use two individual attributes that are measured based on the grocery purchases of the households to which the individuals belong. For each individual in our sample, in addition to observing their on-the-go food and drink purchases, we observe all of the groceries (including food and drink, alcohol and cleaning products) they purchase and take into the home. This household data are from the Kantar Worldpanel and are longitudinal; they contain observation for each household over many weeks.

The first measure is the share of total household calories that are from added sugar. We are able to compute this as the Kantar Worldpanel contains details on the nutrients of all food and drink products. The second measure is the total annual equivalised household grocery expenditure, where we equivalise expenditure using the standard OECD modified equivalence scale (see Hagenaars et al. (1994)). For each measure we show how the effects of the soda tax vary across deciles of the distribution. Table A.4 shows the upper bound for each (of the first 9) deciles.

Table A.4: *Household measure deciles*

	Upper bound of decile								
	1	2	3	4	5	6	7	8	9
Share of calories from added sugar (%)	7.93	9.08	10.01	10.84	11.69	12.46	13.42	14.65	16.71
Annual equivalised household grocery expenditure (£1000)	0.84	1.12	1.33	1.55	1.75	1.94	2.21	2.49	2.95

Notes: Deciles are defined over individual-years.

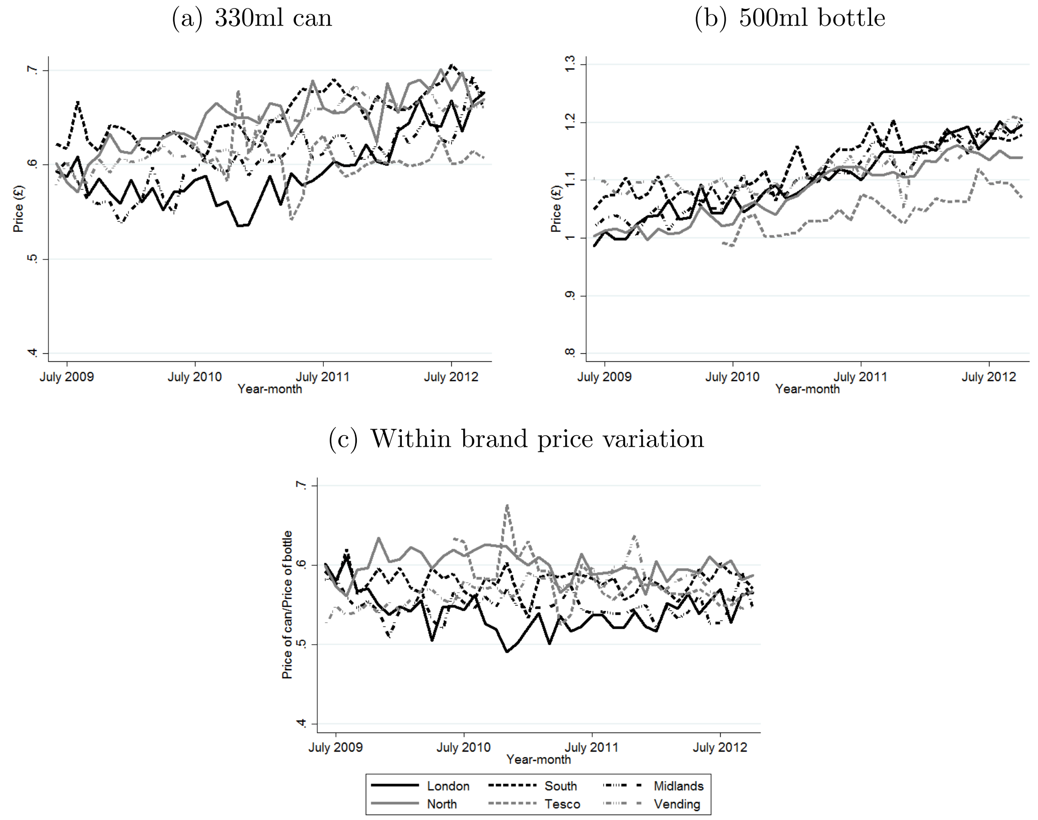
A.2 Prices

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 80% 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 3.3), 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 A.1 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 retailers. 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 A.1: *Price variation for Coca Cola*



Notes: Each line corresponds to a different retailer.

A.3 Demand Model Estimates

In Table A.5 we summarise the parameter estimates – obtained by maximising the likelihood function (equation 3.3). The top panel summarises the estimates of the consumer specific preference parameters for the price, soda and sugar attributes, reporting moments of 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.[14] The reported brand effects are for the first

[14]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.

period in the data (June 2010). We allow each of them to vary through time (from month-to-month).[15]

[15]We do not report the time varying brand effects or the retailer effects in Table A.5. These are available upon request.

Table A.5: *Model estimates*

Moments of distribution of consumer specific preferences

Variable		Estimate	Standard error
Price	Mean	-2.8349	0.0728
	Standard deviation	3.0401	0.0480
	Skewness	-1.4532	0.1051
	Kurtosis	5.8163	0.6329
Soda	Mean	0.1490	0.0965
	Standard deviation	2.3738	0.0387
	Skewness	1.2065	0.0815
	Kurtosis	5.0141	0.3733
Sugar	Mean	0.0550	0.0164
	Standard deviation	1.8340	0.0194
	Skewness	-0.0014	0.0606
	Kurtosis	3.9429	0.2341
Price-Soda	Covariance	-5.6252	0.3427
Price-Sugar	Covariance	-1.0102	0.2236
Soda-Sugar	Covariance	0.4631	0.2928

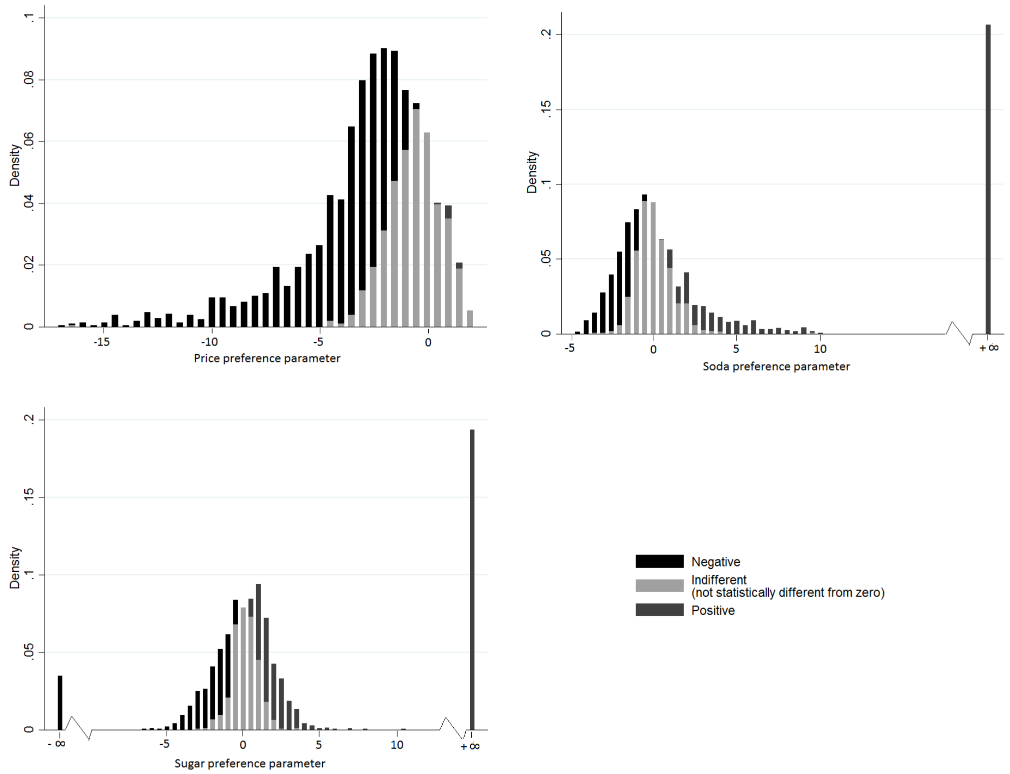
Consumer group specific preferences

Variable	Estimate	Standard error	Estimate	Standard error
	Female - <40		*Female - 40+*	
288ml carton	1.1305	0.0491	0.5030	0.0740
380ml bottle	2.0740	0.0538	2.1254	0.0586
500ml bottle	2.1375	0.0594	2.3207	0.0710
Fanta	-1.8766	0.1614	-1.6256	0.1550
Cherry Coke	-1.6554	0.1483	-2.3570	0.1971
Ribena	-1.3173	0.1330	-1.3315	0.1439
Pepsi	-0.9898	0.0985	-0.9599	0.1068
Lucozade	-1.7899	0.1781	-1.1734	0.1452
Oasis	-2.3789	0.1754	-1.8816	0.1589
Fruit juice	0.2044	0.3039	2.4005	0.3576
Flavoured milk	-3.2606	0.2764	-2.3051	0.3911
	Male - <40		*Male - 40+*	
288ml carton	-0.3100	0.0636	-0.1638	0.0694
380ml bottle	2.0204	0.0462	2.2625	0.0543
500ml bottle	2.3225	0.0551	2.1029	0.0637
Fanta	-1.6338	0.1287	-1.2785	0.1191
Cherry Coke	-2.1061	0.1611	-2.1001	0.1880
Ribena	-2.1274	0.1702	-1.3759	0.1428
Pepsi	-1.5547	0.1127	-0.7731	0.0922
Lucozade	-1.4242	0.1373	-1.2493	0.1236
Oasis	-2.2141	0.1751	-2.6324	0.2162
Fruit juice	1.2629	0.3314	-1.1853	0.4006
Flavoured milk	-2.3016	0.2525	-4.1928	0.3366
Time-demographic-brand effects		Yes		
Retailer-demographic-brand effects		Yes		

Notes: We have an initial sample of 2,563 soda consumer. We estimate demand on this sample and eliminate from the sample a set of consumers with strongly upward sloping demands. We then re-estimate on the remaining 2,183 consumers and 150,426 choice occasions. Estimates are summarised in the table. Moments of distribution of heterogeneous 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.

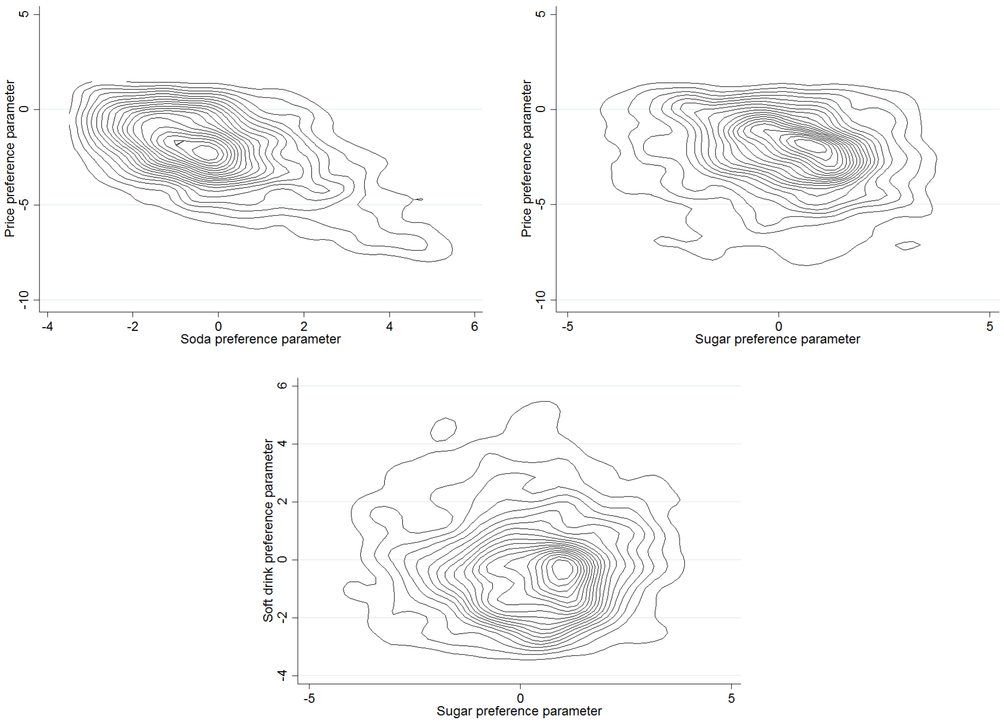
We do not need to impose any distributional assumption on consumer preferences over price, soda and sugar and in particular we do not assume the marginal distributions are normal as is common in random coefficient models. The skewness and kurtosis of the price, soda and sugar preference distributions indicate departures from normality – for instance price preferences are negatively skewed and leptokurtic (i.e. kurtosis above 3 indicating fatter tails than a normal distribution). In addition, the sugar and soda preferences distributions have infinite portions. We plot the marginal distributions in Figure A.2. The shading represents consumers with negative, positive and indifferent (i.e. not statistically significantly different from zero) preferences for each attribute. In Figure A.3 we plot contour plots of the bivariate preference distributions (based on the finite parts of the distribution).

Figure A.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.

Figure A.3: *Bivariate distributions of consumer specific preference parameters*



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

A.4 Substitution to food

The choice model we outline in Section 3.2 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 juice $j = \{j' + 1, ..., J\} = \Omega_{nw}$ The expected utility to the consumer of purchasing a drink is:

$$\begin{aligned}E_{\epsilon_{ijt}}\left[\max_{j\in\Omega_{\mathcal{D}}} U_{ijt}\right] =& \ln\Bigg(\exp\left(\xi_{d(i)0t} + \zeta_{d(i)0t}\right) + \sum_{j\in\Omega_w\cup\Omega_{nw}} \exp\big(\alpha_i p_{jrt} + \beta_i s_j + \\ &\gamma_i w_j + \delta_{d(i)} z_j + \xi_{d(i)b(j)t} + \zeta_{d(i)b(j)r}\big)\Bigg) \\ \equiv& W_{i\mathcal{D}t}.\end{aligned}$$

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:

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

where

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

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:

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

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

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

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

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

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 $\tilde{\boldsymbol{\vartheta}} = (\tilde{\vartheta}_1, ..., \tilde{\vartheta}_B)'$, $\mu_{i\mathcal{D}}$ and $\psi_{i\mathcal{D}}$.

We allow for heterogeneity in the parameters $\mu_{i\mathcal{D}}$ and $\psi_{i\mathcal{D}}$ across age groups. Table A.6 shows estimates of these parameters.

Table A.6: *Upper stage model estimates*

	$\hat{\mu}_{iD}$		$\hat{\psi}_{iD}$	
Age group	Estimate	Standard error	Estimate	Standard error
< 22	0.1325	0.0288	0.1312	0.0062
$22-30$	-0.5120	0.0227	0.2543	0.0050
$31-40$	-0.4609	0.0184	0.2844	0.0041
$41-50$	-0.4683	0.0192	0.2304	0.0044
$51-60$	-1.4209	0.0246	0.4759	0.0048
60+	-0.5353	0.0405	0.2142	0.0092

Notes: Estimates based on sample of 324,818 choice occasions. Chocolate brand effects were also estimated.

Online Appendix

Individual preference heterogeneity, targeting and welfare effects of soda taxes

Pierre Dubois, Rachel Griffith and Martin O'Connell

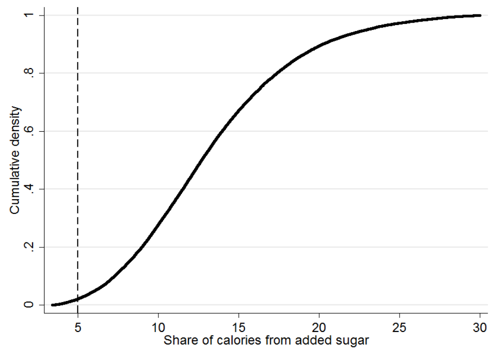
September 11, 2017

A Purchase patterns in US

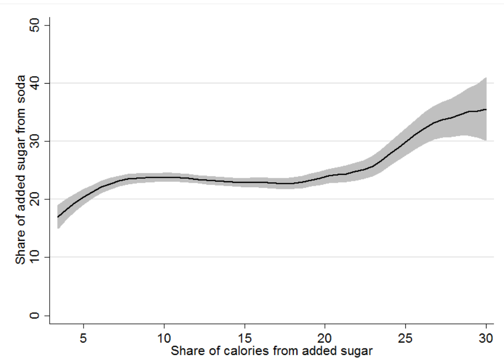
Section 3.1 of the main paper shows that in the UK a) the majority of consumers obtain considerably more of their calories from added sugar than the World Health Organization recommendation of no more than 5% of calories and b) the share of added sugar obtained from soda is increasing in the share of added sugar in individuals' diet and decreasing in age. Using the National Health and Nutrition Examination Study over 2007-2014, a sample of 39,189 adults and children, we show very similar patterns hold for the US. In Figure A.1 we use these data to show this. The patterns shown in Figure A.1 replicate what we see in the UK (Figure 3.1 of the main paper).

Figure A.1: *Added sugar and soda*

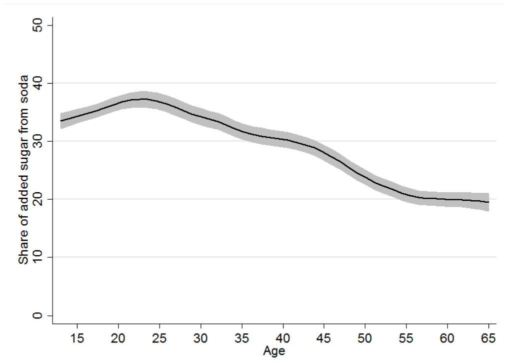
(a) Share of calories from added sugar



(b) Soda and added sugar



(c) Soda and age



Notes: Numbers computed using National Health and Nutrition Examination Study over 2007-2014. Vertical line in panel (a) denotes the WHO target of no more than 5% of calories from added sugar. Shaded areas in panels (b) and (c) denote 95% confidence intervals.

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 fixed effect 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 $\theta = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta})$ based on splitting the sample into two non overlapping random samples. 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 jackknife (bias corrected) estimator is:

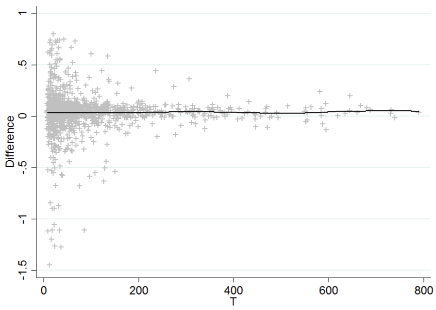
$$\widetilde{\theta}_{split} = 2\widehat{\theta} - \frac{\widehat{\theta}_{(1,T/2)} + \widehat{\theta}_{(T/2,T)}}{2}$$

Figures B.1, B.2 and B.3 show, for the price, soda and sugar preference parameters, how the jackknife ($\widetilde{\theta}_{split}$) 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) age. 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.

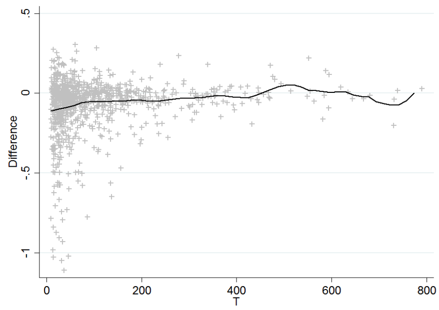
Figures B.4 plots the distributions of price, soda and sugar preference parameter estimates for both the estimators $\widehat{\theta}$ and $\widetilde{\theta}_{split}$, showing there is very little difference in the distributions.

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

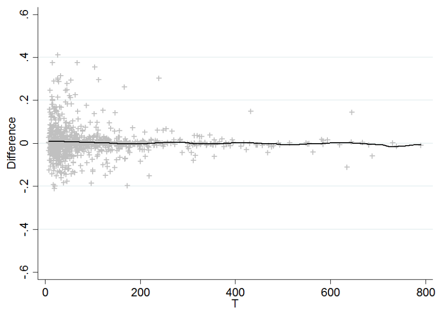
(a) Price



(b) Soda



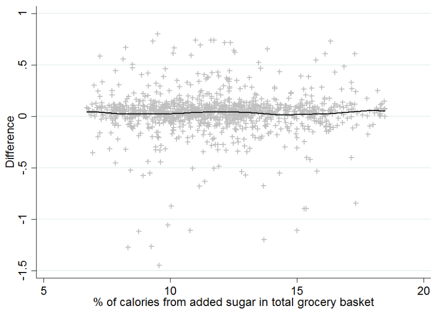
(c) Sugar



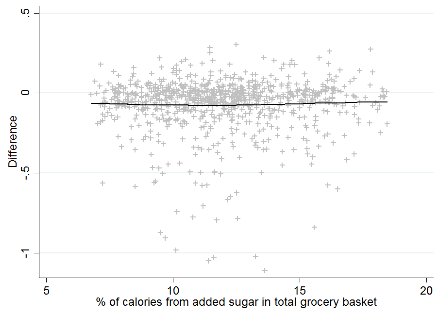
Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

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

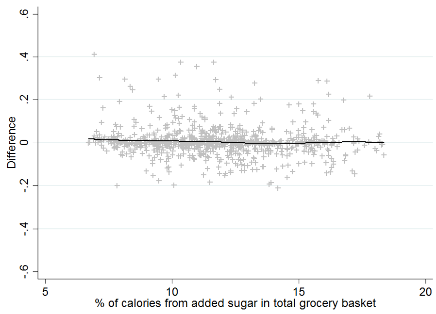
(a) Price



(b) Soda



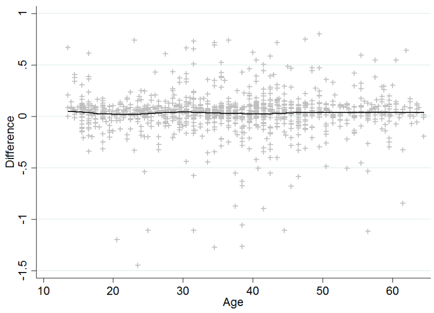
(c) Sugar



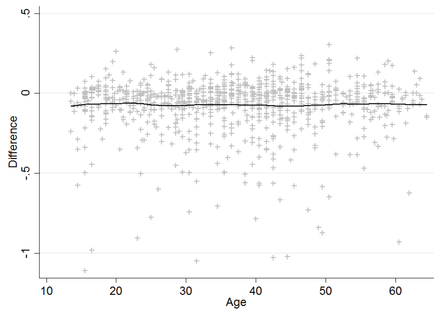
Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

Figure B.3: *Relationship between bias and age*

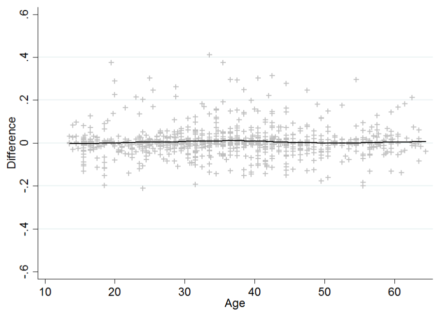
(a) Price



(b) Soda



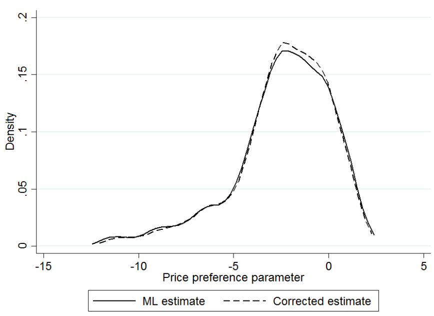
(c) Sugar



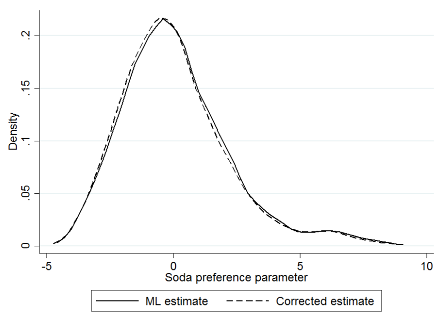
Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

Figure B.4: *Preference parameter distribution*

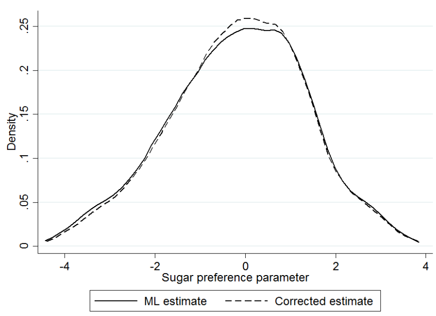
(a) Price



(b) Soda



(c) Sugar



Notes: Lines are kernel density estimates.

C An alternative soda tax

The main paper focuses on the impact of a soda tax incidence only on sugary sodas. We also simulate the impact of a soda tax incidence on all soda products (both regular and diet); this tax takes the form

$$p_{jt}^{cf} = \begin{cases} \tilde{p}_{jt}^{cf} + \tau w_j & \forall j \in \Omega_w \\ \tilde{p}_{jt}^{cf} & \forall j \in \Omega_{nw}. \end{cases}$$

Here we refer to this as a broad soda tax and the tax we focus on in the main paper as a sugary soda tax. We simulate the same rate for the broad soda tax as for the sugary soda tax (25 pence per litre) using the same supply side model estimate sin te first step and conducting the estimation of this tax pass-through in consumer prices.

Table C.1 summarises the impact of the broad soda tax on equilibrium prices and market shares (it is the analogue of Table 5.1 in the main paper). The main difference between a tax incident on only sugary and one incident on all sodas is that the latter leads to prices increases for diet products (that on average are similar to those for sugary products). The result is that the broad soda tax leads to a much smaller reduction in demand for sugary soda and a fall (rather than increase) in demand for diet sodas (relative to the sugary soda tax).

Table C.1: *Impact of "broad" soda tax on market equilibrium*

	Sugary soda	Diet soda	Sugary alternatives	Outside option
Tax (£)	0.11	0.11	0.00	0.00
Δ price (£)	0.12	0.12	0.00	0.00
Δ share (p.p)	-1.83	-1.75	0.73	2.84

Notes: Numbers are mean across products. Tax and price change are weighted by market share.

Table C.2 illustrates that at the 25th, 50th and 75th percentiles (and indeed at all points) of the distribution of sugar reductions the broad soda tax achieves much smaller sugar reductions than a sugary soda tax (see Table **??** in the main paper). Figure C.1 shows that, although the magnitude of sugar reductions is lower for the broad soda tax, the patterns of how sugar reductions vary with both added sugar and age are similar to those for the sugary soda tax (see Figures 5.2(a) and **??**(b) in the main paper).

As the broad soda tax achieves smaller reductions in sugar, raises average drinks prices by more and is not any better targeted than a sugary soda tax it performs unambiguously worse as a corrective tax. The median break even internality (i.e. internality reduction required per 100g sugar reduction to make consumers indifferent to the tax in the long run) is over twice as large for the broad soda tax. Of course, the flip side of this is that the broad soda tax raises more revenue – around 1.9 times as much as the sugary soda tax.

Table C.2: *Effect of "broad" soda tax on grams of sugar from drinks (per year)*

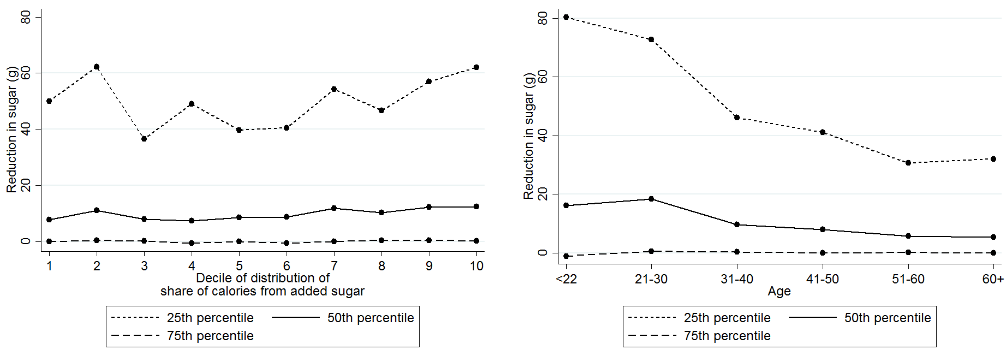
	Percentile of distribution		
	25th	50th	75th
Sugar from soda	1	16	76
Sugar from drinks	0	9	49

Notes: For each year we compute the 25th, 50th and 75th percentile of the sugar reduction distribution. The table presents the man of these across years.

Figure C.1: *Effect of "broad" soda:*

(a) By added sugar decile

(b) By age



Notes: Panel (a) shows variation in sugar reductions across deciles of the distribution of total calories from added sugar in annual household grocery baskets; Panel (b) shows variation across age groups. In age case we show how the 25th, 50th and 75th percentiles of the distribution of sugar reductions varies.