DISCUSSION PAPER SERIES

DP12484

HOW WELL TARGETED ARE SODA TAXES?

Pierre Dubois, Rachel Griffith and Martin O'Connell

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

HOW WELL TARGETED ARE SODA TAXES?

Pierre Dubois, Rachel Griffith and Martin O'Connell

Discussion Paper DP12484
Published 04 December 2017
Submitted 04 December 2017

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

HOW WELL TARGETED ARE SODA TAXES?

Abstract

Soda taxes aim to reduce excessive sugar consumption. Their effectiveness depends on whether they target individuals for whom the harm of consumption is largest. We estimate demand and account for supply-side equilibrium pass-through. We exploit longitudinal data to estimate individual preferences, which allows flexible heterogeneity that we relate to a wide array of individual characteristics. We show that soda taxes are effective at targeting young consumers but not individuals with high total dietary sugar; they impose the highest monetary cost on poorer individuals, but are unlikely to be strongly regressive if we account for averted future costs from over consumption.

JEL Classification: D12, H31, I18

Keywords: preference heterogeneity, discrete choice demand, Pass-Through, soda tax

Pierre Dubois - pierre.dubois@tse-fr.eu
Toulouse School of Economics and CEPR

Rachel Griffith - rgriffith@ifs.org.uk
Institute for Fiscal Studies and University of Manchester and CEPR

Martin O'Connell - martin_o@ifs.org.uk
Institute for Fiscal Studies and CEPR

How well targeted are soda taxes?

Pierre Dubois, Rachel Griffith and Martin O'Connell*

November 2017

Abstract

Soda taxes aim to reduce excessive sugar consumption. Their effectiveness depends on whether they target individuals for whom the harm of consumption is largest. We estimate demand and account for supply-side equilibrium pass-through. We exploit longitudinal data to estimate individual preferences, which allows flexible heterogeneity that we relate to a wide array of individual characteristics. We show that soda taxes are effective at targeting young consumers but not individuals with high total dietary sugar; they impose the highest monetary cost on poorer individuals, but are unlikely to be strongly regressive if we account for averted future costs from over consumption.

Keywords: preference heterogeneity, discrete choice demand, pass-through, soda tax

JEL classification: D12, H31, I18

Acknowledgements: 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 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)). Soda taxes have been proposed as a way to reduce sugar consumption, particularly for individuals whose consumption generates costs that are borne by others (externalities) or for whom the future costs of excess consumption are large and are partially ignored at the point of consumption (internalities). Internality correcting taxes have been advocated for unhealthy foods (O'Donoghue and Rabin (2006), Haavio and Kotakorpi (2011)), as the principal justification for high levels of cigarette taxation (Gruber and Koszegi (2004)), and in energy markets (Allcott et al. (2014)). A growing number of jurisdictions are adopting taxes on soda.[1] Whether such measures will succeed in improving public health depends on how individuals' demand responses correlate with the size of any unanticipated costs that their soda consumption imposes on themselves in future and the costs on others.

Our contribution in this paper is to provide evidence on how well targeted soda taxes are; in particular, are they effective at lowering the sugar consumption of individuals for whom the consequences of high levels of soda consumption are most severe, and for whom internalities are most likely to be important. We estimate consumer choice in the drinks market and simulate the introduction of a soda tax, accounting for pass-through to prices. We show that soda taxes are effective at targeting young consumers, do not target individuals with high total dietary sugar, impose the highest monetary cost on poorer individuals, but are unlikely to be strongly regressive if we account for averted future costs from over consumption. Relative to the existing literature we make two main advances.

First, we model consumer preferences as individual level parameters that we estimate. This departs from the standard approach to modeling consumer preference heterogeneity in discrete choice models, where preferences are treated as random effects drawn from a mixing distribution. The main advantage of our approach is that it enables us to directly relate individual level predictions of the impact of the tax to individual characteristics in a very flexible way. This means that we can assess precisely which individuals respond to the tax and on whom the economic

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

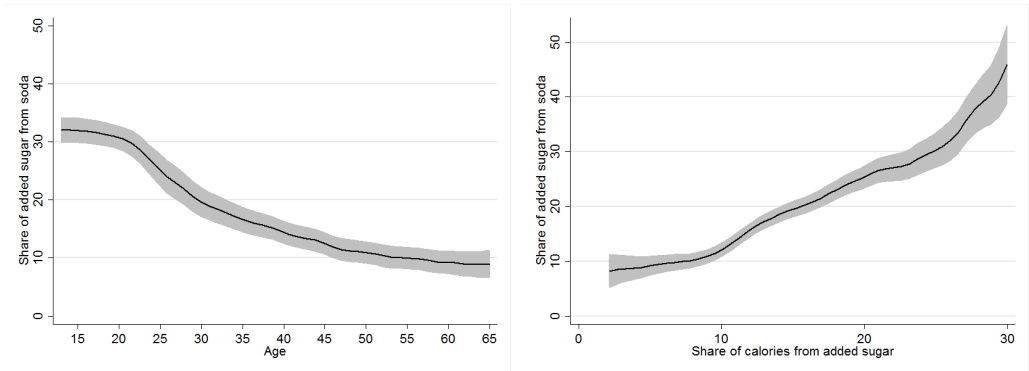
burden of the tax falls most heavily; in other words is a tax well targeted and how regressive is it?

Second, we study individual purchase decisions made for immediate consumption on-the-go using novel longitudinal data on a representative sample of British individuals (including teenagers and young adults). Around half of soda purchases are made on-the-go, making it an important part of the market on which we have little evidence on choice behavior. On-the-go purchases are made by individuals for immediate consumption – most of the literature on choice behaviour studies purchases made in supermarkets and brought into the home for future consumption. A significant advantage of individual level on-the-go data is that they allow us to estimate individual level preferences without the need to place strong restrictions on the intra-household preference structure (see, for example, Adams et al. (2014)). In addition, young adults are a particular group of interest and are typically not identified as a distinct group in data based on household shopping data.

The propensity for people to over consume sugar, the effects that excessive intake has on health and other future outcomes and the role soda plays as a significant contributor to total dietary sugar is well established (see WHO (2015)). Figure 1.1 shows that there is a strong gradient in sugar obtained from soda with both age and total dietary sugar; young people and those for whom sugar represents a high share of the total calories that they purchase (high total dietary sugar individuals) tend to get particularly large amounts of sugar from soda.[2] This suggests that soda taxes are potentially well motivated; sugar consumption is well above medical recommendations, soda represents a substantial share of this, and soda intake is especially high for the young and for individuals with high total dietary sugar.

Figure 1.1: *Sugar from soda*

(a) by age (b) by total dietary sugar



Notes: Numbers using National Diet and Nutrition Survey 2008-2011 for a representative sample of 3,073 UK adults and children. Shaded areas denote 95% confidence intervals.

[2]We show similar patterns hold in the US in Appendix A.1.

However, the effectiveness of a soda tax depends not only on the extent to which individuals consume soda prior to the introduction of the tax, but also on how strongly they switch away from the sugar in soda and what alternatives they switch to. To assess the targeting of the tax we need to know how demand responses vary across markers of likely harm from soda consumption (like age and total dietary sugar); to assess the redistributive consequences we need to know how they vary across the income distribution. We estimate a structural model of demand and supply that allows us to identify individual specific preference parameters and enables us to relate the effects of a soda tax in a flexible way to individual demographics and measures of income.

To model consumer choice we use a discrete choice framework in which consumer preferences are defined over product attributes. Like much of the literature on choice models (Berry 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 consumer level parameters to be estimated (rather than random draws from a mixing – or random coefficient – distribution). This means that we can recover any arbitrary relationship between the individual preference parameters and functions of them, such as the predicted outcomes from a tax simulation, with any attributes of the individual consumers. In contrast, in standard random coefficient models the interactions between consumer attributes and the multivariate preference distribution need to be specified ex ante and a specific functional form imposed. Our approach entails estimating fixed effects in a non-linear model and therefore may suffer from an incidental parameters problem (Hahn and Newey (2004), Arellano and Hahn (2007)). We show robustness to using the split sample jackknife bias correction procedure suggested in Dhaene and Jochmans (2015).

We find that preferences vary over consumer attributes in ways that would be difficult to capture by specifying a priori a random coefficient distribution. For instance, our estimates show that individuals below 30 years of age tend to be more price sensitive and have stronger preferences for sugar than older age groups. Among younger consumers preferences over sugar and prices are negatively correlated – those with strong sugar preferences are most price sensitive – but among older consumers the correlation is reversed. We also show how the effects of a soda tax vary over the joint distribution of age, total dietary sugar and a proxy for income, while placing minimal restrictions on the joint distribution of preferences; we impose that an individual's preferences are stable over choice occasions, which allows us to use the long time dimension of repeated purchases to identify individual preferences.

Our estimates suggest that an excise style tax on sugary soda would be over shifting on to consumer prices and lead to marginally lower prices of diet products. An important factor in driving the patterns of over shifting are strategic complementaries between the pricing of competing soda manufacturers. Firms' pricing response would therefore amplify the price differential that the tax creates between sugary and diet varieties.

We show that the sugary soda tax is well targeted at young people. In response to the tax, young adults (aged 13-21) reduce the amount of sugar they purchase from on-the-go soda by around 80% more than the average consumer. This is driven both by younger consumers being more likely to consume soda than older and, conditional on being soda drinkers, them responding more strongly than older people to the tax in terms of reducing the amount of sugar they purchase from soda. Crucial to finding this pattern is that our modeling allows the joint distribution of preference parameters to vary flexibly across individuals, and in particular our finding that sugar and price preferences are negatively correlated for young groups and positively correlated for older groups.

The tax is less effective at targeting those people with high total dietary sugar; despite getting large amounts of sugar from soda, those individuals with high total dietary sugar do not reduce the amount of sugar that they purchase from soda by any more than those with more moderate amounts of total dietary sugar. This is because individuals with high total dietary sugar have both particularly strong preferences for sugar and are relatively insensitive to price increases.

If consumers fully internalized the future costs of excess sugar consumption, we could measure the full effect on consumer welfare using individuals' revealed preferences to compute their compensating variations. However, if some people do not fully account for the future costs at the point of consumption, then the tax will have a second effect on welfare through averted future unanticipated costs (internalities). We can measure compensating variation, and we show that it is highest among individuals with high total dietary sugar and among young consumers (especially young consumers with high total dietary sugar). While there is experimental evidence that people have behavioural biases with respect to food and drink consumption (see, for instance, Read and Van Leeuwen (1998) and Gilbert et al. (2002)), measuring the extent of the internalities is challenging, and not something we attempt to do in this paper. However, we can get an idea of the full effect on consumer welfare by computing how much internality per reduction in sugar is required to make people indifferent to the introduction of a soda tax. For young consumers this number is

around £0.80 per typical 330ml can of sugary soda, for those in the top decile of the distribution of total dietary sugar the equivalent number is £1.40.

A common criticism of excise style taxes is they are regressive;[3] the poor typically spend a higher share of their income on the taxed good, and so bear a disproportional share of the burden of the tax. However, if the tax plays the role of correcting an internality, then the distributional analysis is more complicated; if low income consumers also save more from averted internalities this may overturn the regressivity of the traditional economic burden of taxation (Gruber and Koszegi (2004)). These redistributive concerns become more subtle when income transfers are considered (Lockwood and Taubinsky (2017)). We show that compensating variation associated with a sugary soda tax is around 40% higher for those in the bottom half of the distribution of total expenditure (based on a wide set of food, drink and non-drink items) compared with those in the top half. However, the reduction in sugar is also larger for these individuals, which leaves open the possibility that they will also benefit more from averted internalities, and so the full effect on consumer welfare is likely to be less negative than the compensating variation suggests.

The rest of this paper is structured as follows. In Section 2 we describe our model of consumer demand in the on-the-go drinks market and oligopoly pricing. In Section 3 we introduce our novel individual level data and summarize estimates of the demand model. In Section 4 we present results of the sugary soda tax simulation, discussing the impact on equilibrium pricing, how well targeted the measure is, the effects on consumer welfare and its distributional implications. In Section 5 we extend our demand model to incorporate broader patterns of consumer switching, including towards food, and show that our results are robust to inclusion of these additional margins of consumer response. A final section concludes.

2 Model

In this section we develop a model of demand and supply for sodas. We start by describing our demand model and discussing identification. We then describe a supply side oligopoly model that allows us to evaluate the equilibrium pass-through of a soda tax to consumer prices.

We estimate demand using novel longitudinal data on food and drink purchases that consumers make on-the-go (i.e. food and drinks bought from retailers for immediate consumption). This is both an important and understudied segment of

[3]For instance, see Senator Sanders op-ed on the Philadelphia soda tax, Sanders (2016).

the market. We have data on a sample of over 5000 individuals for which we observe a long history of on-the-go purchases that allows us to identify and estimate the demand model presented here; we describe this in Section 3.1.

2.1 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 take the decision to purchase a drink as exogenous. In Section 5 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, $j \in \{1, ..., j'\} = \Omega_a$, and alternative drinks (fruit juice and flavored milk), $j \in \{j' + 1, ..., J\} = \Omega_n$. We distinguish the set of sugary sodas, Ω_s, from diet sodas, Ω_d; $\Omega_a = \Omega_s \cup \Omega_d$. Each product belongs to a brand – we denote the brand that product j belongs to as $b(j)$. There are fewer brands than products; soda brands typically comprise two different sizes and a sugary and diet variety. We denote the outside option, bottled water, as $j = 0$.

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

$$U_{ijt} = \alpha_i p_{jrt} + \beta_i s_j + \gamma_i w_j + \delta_{d(i)} z_j + \xi_{d(i)b(j)t} + \zeta_{d(i)b(j)r} + \epsilon_{ijt}, \tag{2.1}$$

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).[4] 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. 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 to vary by gender and age – we denote the gender-age groups by $d \in \{1, ..., D\}$ 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:

$$U_{i0t} = \xi_{d(i)0t} + \zeta_{d(i)0r} + \epsilon_{i0t}, \tag{2.2}$$

[4]Here and below, for simplicity, we do not specify an r subscript to U_{ijt} and ϵ_{ijt}.

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.

We denote $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_N)'$ and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_N)'$ the vectors of individual preference parameter whose distribution needs not be a priori restricted. We use the large T dimension of our data to recover estimates of individual specific parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ while the large N dimension allows us to identify nonparameterically the joint probability distribution function $f(\alpha_i, \beta_i, \gamma_i)$ using the empirical probability distribution function of estimated $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. 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 behavior 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 is important for enabling choice models to capture switching patterns across products,[5] while a few papers have used non-parametric methods to relax parametric restrictions on random coefficients.[6] Like these papers we model consumer specific preferences, however, in contrast to them, we treat the preferences as parameters to be estimated and so we do not need to add conditional independence assumptions to integrate out their distribution conditional on some chosen observed characteristics. 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 $T \to \infty$, asymptotic bias may remain, although it shrinks as the sample size rises (Hahn and Newey (2004), Arellano and Hahn (2007)). The long T dimension of our data helps lower 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 3.3 show that our maximum likelihood and jackknife estimates are similar and that the bias correction does not 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 from

[5]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.

[6]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.

consumer stockpiling (a situation considered in Wang (2015)); by definition the consumption occasions that we are modeling do not involve storage. Another form of dynamics would arise if drinks preferences were intertemporally nonseperable; we assume this is not the case. However, our model is able to capture the propensity that different individuals may have to buy similar products over time through the rich individual level preference heterogeneity. Failure to account for such rich unobserved heterogeneity may lead to spurious state dependence.

Another benefit of having large T for each individual is that we can allow for some consumers who may have sufficiently strong distaste for some product sets that they endogenously will never choose to buy them. Contrary to standard logit discrete choice models, we allow for this possibility by allowing some consumers to have zero probability of purchasing certain products. We use the long time dimension of our data to identify consumers that never purchase products with particular characteristics (for example, products that are sodas, or products that are diet varieties) and allow them to have zero purchase probabilities for products that have that attribute. Assuming that the unobservable error term has 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, flavored 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 $j \in \Omega_a$ and $\sum_{j \in \Omega_n} P_{it}(j) = 1$. Consumers that never purchase non soda drinks can be thought of as having negative infinite preferences for non soda (which we denote by $\gamma_i = \infty$) 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 never buy diet products (or the outside option) have negatively infinite diet preferences (which we denote $\beta_i = \infty$). Those consumers observed purchasing both diet and sugary soda across their choice occasions have finite sugar preferences ($\beta_i \in (-\infty, \infty)$).

Since we incorporate the possibility that a consumer will never buy some products that contain a characteristic he sufficiently dislikes, the 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. As a consequence, it is convenient to define the consumer i specific set of products with

non-zero purchase probabilities, denoted by Ω_i, as

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

We assume that the consumer level products sets Ω_i are measured exactly due to the large T dimension of observed consumer level choices. However, our sample is finite and thus a finite sample measurement error is introduced on Ω_i. We will ignore such measurement error on the discrete set Ω_i for simplicity and also because Monte Carlo simulations show that such error is negligible in our application where T is relatively large.[7]

Then, we denote :

$$v_{ijt} \equiv \alpha_i p_{jrt} + \beta_i s_j 1_{\{\beta_i \in (-\infty,\infty)\}} + \gamma_i w_j 1_{\{\gamma_i \in (-\infty,\infty)\}},$$

$$\eta_{ijrt} \equiv \delta_{d(i)} z_j + \xi_{d(i)b(j)t} + \zeta_{d(i)b(j)r}, \ \eta_{i0rt} \equiv \xi_{d(i)0t} + \zeta_{d(i)0r}$$

such that (2.1) and (2.2) can be written

$$U_{ijt} = v_{ijt} + \eta_{ijrt} + \epsilon_{ijt}, \ U_{i0t} = \eta_{i0rt} + \epsilon_{i0t}.$$

Then, the assumption that ϵ_{ijt} is an idiosyncratic shock distributed type I extreme value means that the consumer level choice probabilities are given by the multinomial logit formula, such that the choice probability of consumer i purchasing any good j can be written:

$$P_{it}(j) = \frac{1_{\{\gamma_i \in (-\infty,\infty), j=0\}} \exp(\eta_{i0rt}) + 1_{\{j \in \Omega_i, j>0\}} \exp(v_{ijt} + \eta_{ijrt})}{1_{\{\gamma_i \in (-\infty,\infty)\}} \exp(\eta_{i0rt}) + \sum_{k \in \Omega_i} \exp(v_{ikt} + \eta_{ikrt})}$$

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

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

[7]Further details available from authors on request.

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

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}) = \sum_i \ln \mathcal{P}_i(\boldsymbol{y}_i), \tag{2.3}$$

which is globally concave with respect to all parameters.

2.2 Identification

Our main identification challenge is to pin down the causal impact of price on demand. Our strategy for doing this relies on two sources of price variation. First, 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, but rather is driven by daily life in which individuals move between home, school, leisure or work. Second, we exploit variation in prices within brand across different containers and sizes. While there may be some aggregate shock to demand for a specific brand (that manufacturers observe and change 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 at the point of purchase 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 (e.g. driven by national advertising) and the latter capture any differential propensity of consumers to choose different drink types across retailers. Conditional on these, the cross-retailer differences in prices provides a useful source of price variation that will be driven by cost shocks and idiosyncratic factors.

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 behavior would occur either if consumers could predict fluctuating relative prices across retailers or if they visited several retailers in search

[8] In Appendix A.2 we describe some of the variation in product prices.

of a low price draw for the product they are seeking. We find either scenario highly unlikely in the case of 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 little retailer or regional advertising of specific drinks products. 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 local discounts related to excess stock.

The second source of price variation we exploit is non-linear pricing across container sizes (this is common in the UK). This price variation is not collinear with the size effects and the extent of non-linear pricing varies over time and retailers. This source of identification would be invalidated if there were systematic shocks to consumers' valuation of container 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 products that belong to the same model. We assume that, conditional on time-varying brand characteristics, unobserved size specific attributes do not vary differentially across brands.

2.3 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 model a tax of 25 pence per liter, (or

[9] In Appendix C 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.

33 US cents per liter, which is 1.2 cents per ounce – similar to the US taxes of 1-1.5 cents per ounce).

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 manufacturers and F_f denote the set of products owned by f. For simplicity, we assume that prices are set by manufacturers and abstract from modeling manufacturer-retailer relationships – efficient vertical contracting would lead to such a price equilibrium. Normalizing 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. Firm f's (variable) profits in market t are given by:

$$\Pi_{ft} = \sum_{j \in F_f} (p_{jt} - c_{jt}) q_{jt}(\mathbf{p}_t) \tag{2.4}$$

and the firm's first order conditions are:

$$q_{jt}(\mathbf{p}_t) + \sum_{k \in F_f} (p_{kt} - c_{kt}) \frac{\partial q_{kt}(\mathbf{p}_t)}{\partial p_{jt}} = 0 \quad \forall j \in F_f. \tag{2.5}$$

Under the assumption that observed market prices are an equilibrium outcome of the Nash-Bertrand game played by firms and given our estimates of the demand function, we can invert the first order conditions to infer marginal costs c_{jt}.

The introduction of a soda tax creates a wedge between post-tax prices, $\mathbf{p}$, and pre-tax prices, which we denote $\tilde{\mathbf{p}}$. The volumetric tax on sugary soda implies pre-tax and post-tax prices are related by:

$$p_{jt} = \begin{cases} \tilde{p}_{jt} + \tau l_j & \forall j \in \Omega_s \\ \tilde{p}_{jt} & \forall j \in \Omega_d \bigcup \Omega_n \end{cases}$$

where l_j is the volume of product j.

In the counterfactual equilibrium, prices satisfy the conditions:

$$q_{jt}(\mathbf{p}_t) + \sum_{k \in F_f} (\tilde{p}_{kt} - c_{kt}) \frac{\partial q_{kt}(\mathbf{p}_t)}{\partial p_{jt}} = 0 \quad \forall j \in F_f. \tag{2.6}$$

for all firms. We solve for the new equilibrium prices as the vector that satisfies the set of first order conditions (equation 2.6) when $\tau = 0.25$. Tax pass-through describes how much of the tax is shifted through to post-tax prices, for products $j \in \Omega_s$, we measure this as the difference in the post-tax and pre-tax equilibrium consumer price over the amount of tax levied, τl_j.[11]

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

3.1 Data

A substantial portion of soda is consumed on-the-go; in the UK, half of soda is consumed outside the home, the same is true in the US (Han and Powell (2013)). These purchases are for immediate consumption, in contrast to purchases brought into the home, which are for future consumption. Despite the importance of this market segment there are few studies that model choice behavior on-the-go, 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 (not including restaurant or bars), 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 (UPC) level using mobile phones. The data contain product and store information, transaction level prices and the age and gender of each participant. A significant advantage of these data is that we can model individual choice behavior (rather than household level) and that we are able to include young adults (aged 13-21). 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.

In order to measure total dietary sugar and income we use information from the Kantar Worldpanel, which 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. These include all food, drink, alcohol, toiletries, cleaning produce and pet foods. 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

[11] In principle we could solve for a separate price equilibrium in each time period and retailer market (246 markets). Instead we average all markets to the year level and solve the supply side model and conduct the counterfactuals for these representative average markets. We then simulate the counterfactual demands on all markets using the equilibrium price changes obtained for each product on these representative average markets.

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), Kaplan and Menzio (2015) and Dubois et al. (2014)). Data on food purchased on-the-go have, to our knowledge, been much less exploited.

We have information on 4,993 individuals over the period June 2009-October 2012. We observe each person making on average 81 purchases on different days with a minimum of 25. We model demand for the soda products that belong to the main brands, as well as for alternative drink products, see Table 3.5. We observe 2,183 individuals purchasing one of these sodas on at least three occasions – together these individuals account for 99% of all main brand soda purchases. We use this sub-sample of individuals (which we refer to as the set of soda purchasers) to estimate soda demand. When we describe the effect of soda taxes we use the full sample of 4,993 individuals. We gross up so they correspond to annual effects reflective of all on-the-go soda purchases.

A principal motivation policymakers have provided for introducing soda taxes is to lower the sugar consumption of young people, while public health advocates of such taxes have also highlighted the importance of lowering the consumption of sugar among those with high total dietary sugar. We describe how the effects of a tax vary both across the distribution of age and a measure of total dietary sugar. We measure this latter variable as the share of total household calories from added sugar based on all the annual (food and drink) grocery purchases taken into the home of the households to which the individuals in our sample belong.

One potential concern about the adoption of a soda tax is that it might be regressive, with the burden falling disproportionately on those with low incomes. We therefore also describe the effects of the tax across the distribution of total annual equivalized household grocery expenditure; we equivalize using the standard OECD modified equivalence scale (see Hagenaars et al. (1994)). In Appendix A.3 we show that equivalized grocery expenditure is strongly correlated with current income, while expenditure is often viewed as a better proxy for lifetime income than current income is (e.g. Poterba (1989)). Tables 3.1, 3.2 and 3.3 describe the distribution of age, total dietary sugar and equivalized expenditure in our sample.

Young adults comprise 9% of our sample, see Table 3.1. Individuals up to the age of 40 are more likely to be soda purchasers than older individuals – for instance, 68% of those aged 13-21 are soda purchasers, while only 38% of those over 60 are. Conditional on being a soda purchaser, the amount of sugar obtained from soda is negatively related to age; soda purchasing individuals in the youngest age group

(13-21) get, on average, 1754g of sugar from on-the-go soda, purchased in store and vending machines, annually; those in the oldest group (60+) get 886g.

Table 3.1: *Descriptive statistics by age groups*

	Age group					
	13-21	22-30	31-40	41-50	51-60	60+
% of sample	9	14	22	23	18	14
Soda purchasers	0.68	0.67	0.66	0.59	0.49	0.38
Sugar from soda (g)	1754	1439	1181	1235	1054	886

Notes: Row 1 shows fraction of individual-year observations in each age group. Row 2 shows the fraction of each age group that is ever observed purchasing soda. Row 3 is total annual sugar from soda on-the-go from stores and vending machines for those individuals who are soda purchasers.

Table 3.2 shows that the relationship between soda purchasers on-the-go and total dietary sugar is also strong; those with high total dietary sugar overall are both more likely to be soda purchasers (67% of those in the top added sugar decile compared with 46% of those in the bottom one) and, conditional on being soda purchasers, get more sugar from soda than those with more moderate amounts of sugar in their overall diet (1832g for the top decile and 797g for the bottom). The relationship between both age and total dietary sugar with being a soda purchaser and, conditional on this, quantity of sugar from soda are what drive the relationships shown in Figure 1(a) and (b) in the Introduction. These patterns suggest that a tax on the sugar in soda may be well targeted at the sugar consumption of both the young and those with high total dietary sugar.

Table 3.2: *Descriptive statistics by total dietary sugar*

	Decile of distribution of share of calories from added sugar									
	1	2	3	4	5	6	7	8	9	10
Decile upper bound	7.44	8.63	9.57	10.41	11.25	12.11	13.08	14.29	16.3	22.28
Soda purchasers	.46	.53	.52	.56	.58	.59	.61	.63	.63	.67
Sugar from soda (g)	797	962	951	892	1037	1053	1175	1095	1298	1832

Notes: Row 1 gives the upper bound of the decile. Row 2 shows the fraction of each added sugar group that is ever observed purchasing soda. Row 3 is total annual sugar from soda on-the-go from stores and vending machines for those individuals who are soda purchasers.

Table 3.3 shows that individuals in lower deciles of the equivalized grocery expenditure distribution are both more likely to be soda purchasers and, conditional on being so, get more sugar from soda. This raises a possible concern that the economic burden of the soda tax will fall disproportionately on low income individuals.

Table 3.3: *Descriptive statistics by equivalized expenditure*

	Decile of distribution of total equivalized grocery expenditure									
	1	2	3	4	5	6	7	8	9	10
Decile upper bound	.87	1.14	1.37	1.58	1.78	1.99	2.23	2.55	3.05	4.94
Soda purchasers	.63	.6	.61	.57	.6	.58	.55	.58	.56	.5
Sugar from soda (g)	1213	1270	1247	1240	1264	1087	1059	908	927	1056

Notes: Row 1 gives the upper bound of the decile, measured in £1000. Row 2 shows the fraction of each group that is ever observed purchasing soda. Row 3 is total annual sugar from soda on-the-go from stores and vending machines for those individuals who are soda purchasers.

In Section 2.1 we outlined how we can use the long T dimension of our data to identify consumers that never purchase specific subsets of products. In particular, we distinguish between those that only buy soda and that sometimes buy soda and sometimes an alternative drink, and we distinguish between those that buy only diet soda, only sugary soda, and sometimes buy both. Table 3.4 shows which fraction of soda purchasers are in each group. Consumers who are observed buying both soda and non soda, and both diet and sugary soda, over all their choice occasions make up 62.9% of the sample.

Table 3.4: *Consumer specific product sets*

	Purchase:	
	Soda and non soda	Only soda
Only diet	1.7%	1.9%
Both diet and sugary	62.9%	13.5%
Buy only sugary	14.3%	5.6%

Notes: For each soda purchaser, across all their choice occasions, we distinguish between those that only buy soda or buy soda and non soda drinks and that only buy diet, sugary or both diet and sugary soda. Numbers are % of soda purchasers in each cell.

The UK soda market includes a number of large brands and a much larger set of small brands. For the purposes of having a tractable demand model we focus on choice among the large brands. Together these brands make up over two-thirds of the market; omitted brands have share of the drinks market below 3%. We model choice between the major soda products, fruit juice, flavored milk and bottled water (the outside option).

Table 3.5 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. Most brands are available in both a sugary (i.e. regular) and diet variety, and often in two different container sizes. The fruit juice and flavored 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 3.5: *Drinks products*

	Product						
	Firm	Brand	Variety	Size	Market share	Price (£)	g sugar per 100ml
Sodas							
	Coca Cola Company				**51.1%**		
		Coca Cola			*36.1%*		
			Regular	330ml can	7.4%	0.62	10.6
			Regular	500ml bottle	8.8%	1.08	10.6
			Diet	330ml can	8.4%	0.63	0.0
			Diet	500ml bottle	11.5%	1.09	0.0
		Fanta			*5.7%*		
			Regular	330ml can	1.1%	0.60	6.9
			Regular	500ml bottle	4.1%	1.08	6.9
			Diet	500ml bottle	0.5%	1.07	0.6
		Cherry Coke			*3.5%*		
			Regular	330ml can	0.9%	0.63	11.2
			Regular	500ml bottle	2.0%	1.08	11.2
			Diet	500ml bottle	0.7%	1.08	0.0
		Oasis			*5.9%*		
			Regular	500ml bottle	5.4%	1.07	4.1
			Diet	500ml bottle	0.5%	1.06	0.5
	Pepsico				**17.1%**		
		Pepsi			*17.1%*		
			Regular	330ml can	1.8%	0.61	11.0
			Regular	500ml bottle	4.0%	0.96	11.0
			Diet	330ml can	2.3%	0.62	0.0
			Diet	500ml bottle	9.0%	0.95	0.0
	GSK				**11.8%**		
		Lucozade			*7.6%*		
			Regular	380ml bottle	4.4%	0.93	13.8
			Regular	500ml bottle	3.1%	1.13	13.8
		Ribena			*4.3%*		
			Regular	288ml carton	1.2%	0.65	10.5
			Regular	500ml bottle	2.2%	1.12	10.5
			Diet	500ml bottle	0.9%	1.10	0.5
Non-sodas		Fruit juice		330ml	3.6%	1.39	10.6
		Flavoured milk		500ml	2.3%	0.96	10.6
Outside option		Water		14.0%			

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

3.2 Demand estimates

3.2.1 Preference distribution and elasticities

In Table 3.6 we summarize the parameter estimates obtained by maximizing the likelihood function (equation 2.3). The top panel summarizes 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 normalize 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. The reported brand effects are for the first period in the data (June 2010). We allow each of them to vary through time (from month-to-month).[12]

In Figure 3.1 we plot the marginal preference distributions for price, and the soda and sugar product attributes. These are based on individual level preference estimates, so we have a measure of statistical significance for each individual; this is represented by the shading, which indicates consumers with negative, positive and indifferent (i.e. not statistically significantly different from zero) preferences for each attribute. Table 3.6 shows that moments of each of these distributions are estimated with a high degree of statistical significance. Figure 3.1 makes clear that the univariate preference distributions depart significantly from normality (which is typically imposed in random coefficient models) – this is apparent both in the negative and positive skew in the price and soda preference distributions, and also in the infinite portions of the soda and sugar preference distributions.

The estimates of the consumer specific preference parameters (on price, sugar and soda) reveal a large degree of heterogeneity in preferences across individuals – the standard deviation for price preferences is 3.0 (with a coefficient of variation of 1.1), while the standard deviation for sugar and soda is 1.8 and 2.4. The preferences also exhibit interesting correlations – price sensitive consumers tend to have relatively strong sugar preferences (the correlation coefficient between price and sugar preferences is -0.1), as well as relatively strong preferences for the soda product attribute (the correlation coefficient between price and soda preferences is -0.7). We show contour plots of the bivariate preference distributions in Appendix B.1.

[12] We do not report the time varying brand effects or the retailer effects in Table 3.6. These are available upon request.

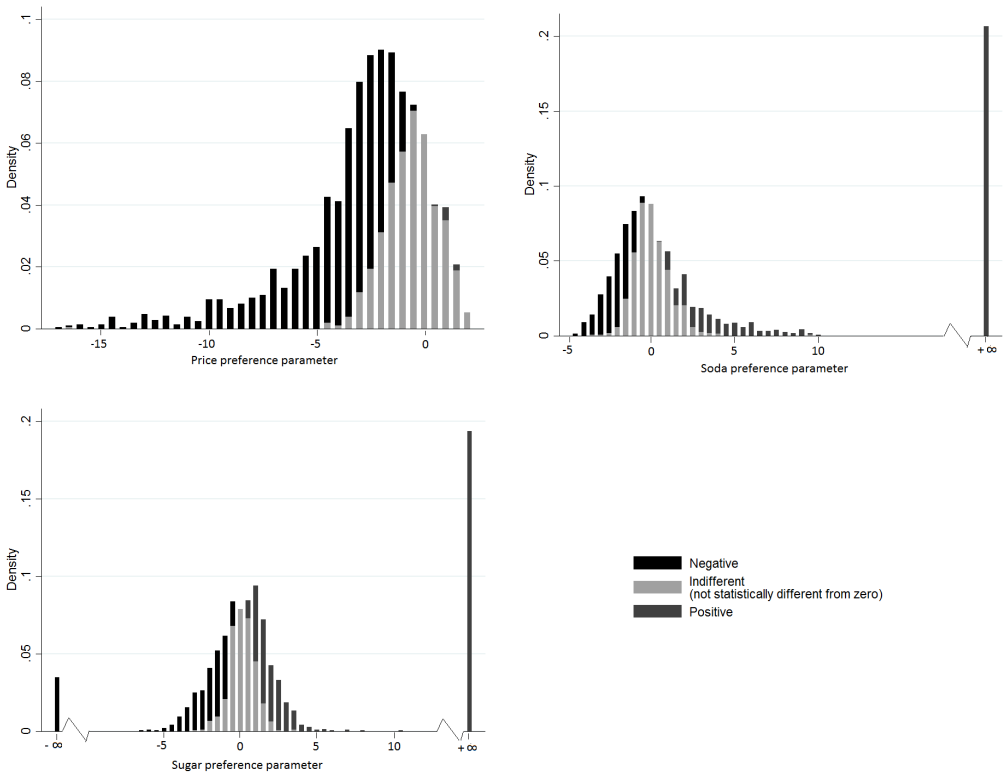
Table 3.6: *Demand model estimates*

Moments of distribution of consumer specific preferences			
Variable		Estimate	Standard error
Price (α_i)	Mean	-2.8349	0.0728
	Standard deviation	3.0401	0.0480
	Skewness	-1.4532	0.1051
	Kurtosis	5.8163	0.6329
Soda (γ_i)	Mean	0.1490	0.0965
	Standard deviation	2.3738	0.0387
	Skewness	1.2065	0.0815
	Kurtosis	5.0141	0.3733
Sugar (β_i)	Mean	0.0550	0.0164
	Standard deviation	1.8340	0.0194
	Skewness	-0.0014	0.0606
	Kurtosis	3.9429	0.2341
Price-Soda	Covariance	-4.3691	0.2030
Price-Sugar	Covariance	-0.4597	0.0711
Soda-Sugar	Covariance	-0.5170	0.0628

Consumer group specific preferences				
Variable	Estimate	Standard error	Estimate	Standard error
	Female - <40		*Female - 40+*	
Carton-size demographic effects ($\delta_{d(i)}$)				
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
Demographic-brand baseline effects ($\xi_{d(i)b(j)\bar{t}}$)				
Fanta	-1.8766	0.1614	-1.6256	0.1550
Cherry Coke	-1.6554	0.1483	-2.3570	0.1971
Oasis	-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
Ribena	-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+*	
Carton-size demographic effects ($\delta_{d(i)}$)				
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
Demographic-brand baseline effects ($\xi_{d(i)b(j)\bar{t}}$)				
Fanta	-1.6338	0.1287	-1.2785	0.1191
Cherry Coke	-2.1061	0.1611	-2.1001	0.1880
Oasis	-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
Ribena	-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 ($\xi_{d(i)b(j)t}$)		Yes		
Retailer-demographic-brand effects ($\zeta_{d(i)b(j)r}$)		Yes		

Notes: We estimate demand on a sample of 2,183 consumers and 150,426 choice occasions. Estimates are summarized 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. Brand effects are shown for a baseline period $\bar{t}$.

Figure 3.1: *Univariate distributions of consumer specific preference parameters*



We report a selection of price elasticities in Table 3.7. 95% confidence bands are given in brackets.[13] The top panel of the table reports elasticities for products that belong to the two most popular brands, Coca Cola and Pepsi.[14] 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 drinks (soda, fruit juice plus flavored milk). For example, a 1% increase in the price of the most popular sugary product, Coca Cola 500ml, would result in a reduction in demand for that product of around 1.75%. Demand for alternative sugary sodas would rise by around 0.37%, demand for diet sodas would rise by 0.12% and demand for non soda sugary drinks would rise by 0.18%. Demand for soda and juice as a whole would fall by 0.07%.

A couple of interesting patterns are apparent, First, consumers are more willing to switch from sugary soda products to alternative sugary sodas and from diet products to diet alternatives, than they are between sugary and diet products.

[13]To calculate the confidence intervals we obtain the variance-covariance matrix for the parameter vector estimates using standard asymptotic results. We then take 100 draws of the parameter vector from the joint normal asymptotic distribution of the parameters and, for each draw, compute the statistic of interest, using the resulting distribution across draws to compute Monte Carlo confidence intervals (which we recentre).

[14]We show elasticities for all products in Appendix B.2.

Second, the price elasticities for the 500ml products are smaller in magnitude than for the 330ml versions; consumers that choose to buy the larger bottle variants rather than smaller cans, tend to be less willing to switch away from their chosen product in response to a price increase. This is precisely the opposite pattern from what we would get in a logit choice model without preference heterogeneity, in which the functional form imposes that own price elasticities are approximately proportional to price and therefore the higher price bottles would be more price elastic than cans.

Table 3.7: *Price effects*

	Effect of 1% price increase on:				
	Own demand	cross demand for: sugary soda	diet soda	sugary alternatives	Total demand
Coca Cola 330	-2.56	0.25	0.08	0.05	0.01
	[-2.62, -2.51]	[0.25, 0.26]	[0.08, 0.08]	[0.05, 0.06]	[0.01, 0.01]
Coca Cola 500	-1.75	0.37	0.12	0.18	-0.07
	[-1.87, -1.63]	[0.35, 0.40]	[0.11, 0.13]	[0.17, 0.20]	[-0.07, -0.07]
Coca Cola Diet 330	-2.43	0.08	0.29	0.01	0.01
	[-2.49, -2.36]	[0.07, 0.08]	[0.29, 0.30]	[0.01, 0.01]	[0.01, 0.01]
Coca Cola Diet 500	-1.47	0.11	0.37	0.06	-0.05
	[-1.56, -1.37]	[0.10, 0.12]	[0.34, 0.39]	[0.05, 0.07]	[-0.05, -0.05]
Pepsi 330	-3.12	0.11	0.03	0.02	0.00
	[-3.20, -3.05]	[0.11, 0.12]	[0.03, 0.03]	[0.02, 0.02]	[0.00, 0.01]
Pepsi 500	-2.13	0.20	0.07	0.09	-0.04
	[-2.24, -2.03]	[0.19, 0.21]	[0.06, 0.07]	[0.08, 0.10]	[-0.05, -0.04]
Pepsi Diet 330	-3.43	0.03	0.18	0.00	0.00
	[-3.52, -3.34]	[0.03, 0.03]	[0.17, 0.18]	[0.00, 0.00]	[0.00, 0.00]
Pepsi Diet 500	-1.89	0.06	0.23	0.03	-0.04
	[-1.99, -1.80]	[0.06, 0.07]	[0.22, 0.25]	[0.02, 0.03]	[-0.04, -0.04]
Soda	-0.34			0.77	-0.27
	[-0.35, -0.32]			[0.69, 0.84]	[-0.28, -0.26]
Sugary soda	-0.72		0.50	0.63	-0.15
	[-0.76, -0.68]		[0.47, 0.52]	[0.57, 0.69]	[-0.16, -0.15]

Notes: For each of the eight 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. 95% confidence bands are shown in brackets.

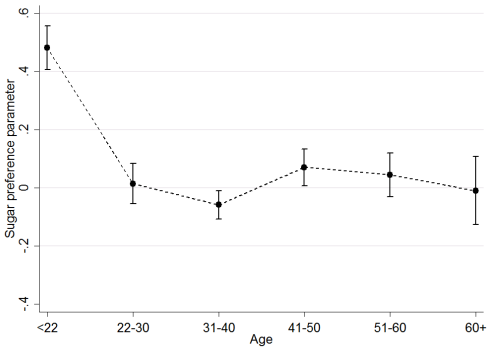
The bottom panel of Table 3.7 reports the effect on demand of a marginal increase in the price of all sugary soda and in the price of all soda (i.e. both sugary and diet). The own price elasticity for soda is -0.34. This is smaller than the own price elasticity of any individual soda product. The own price elasticity for sugary soda is -0.72. This is larger than for all soda, reflecting that some consumers respond to an increase in the price of sugary soda by switching to diet alternatives.

3.2.2 Relationship with individual attributes

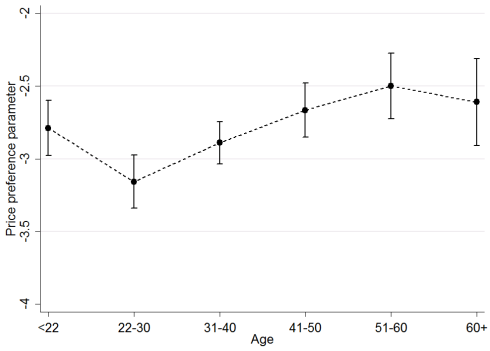
A key feature of our model is that it allows us to flexibly relate preference parameters to characteristics of consumers. This enables us to address the question of how well targeted soda taxes are, and to what extent they disproportionately impact the young and the poor.

Figure 3.2: *Preferences variation with age*

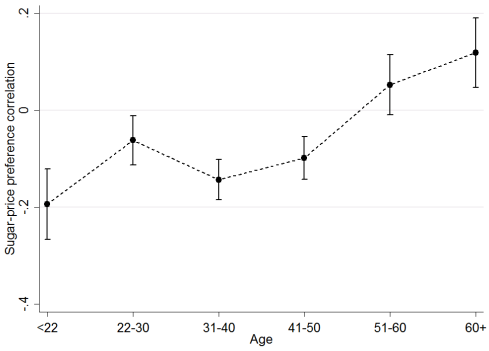
(a) preferences for sugar



(b) preferences for price



(c) sugar-price preference correlation



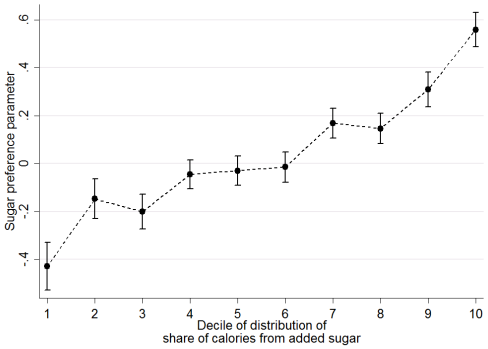
Notes: Figure shows how the mean sugar and price preferences and the correlation between sugar and price preferences vary by age groups. 95% confidence intervals are shown by bars.

In Figure 3.2 we show how features of the preference distribution vary with age. Panel (a) shows that the mean preference for sugar varies with age, with younger individuals (aged 13-21) having stronger sugar preferences than older individuals. Panel (b) shows that younger consumers tend to have price coefficients that are larger in magnitude than those aged over 40, indicating they are more sensitive to changes in prices. Panel (c) shows how the *within* age group correlation in sugar and price preferences varies across age groups. There is an increasing relationship; among individuals below 50 years of age the preferences exhibit a negative correlation within their age group, and this is more negative for the youngest group, and among older groups the correlation is positive. This indicates that within the younger groups (and especially so for the youngest group, aged 13-21), those

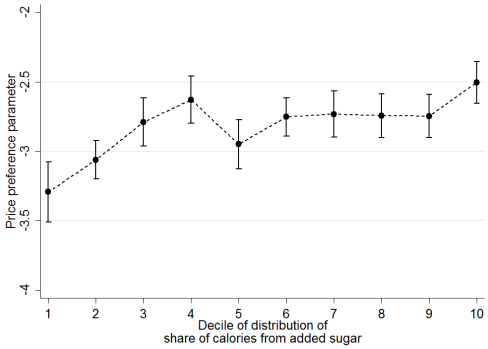
consumers with strong sugar preferences tend also to be the most price sensitive, whereas within older groups those with the strongest sugar preferences are least price sensitive. These preference patterns are important in driving the shape of demand responses across the age distribution to soda taxes.

Figure 3.3: *Preferences variation with total dietary sugar*

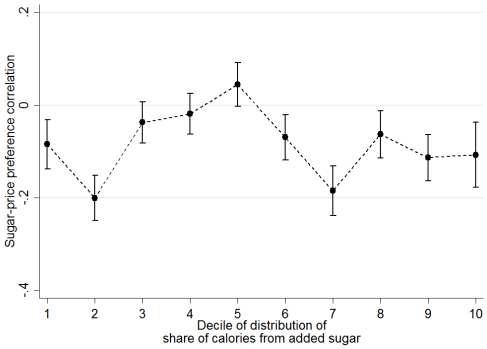
(a) preferences for sugar



(b) preferences for price



(c) sugar-price preference correlation



Notes: Figure shows how the mean sugar and price preferences and the correlation between sugar and price preferences vary by deciles of the distribution of total calories from added sugar. 95% confidence intervals are shown by bars.

Figure 3.3 shows how price and sugar preferences vary across deciles of the distribution of total dietary sugar (measured as the share of total calories brought into the home over a year that are from added sugar), and how the within decile correlation in these preference patterns varies. It shows that individuals with high total dietary sugar (based on their average annual at home grocery basket) tend to have relatively strong preferences for sugar when buying drinks on-the-go, but that their price preferences tend to be smaller in magnitude, suggesting they might respond relatively weakly to price changes. Unlike for age, in the case of total dietary sugar, there is no obvious pattern in how the within group (decile) preference correlations vary across deciles.

In Appendix B.1 we show how preferences vary across deciles of the distribution of total equivalized grocery expenditure (a proxy for income). There is a clear gra-

dient in income for both sugar and price preference parameters; poorer individuals typically have stronger sugar preferences and more negative preferences for price than richer individuals.

3.3 Bias correction for incidental parameters problem

In our non-linear model with fixed effects, maximum likelihood estimates of the parameters may suffer from an incidental parameters problem, noted by Neyman and Scott (1948). 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 $1/T^2$.

A range of bias correction methods exist (see surveys in Arellano and Hahn (2007), Arellano and Bonhomme (2011)). We use panel jackknife methods (Hahn and Newey (2004)), employing the split sample procedure 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 sub-samples. Each sub-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:

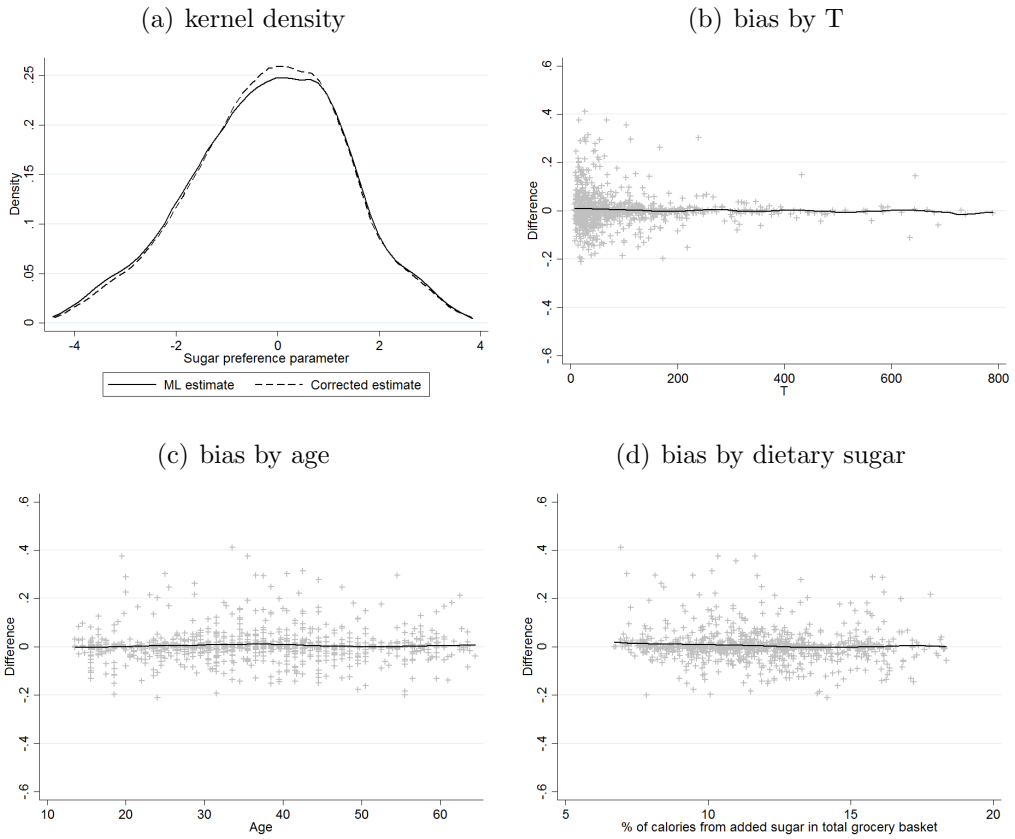
$$\widetilde{\theta}_{split} = 2\widehat{\theta} - \frac{\widehat{\theta}_{(1,T/2)} + \widehat{\theta}_{(T/2,T)}}{2}.$$

In Figure 3.4 we graph the difference between the jackknife (bias corrected) and maximum likelihood sugar preference parameters. Panel (a) shows the distribution of estimates (for those with finite sugar preferences) for the maximum likelihood and jacknife estimates. Panel (b) shows how the difference in these estimates relates to the time a consumer is in the sample. Panels (c) and (d) show how the difference relates to consumers' age and total dietary sugar.

The figure shows that the difference between the two estimates is 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 0.06. The difference is decreasing in T; those in the sample for a relatively small number of choice occasions tend to have higher differences than those in the sample relatively many times. However, 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

and the difference between the jackknife and maximum likelihood estimates is completely unrelated to individuals' age or total dietary sugar.

Figure 3.4: *Sugar preference parameters*



Notes: In panels (b)-(d) markers represent consumer level differences. Lines are local polynomial regressions.

In Appendix B.3 we show that 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 or negative and there is no systematic relationship with the key variables that we relate our demand effects to. For instance, the standard deviation of the price preference parameter estimates is around 3.0, while the average absolute difference between the jackknife and maximum likelihood estimates is 0.2. For the soda preferences the numbers are 2.4 and 0.1. As a consequence, our results regarding the effectiveness of soda taxes are completely robust to the bias correction procedure.

4 The effects of a soda tax

4.1 Market equilibrium

We use our demand estimates, along with the supply side model outlined in Section 2.3, to simulate the introduction of a tax levied on sugary soda.[15] We consider the introduction of a tax of 25 pence per liter. This tax is similar to what has been implemented in some cities and counties in the US and also implies equilibrium price changes that are of a roughly similar order to the observed price changes in our data period. By construction, for soda brands with sugar, such a tax will be larger for larger sized products, imposing more tax on a 500ml bottle than on a 330ml can. For simplicity, in solving for the post-tax equilibrium we hold fixed the prices of the non soda composite products, fruit juice and flavored milk, as well as the outside option. We model the pricing response of soda manufacturers, including changes in prices for soda products not directly subject to the tax (i.e. diet sodas).

In Table 4.1 we report the mean tax levied per product, 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 10.55 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 15 pence – average equilibrium pass-through of the tax is therefore around 140%. Important in driving this over shifting of the tax is the strong strategic complementaries between the prices of soda products owned by different firms. For instance, if we impose a sugary soda tax only on the products owned by the largest firm in the market, Coca Cola, the average pass-through of the tax onto its products is less than 100%.

Pass-through rates vary across products; the larger 500ml bottled products typically have rates in excess of 150% and smaller 330ml canned products have rates of around 100%. This means manufacturers respond to the tax by increasing margins on the 500ml products and maintaining them at around the pre-tax level for the 330ml cans. Our demand estimates imply that the bottled products have less elastic demands than the cans. By raising margins on these products, firms sacrifice some marginal consumers, who switch to alternatives, but earn more profits on the infra-marginal consumers who still buy bottles. Nevertheless, profits on the bottled products fall, while profits on the canned products, in some cases, rise as some consumers respond to the tax by downsizing (i.e. switching from bottles to cans).

[15] In Appendix C we show results for a tax levied on all soda.

Table 4.1: *Effects of sugary soda tax at product level*

	Tax (pence)	Δ price (pence)	Δ share (p.p.)
Sugary soda	10.55	14.98	-5.50
		[14.05, 15.90]	[-5.48, -4.96]
Coca Cola 330	8.25	8.36	0.00
Coca Cola 500	12.50	20.42	-1.98
Fanta 330	8.25	8.39	-0.02
Fanta 500	12.50	20.89	-0.37
Cherry Coke 330	8.25	8.29	0.00
Cherry Coke 500	12.50	19.34	-0.22
Oasis 500	12.50	20.93	-0.56
Pepsi 330	8.25	8.19	0.00
Pepsi 500	12.50	20.11	-1.25
Lucozade 380	9.50	13.24	-0.45
Lucozade 500	12.50	19.08	-0.46
Ribena 288	7.20	6.94	0.09
Ribena 500	12.50	20.52	-0.28
Diet soda	0.00	-3.12	3.41
		[-3.66, -2.58]	[2.92, 3.30]
Coca Cola Diet 330	0.00	0.55	0.19
Coca Cola Diet 500	0.00	-4.59	1.37
Fanta Diet 500	0.00	-5.37	0.30
Cherry Coke Diet 500	0.00	-4.52	0.16
Pepsi Diet 330	0.00	0.04	0.17
Pepsi Diet 500	0.00	-2.89	0.68
Ribena Diet 500	0.00	-2.97	0.13
Sugary alternatives	0.00	0.00	0.59
			[0.53, 0.62]
Outside option	0.00	0.00	1.51
			[1.45, 1.59]

Notes: Panels 2 and 4 show the mean effect of the sugary soda tax on products' price and market share. Panels 1, 3 and 4 show the mean effects of the tax on all sugary soda, all diets soda and on alternative drinks. 95% confidence intervals are shown by bars.

The tax on sugary sodas thus increases equilibrium prices for sugary sodas, doing so by more for the larger sized products due to a higher tax rate and over shifting. The market share of sugary sodas falls by 5.5 percentage points. Soda manufacturers also optimally respond to the tax by lowering the price of diet products. The average reduction in price is 3 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. Most of the demand reduction in sugary sodas represents substitution to diet sodas.

The pricing response of soda manufacturers acts to magnify the price differential that the tax creates between sugary and diet products. Relative to the case in which

producers simply increase consumer prices by an amount exactly equal to the tax (so pass-through of tax is 100%), firms' equilibrium pricing response induces more switching away from sugary soda and more towards diet soda; the share of diet soda increases by 3.4 percentage points. Alternative sugary drinks and the outside option also see increases in market share of 0.6 and 1.51 percentage points.

A number of papers use observed tax changes to estimate pass-through of taxes to prices. These include Besley and Rosen (1999), which exploits variation in state and local sales taxes in the US and looks at the impact on prices of a number of products and finds over shifting for soda products, Delipalla and O'Donnell (2001), which analyzes the incidence of cigarette taxes in several European countries and Kenkel (2005), which uses data on how the price of alcoholic beverages changed in Alaska. Results from the literature vary, but typically these papers find complete or over shifting of specific taxes, which broadly accord with our pass-through results.

Evidence from papers that study recently implemented taxes imposed on soda is mixed; comparing taxed and non-tax products, Grogger (2015) finds that prices rose by more than the amount of the tax following the adoption of the Mexican soda tax in 2014, while Cawley and Frisvold (2017) find under-shifting of the Berkeley soda tax, which they rationalize as due to the ease with which consumers can avoid the tax by shopping in neighbouring municipalities. In an ex ante study of the effects of a sugary soda tax in France, Bonnet and Réquillart (2013) find pass-through that exceeds 100% and also reductions in the prices of diet products. The empirical literature on pass-through of cigarette taxes is similarly mixed; Harding et al. (2012) find taxes in the US are under-shifted and that avoidance opportunities have a sizeable effect on purchases, while Lillard and Sfekas (2013) find evidence of over shifting once the implicit tax in state lawsuits is taken account off.

There is also a related literature that looks at pass-through of cost shocks. Much of this finds under-shifting (see, for instance, Goldberg and Hellerstein (2013) and Nakamura and Zerom (2010)). An important reason for incomplete pass-through of cost shocks is that often not all cost components are affected by the shock. For instance, exchange rate movements do not directly impact the cost of non-traded inputs (Goldberg and Hellerstein (2013)). In a context where firms' marginal costs are observable (in the wholesale electricity market), Fabra and Reguant (2014) find changes in marginal costs are close to fully shifted to prices.

4.2 How well targeted is the tax?

Our tax simulation suggests that, on average, consumers will lower the total amount of sugar they purchase from drinks on-the-go by around 160g per annum. However,

there is a substantial amount of variation in this – 43% of consumers do not purchase soda and therefore the tax induces zero reduction in sugar for this group. On the other hand, within the 57% of consumers that do purchase soda the reductions in sugar are dispersed; the 25th, 50th and 75th percentiles of the distribution of reduction in sugar purchased are 17g, 103g and 325g – corresponding to the equivalent of around 0.5, 3 and 9 cans of Coca Cola per annum.

Key to understanding the effectiveness of a soda tax is whether it successfully achieves reductions in sugar amongst the targeted groups of consumers – the young and those with high total dietary sugar. In Figure 4.1 we show how reductions in sugar achieved by the soda tax (in grams per year) vary with age and total dietary sugar. This figure is based on the full sample of individuals. Variation in sugar reductions across consumer age and total dietary sugar reflect differential propensities of different groups to consume soda (i.e. be in the sample of soda purchasers), which we measure in our longitudinal data (see Tables 3.1 and 3.2), and how strongly the soda purchasers switch from sugary soda and to what alternatives they switch, which we estimate based on our equilibrium demand and supply model. We estimate individual specific preferences, so we allow the strength of switching to vary in a flexible way across the distribution of age and total dietary sugar.

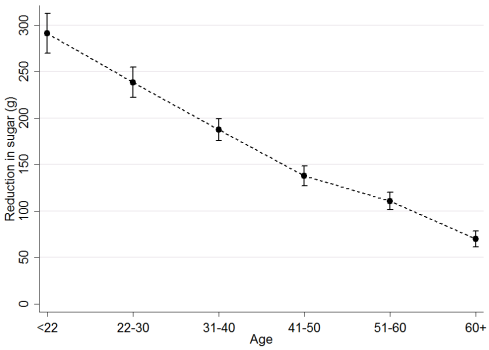
Panel (a) shows that the reduction in sugar falls monotonically across age groups. The mean reduction among the group aged 13-21 is 291g; this falls to a mean reduction of 70g for the oldest group, aged above 60. This relationship is driven by two things. First, young people are more likely to be soda purchasers, and therefore are more likely to be affected by the tax. This does not explain the entire relationship. The second reason is that among the soda purchasers the young respond more strongly to the tax; conditional on being soda purchasers the average sugar reduction among the youngest groups is around 2.3 times of that for the oldest group. Important in driving this is the way the joint distribution of preferences varies across the age groups, and in particular that among younger groups those individuals with strong sugar preferences tend to be relatively price sensitive.

While the tax does succeed in achieving relatively large responses among the young, it is much less successful at achieving large demand responses among those with high total dietary sugar (see panel (b)). Across the deciles of the distribution of total dietary sugar, the smallest average reduction is 131g (decile 4) and the largest is 197g (decile 10). The higher sugar reductions among the top few deciles is driven entirely by a larger fraction of those deciles being comprised of soda purchasers relative to the lower deciles. Conditional on being a soda purchasers, individuals with high total dietary sugar do not reduce purchases of sugar in soda on-the-go by

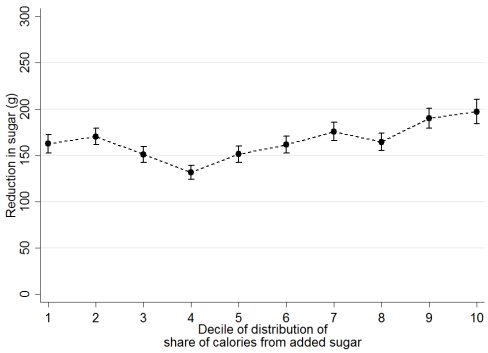
more as a consequence of the tax compared with those with more moderate amounts of total dietary sugar. The reason is that individuals with high total dietary sugar both tend to have relatively strong sugar preferences and to be relatively insensitive to price (when making on-the-go drinks decision).

Figure 4.1: *Reductions in sugar*

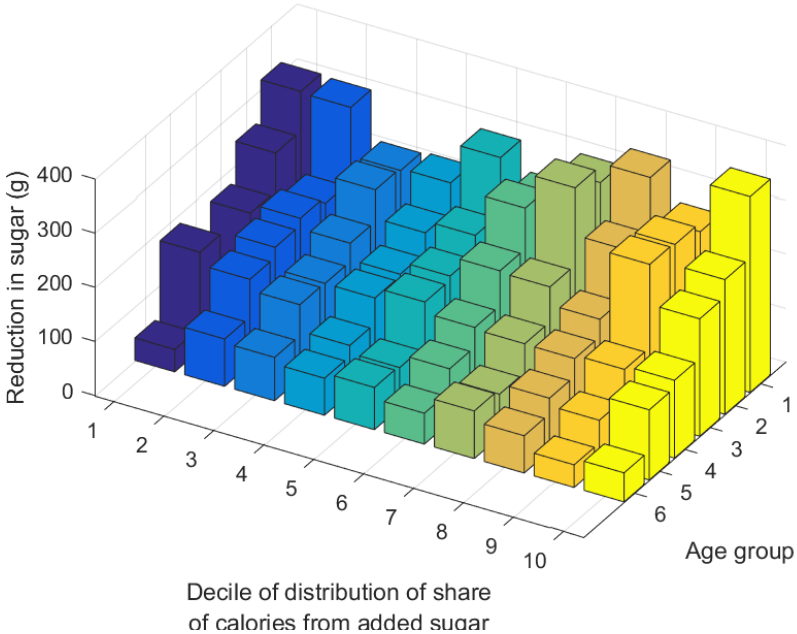
(a) by age



(b) by total dietary sugar



(c) by age and total dietary sugar



Notes: Sample includes soda purchasers and non soda purchasers. Numbers show how the mean reduction in sugar from soda varies by age and deciles of the distribution of share of calories from added sugar. In panels (a) and (b) 95% confidence intervals are shown by bars. In panel (c) age groups are 1=<22, 2=22-30, 3=31-40, 4=41-50, 5=51-60, 6=60+.

Panel (c) shows how the reduction in sugar varies jointly with age and total dietary sugar. It shows that, within each decile of the distribution of total dietary sugar, the strong relationship between sugar reductions and age holds, with the young responding more strongly.

4.3 Consumer welfare

Higher taxes, to the extent they raise prices, impose an economic burden on consumers; after a tax is introduced consumers can obtain less produce for a given amount of expenditure than before. In the case of a tax on sugary soda, consumers that buy sugary soda will incur a welfare loss through this channel. Those consumers that never buy soda will see no change in their welfare (we assume that the prices of non sodas are unaffected by the tax), while those individuals that drink diet soda may actually benefit slightly as the optimal pricing response to the tax is to lower the price of diet sodas.

In Figure 4.2 we describe this effect; we use the preference estimates to compute compensating variation – the monetary amount an individual would require to be paid to be indifferent to the imposition of the tax based on their estimated preferences. Letting p_{jrt} and p'_{jrt} denote the region r time t price of product j prior to and following the introduction of the tax, the expected compensating variation for individual i on a choice occasion is given by (Small and Rosen (1981)):

$$cv_{it} = \frac{1}{\alpha_i}\left[\ln\left(1_{\{\gamma_i\in(-\infty,\infty)\}}\exp(\eta_{i0rt}) + \sum_{k\in\Omega_i}\exp(v_{ikt}+\eta_{ikrt}-\alpha_i(p_{jrt}-p'_{jrt}))\right) - \ln\left(1_{\{\gamma_i\in(-\infty,\infty)\}}\exp(\eta_{i0rt}) + \sum_{k\in\Omega_i}\exp(v_{ikt}+\eta_{ikrt})\right)\right]$$

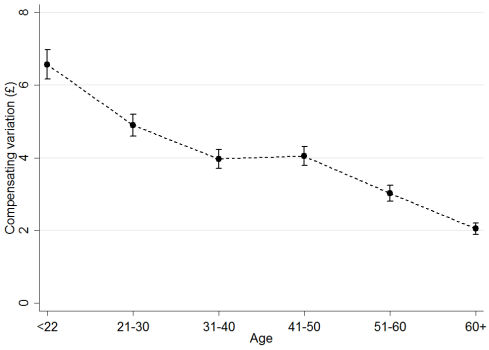
where v_{ijt}, η_{i0rt}, and η_{ijrt} are defined in Section 2.1. Summing cv_{it} over all the consumer's choice occasions in the year gives their annual compensating variation. We show how this varies by an individual's age (panel (a) of Figure 4.2), position in the distribution of total dietary sugar (panel (b)), and jointly with these variables (panel (c) of Figure 4.2). Compensating variation is falling across age groups and rising across total dietary sugar deciles; on average, both the youngest and those with high total dietary sugar have the highest compensating variations. This is because both groups are more likely to be soda purchasers and, conditional on being a soda purchaser, are more likely to purchase large quantities. Panel (c) of Figure 4.2 shows that the highest compensating variations are among those in the top total dietary sugar decile and aged below 60, and those aged below 30 and in the top half of the total dietary sugar distribution.

If consumers fully took account of all the costs associated with their soda consumption, then compensating variation would capture the total effects of the tax on consumer welfare and we could conclude that the tax makes all purchasers of sugary soda worse off, with the largest effects being among the young and those with

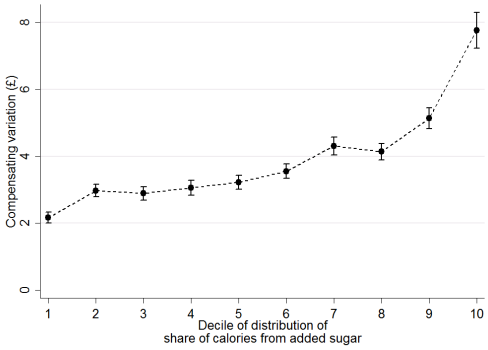
high levels of dietary sugar. However, if soda consumption is associated with future costs that are not taken account of by the individual at the point of consumption (internalities), or with costs imposed on others (externalities), then compensating variation measured based on revealed preference captures only part of the total consumer welfare effect of the tax.

Figure 4.2: *Revealed consumer welfare effect*

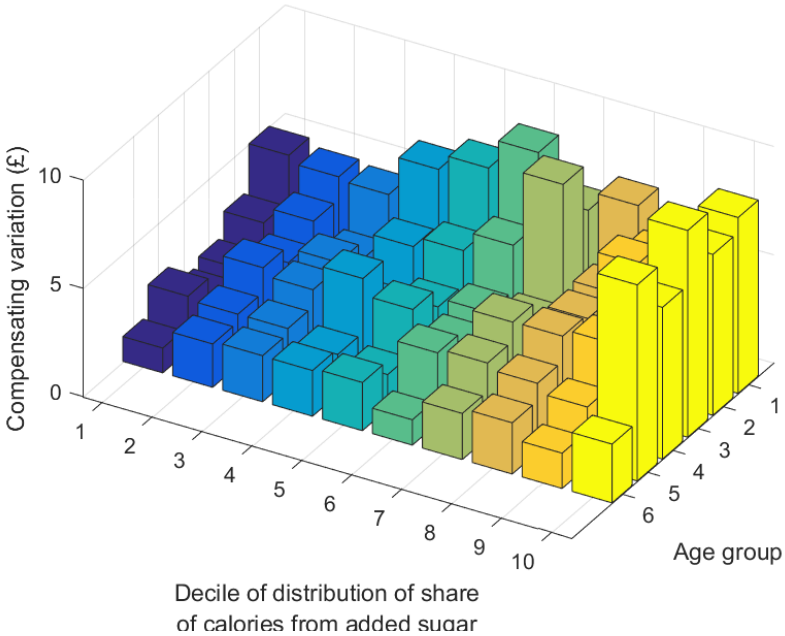
(a) by age



(b) by total dietary sugar



(c) by age and total dietary sugar



Notes: Sample includes soda purchasers and non soda purchasers. Numbers show how the mean compensating variation varies by age and deciles of the distribution of share of calories from added sugar. In panels (a) and (b) 95% confidence intervals are shown by bars. In panel (c) age groups are 1=<22, 2=22-30, 3=31-40, 4=41-50, 5=51-60, 6=60+.

The potential consequences of consuming sugar in excess are well established. It may be that some individuals fully internalize these future costs when deciding whether to consume sugary soda. However, there is a large theory literature that posits that not all individuals fully account for future costs of consumption (for a

survey see Rabin (1998)) and there is evidence this is particularly relevant for food, both experimental (for instance Read and Van Leeuwen (1998) and Gilbert et al. (2002)) and circumstantial, through the existence of a multi-billion pound dieting industry (Cutler et al. (2003)).

The young are particularly susceptible to suffer from internalities from excess sugar. The consequences of poor nutrition early in life are profound: with excess sugar associated with poor mental health and school performance in children, and poor childhood nutrition thought to be an important determinant of later life health, social and economic outcomes and of persistent inequality. (see, for instance Cawley (2010)).[16] Few would argue that these significant costs are fully taken account of by children and young adults when making consumption decisions. The average compensating variation for individuals aged 13-21 and who are soda purchasers is around £10, while the average reduction in sugar for this group is around 430g (equivalent to around 12 cans of Coca Cola).[17] If the internality associated with drinking the amount of sugar in a can of Coca Cola is above £0.80, then, for the average person aged 13-21, the soda tax will be welfare improving.

While there is not much direct evidence that those individuals that have high total dietary sugar suffer from internalities, it is well understood that the health consequences of such a diet are severe and there is evidence that excess consumption can have convexly increasing health costs (for instance Hall et al. (2011) show that adults with greater adiposity experience larger health gains from a given reduction in energy intake). For those in the top decile of the distribution of total dietary sugar and who are soda purchasers, the average compensating variation from the tax is £11.50, and the average reduction in sugar is 290g. For this group, they would need to avoid internalities of around £1.40 per can of Coca Cola avoided (or a drink with the equivalent sugar content) to be made better off by the tax.

4.4 Redistribution

A common criticism of excise taxes is that they are regressive. This is typically based on the observation that those with lower incomes tend to be relatively heavy consumers of the taxed product (which, for a small change in price, is a good

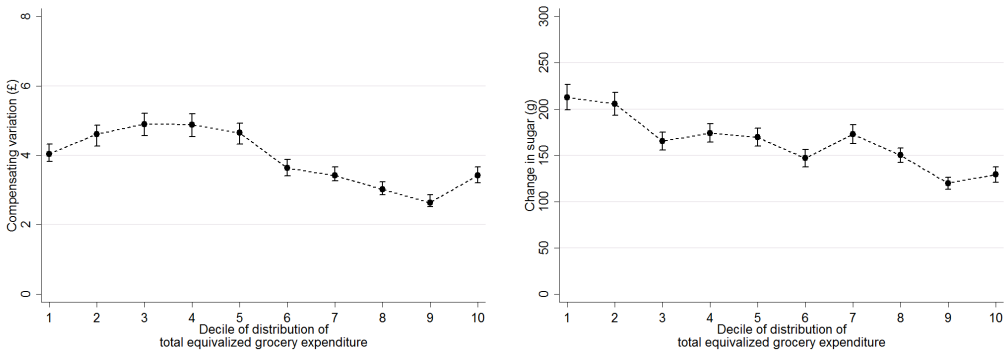
[16] Cavadini et al. (2000) document an increase in soft drink consumption in the US for 11-18 year olds of almost 300% for boys, and over 200% for girls between 1965 and 1996. Nielsen and Popkin (2004) document a contemporaneous fall in the share of calories children get from milk. Medical evidence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

[17] Note, these numbers differ from those in Figures 1(a) and 2(a) because in the figures we show numbers based on all individuals – whether or not they are soda purchasers.

approximation to the revealed consumer welfare cost). Table 3.3 confirms that, in the case of sugary soda, poorer individuals (those with low total annual equivalized grocery expenditure) are more likely to be soda purchasers and to get more sugar from these products. Based on our demand and supply estimates we can estimate the true revealed welfare cost from the tax – Panel (a) of figure 4.3 shows how this varies across deciles of the equivalized grocery expenditure distribution. It is noticeable that the gradient in compensating variation in the figure is less steep than are the purchase patterns in the descriptive table (Table 3.3). The reason is that the relatively strong demand responses of those in the lower expenditure deciles limits the magnitude of their compensating variations.

Figure 4.3: *Effects by total equivalized grocery expenditure*

(a) compensating variation (b) reduction in sugar



Notes: Sample includes soda purchasers and non soda purchasers. Panel (a) shows how compensating variation and panel (b) shows how reductions in sugar, from tax varies across deciles of the distribution of total equivalized grocery expenditure. 95% confidence intervals are shown by bars.

However, if some consumers impose internalities on themselves, then the standard revealed consumer welfare loss (compensating variation) provides an incomplete picture of the redistributive effects of the tax (a point made by Gruber and Koszegi (2004) in the case of cigarette taxation). Panel (b) of Figure 4.3 shows that sugar reductions from the tax are somewhat higher on average among those towards the bottom of the equivalized grocery expenditure distribution compared to those further up (200g for the bottom decile versus 120g for the top). Therefore, if the prevalence of internalities is broadly constant across the expenditure distribution, the larger reductions in sugar among low spending individuals may be enough to offset the slightly higher compensating variation.

In addition, there is some evidence that low income people are more likely to exhibit behavior that creates internalities. For example, Haushofer and Fehr (2014) and Mani et al. (2013) suggest that the stress and cognitive load of being in poverty means people are more likely to make unwise decisions and underweight the future. Focusing on asset accumulation Bernheim et al. (2015) argue that poverty can per-

petuate 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. Any propensity for self-control problems, or other sources of internality generating behavior, that are concentrated among poorer individuals is likely to result in a soda tax being less regressive.

5 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 that contain sugar. 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. It would be numerically difficult to estimate our model with all food on-the-go items being simultaneous choices. Instead we embed our drinks model into a two stage food on-the-go choice model. We assume that the idiosyncratic unobserved shocks that affect the choice of which drink to consume are unknown in the first stage, so that we can simplify the choice model between drinks and non drinks, but still take into account the heterogeneity of tastes and preferences of consumers for drinks. The idiosyncratic i.i.d. extreme value shock in the first stage are taken in expectation in the first stage.

Thus, we suppose the choice model of Section 2 is a second stage of a two-stage decision process, which governs which drink to select, conditional on choosing a drink. Consider a first stage in which the consumer chooses between chocolate products, a non-sugary snack and 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:

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

where $W_{i\mathcal{D}t}$ is the expected utility from choosing the preferred 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 Appendix D 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 six age groups used previously. 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 the preferred drink) does induce some switching away from drinks and towards foods. However, the strength of this switching to food between the baseline model (results presented in Section 4) and the extended two-stage model is relatively small. Taking account of switching to food sources of sugar dampens the mean overall reduction in sugar by between 4% (for those aged 21 or below) to 11% (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. Appendix D provides further details.

6 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 large future costs, while almost all individuals exceed official recommendations 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 total dietary sugar 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 total dietary sugar 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 effected by such policies.

References

Adams, A., L. Cherchye, B. De Rock, and E. Verriest (2014). Consume now or later? Time inconsistency, collective choice, and revealed preference. *American Economic Review 104*(12), 4027–4070.

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

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

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

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

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

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

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.

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.

Cawley, J. (2010). The economics of childhood obesity. *Health Affairs 29*(3), 364–371.

Cawley, J. and D. E. Frisvold (2017, March). The Pass-Through of Taxes on Sugar-Sweetened Beverages to Retail Prices: The Case of Berkeley, California. *Journal of Policy Analysis and Management 36*(2), 303–326.

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? *Journal of Economic Perspectives 17*(3), 93–118.

Delipalla, S. and O. O'Donnell (2001). Estimating tax incidence, market power and market conduct: The European cigarette industry. *International Journal of Industrial Organization 19*(6), 885–908.

Dhaene, G. and K. Jochmans (2015, July). Split-panel Jackknife Estimation of Fixed-effect Models. *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.

Fabra, N. and M. Reguant (2014). Pass-Through of Emissions Costs in Electricity Markets. *American Economic Review 104*(9), 2872–2899.

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.

Goldberg, P. K. and R. Hellerstein (2013). A structural approach to identifying the sources of local currency price stability. *Review of Economic Studies 80*(1), 175–210.

Grogger, J. (2015, May). Soda Taxes and the Prices of Sodas and Other Drinks: Evidence from Mexico. Working Paper 21197, National Bureau of Economic Research.

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.

Harding, M., E. Leibtag, and M. F. Lovenheim (2012). The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data. *American Economic Journal: Economic Policy 4*(4), 169–198.

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

Kaplan, G. and G. Menzio (2015). The morphology of price dispersion. *International Economic Review 56*(4), 1165–1206.

Kenkel, D. S. (2005). Are alcohol tax hikes fully passed through to prices? Evidence from Alaska. *American Economic Review: Papers & Proceedings 95*(2).

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

Lillard, D. R. and A. Sfekas (2013). Just passing through: The effect of the Master Settlement Agreement on estimated cigarette tax price pass-through. *Applied economics letters 20*(4), 353–357.

Lockwood, B. B. and D. Taubinsky (2017, January). Regressive Sin Taxes. Working Paper 23085, National Bureau of Economic Research.

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.

Nakamura, E. and D. Zerom (2010). Accounting for Incomplete Pass-Through. *Review of Economic Studies 77*(3), 1192–1230.

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.

Poterba, J. M. (1989). Lifetime Incidence and the Distributional Burden of Excise Taxes. *American Economic Review 79*(2), 325–330.

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.

Sanders, B. (2016, April). Bernie Sanders Op-Ed: A Soda Tax Would Hurt Philly's Poor.

Small, K. A. and H. S. Rosen (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica 49*(1), 105–130.

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.

Appendix

How well targeted are soda taxes?

Pierre Dubois, Rachel Griffith and Martin O'Connell

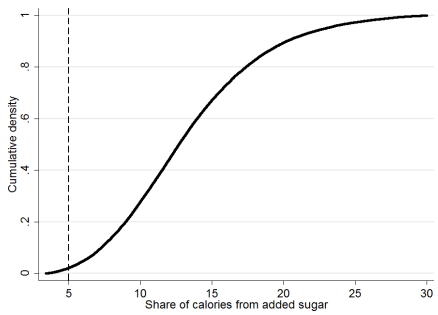
A Data appendix

A.1 Purchase patterns in US

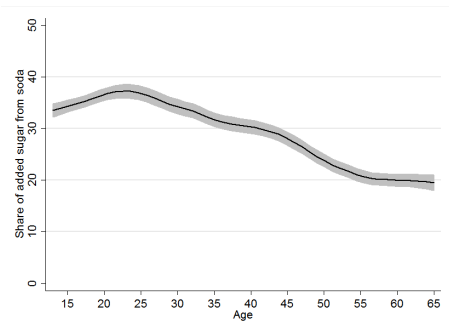
In the paper we show that in the UK the share of added sugar obtained from soda is increasing in individuals' total dietary sugar and decreasing in age. Using the National Health and Nutrition Examination Study over 2007-2014, a sample of 39,189 adults and children, we show similar patterns hold in the US. In Figure A.1 we use these data to show this. Panel (a) shows in the US, like the UK, the majority of the population get more calories from added sugar than the WHO guideline. The remaining panels replicate what we see in the UK (Figure 1.1 of the main paper).

Figure A.1: *Added sugar and soda (US)*

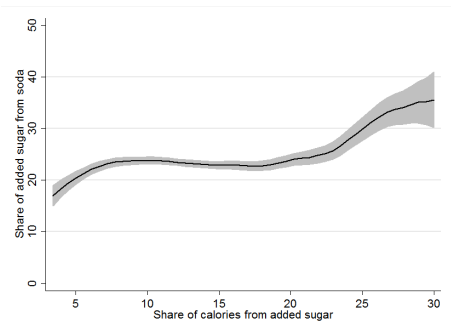
(a) Share of calories from added sugar



(b) Soda and age



(c) Soda and added sugar



Notes: Numbers based on 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.

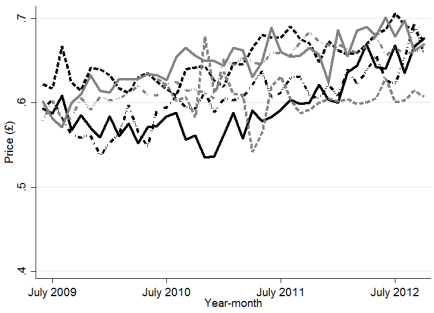
A.2 Prices

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.

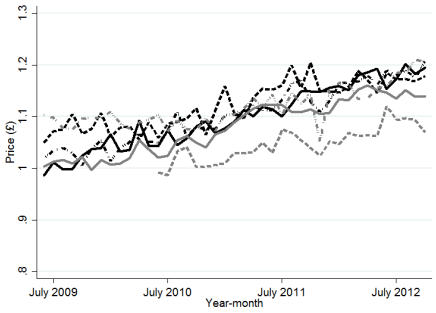
The other four price series are based on prices set by mainly smaller local stores, which make up around 80% of on-the-go purchases of soda. These vary geographically. We compute regional prices for the North, Midlands, South and London. On each choice occasion we observe where an individual shops, we assume that this is independent of demand shocks (see Section 2.2), and we assume that the consumer faces the vector of prices for products in the retailer that we observe them shopping in.

Figure A.2: *Price variation for Coca Cola*

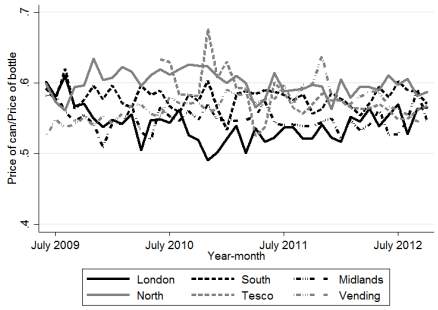
(a) 330ml can



(b) 500ml bottle



(c) Within brand price variation



Notes: Each line corresponds to a different retailer.

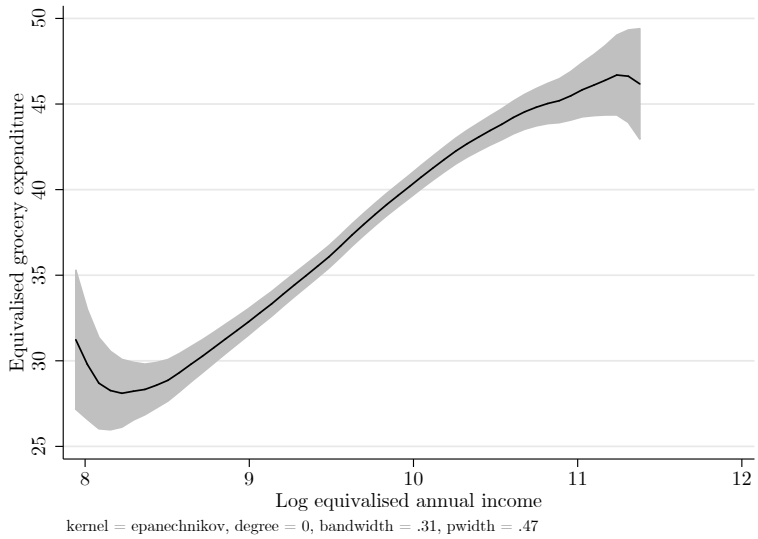
To illustrate the variation in prices that we use, in Figure A.2 we plot the evolution of prices over time for the 330ml can (panel (a)) and 500ml bottle (panel (b)) of Coca Cola. We control for time varying brand effects in the demand estimates, so this means that we exploit differential time series variation in prices across the

two container sizes and across 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.

A.3 Relationship between equivalized expenditure and income

We use total household grocery expenditure to proxy for household income. The Living Costs and Food Survey (LCFS) is an expenditure survey that collects data on spending for a repeated cross-section of households. It also contains information on household income. Figure A.3 shows that there is a strong relationship between households' annual equivalized income and equivalized weekly grocery spending.

Figure A.3: *Relationship between household income and grocery expenditure*



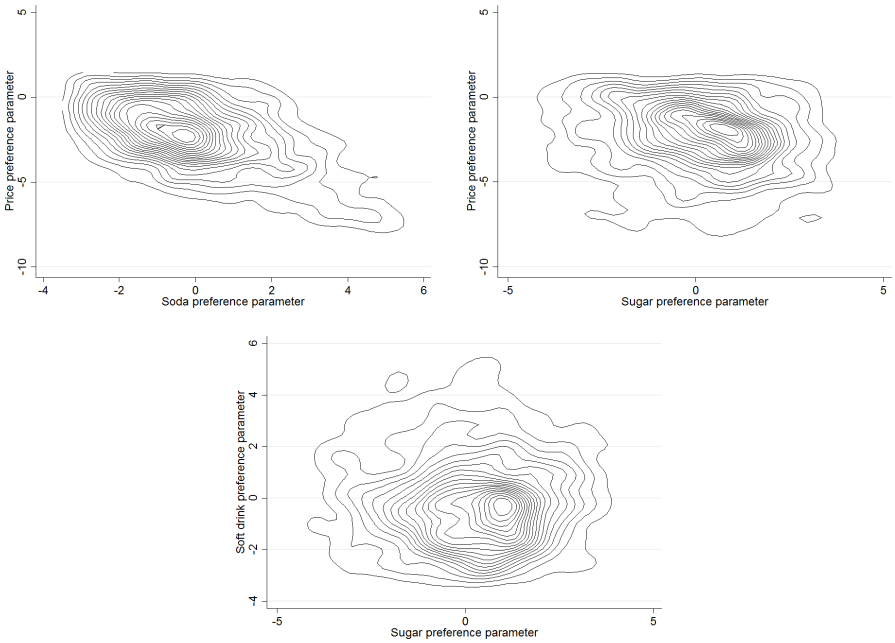
Notes: Figure drawn using data on 4937 households in the Living Costs and Food Survey 2011. The horizontal axis shows logged equivalized annual income of the household, and the vertical axis shows equivalized weekly grocery expenditure. Figure trims the 5th and 95th percentiles of the logged equivalized annual income distribution. We equivalise using the standard OECD modified equivalence scale (see Hagenaars et al. (1994)).

B Further details of demand estimates

B.1 Distributions of preference parameters

In Figure B.1 we plot contour plots of the bivariate preference distributions (based on the finite parts of the distribution). Figure B.2 shows how price and sugar preferences varies across the distribution of total equivalized grocery expenditure.

Figure B.1: *Bivariate distributions of consumer specific preference parameters*

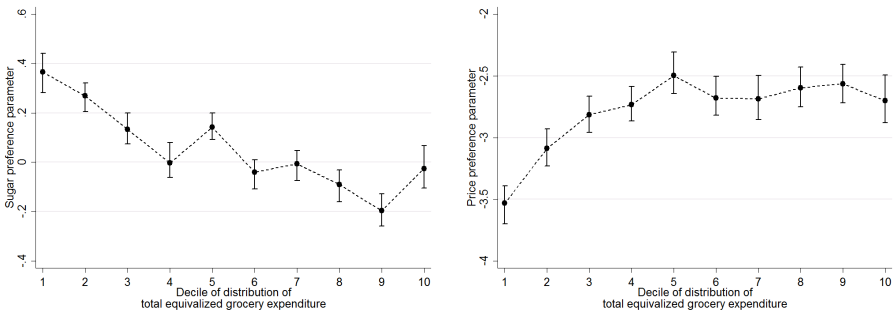


Notes: Distribution plots use consumers with finite preference parameters, those having infinite distaste for sodas or sugar are not included in this graph.

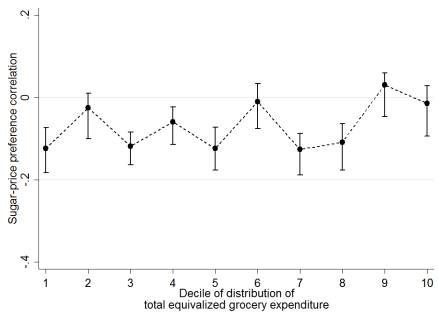
Figure B.2: *Preferences variation with equivalized expenditure*

(a) preferences for sugar

(b) preferences for price



(c) sugar-price preference correlation



Notes: Figure shows how the mean sugar and price preferences and the correlation between sugar and price preferences vary by equivalized expenditure deciles. 95% confidence intervals are shown by bars.

B.2 Price effects on demand

Table B.1: *Price Effects on Demand*

	Effect of 1% price increase on:				
	Own demand	cross demand for: sugary soda	diet soda	sugary alternatives	Total demand
Coca Cola 330	-2.56	0.25	0.08	0.05	0.01
	[-2.62, -2.51]	[0.25, 0.26]	[0.08, 0.08]	[0.05, 0.06]	[0.01, 0.01]
Coca Cola 500	-1.75	0.37	0.12	0.18	-0.07
	[-1.87, -1.63]	[0.35, 0.40]	[0.11, 0.13]	[0.17, 0.20]	[-0.07, -0.07]
Coca Cola Diet 330	-2.43	0.08	0.29	0.01	0.01
	[-2.49, -2.36]	[0.07, 0.08]	[0.29, 0.30]	[0.01, 0.01]	[0.01, 0.01]
Coca Cola Diet 500	-1.47	0.11	0.37	0.06	-0.05
	[-1.56, -1.37]	[0.10, 0.12]	[0.34, 0.39]	[0.05, 0.07]	[-0.05, -0.05]
Fanta 330	-3.31	0.06	0.02	0.01	0.00
	[-3.39, -3.22]	[0.06, 0.07]	[0.02, 0.02]	[0.01, 0.01]	[0.00, 0.00]
Fanta 500	-1.92	0.06	0.02	0.03	-0.01
	[-2.07, -1.78]	[0.06, 0.07]	[0.02, 0.02]	[0.03, 0.04]	[-0.01, -0.01]
Fanta Diet 500	-1.71	0.02	0.07	0.01	-0.01
	[-1.82, -1.60]	[0.02, 0.02]	[0.06, 0.07]	[0.01, 0.01]	[-0.01, -0.01]
Cherry Coke 330	-3.33	0.04	0.01	0.01	0.00
	[-3.41, -3.25]	[0.04, 0.04]	[0.01, 0.01]	[0.01, 0.01]	[0.00, 0.00]
Cherry Coke 500	-2.07	0.04	0.01	0.02	-0.01
	[-2.21, -1.94]	[0.04, 0.05]	[0.01, 0.01]	[0.02, 0.03]	[-0.01, -0.01]
Cherry Coke Diet 500	-1.73	0.01	0.04	0.01	-0.01
	[-1.86, -1.60]	[0.01, 0.01]	[0.04, 0.04]	[0.01, 0.01]	[-0.01, -0.01]
Oasis 500	-1.97	0.10	0.03	0.05	-0.02
	[-2.13, -1.82]	[0.08, 0.13]	[0.02, 0.04]	[0.04, 0.06]	[-0.02, -0.02]
Oasis Diet 500	-1.75	0.03	0.09	0.01	-0.01
	[-1.87, -1.64]	[0.02, 0.04]	[0.07, 0.11]	[0.01, 0.02]	[-0.02, -0.01]
Pepsi 330	-3.12	0.11	0.03	0.02	0.00
	[-3.20, -3.05]	[0.11, 0.12]	[0.03, 0.03]	[0.02, 0.02]	[0.00, 0.01]
Pepsi 500	-2.13	0.20	0.07	0.09	-0.04
	[-2.24, -2.03]	[0.19, 0.21]	[0.06, 0.07]	[0.08, 0.10]	[-0.05, -0.04]
Pepsi Diet 330	-3.43	0.03	0.18	0.00	0.00
	[-3.52, -3.34]	[0.03, 0.03]	[0.17, 0.18]	[0.00, 0.00]	[0.00, 0.00]
Pepsi Diet 500	-1.89	0.06	0.23	0.03	-0.04
	[-1.99, -1.80]	[0.06, 0.07]	[0.22, 0.25]	[0.02, 0.03]	[-0.04, -0.04]
Lucozade 380	-2.30	0.16	0.05	0.07	0.00
	[-2.41, -2.19]	[0.16, 0.17]	[0.05, 0.05]	[0.06, 0.07]	[-0.00, 0.00]
Lucozade 500	-1.96	0.10	0.03	0.05	-0.02
	[-2.10, -1.82]	[0.09, 0.10]	[0.03, 0.04]	[0.05, 0.06]	[-0.02, -0.02]
Ribena 288	-3.10	0.05	0.01	0.01	0.01
	[-3.31, -2.90]	[0.05, 0.06]	[0.01, 0.01]	[0.01, 0.01]	[0.01, 0.01]
Ribena 500	-1.97	0.05	0.01	0.03	-0.01
	[-2.17, -1.78]	[0.04, 0.06]	[0.01, 0.02]	[0.02, 0.03]	[-0.01, -0.01]
Ribena Diet 500	-1.70	0.01	0.04	0.01	-0.01
	[-1.83, -1.58]	[0.01, 0.02]	[0.04, 0.05]	[0.00, 0.01]	[-0.01, -0.00]
Fruit juice	-1.21	0.05	0.02	0.19	0.00
	[-1.34, -1.09]	[0.05, 0.06]	[0.02, 0.02]	[0.16, 0.21]	[0.00, 0.00]
Flavored milk	-1.48	0.04	0.01	0.09	-0.01
	[-1.59, -1.36]	[0.04, 0.05]	[0.01, 0.01]	[0.07, 0.10]	[-0.02, -0.01]

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. Numbers are means across time. 95% confidence bands in brackets.

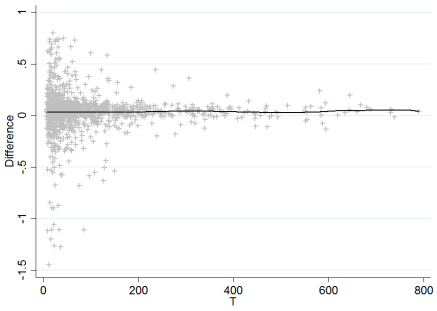
B.3 Incidental parameters problem

Figures B.3, B.4 and B.5 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) age and c) total dietary sugar. 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.

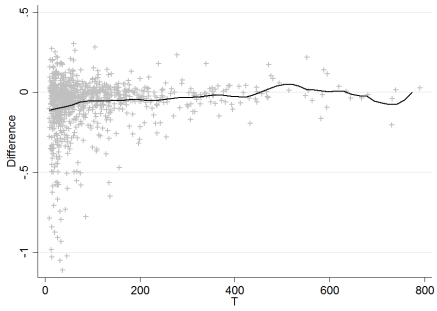
Figures B.6 plots the distributions of price, soda and sugar preference parameter estimates for both the estimators $\widehat{\theta}$ and $\widetilde{\theta}_{split}$, showing there is little difference in the distributions.

Figure B.3: *Relationship between bias and time in sample*

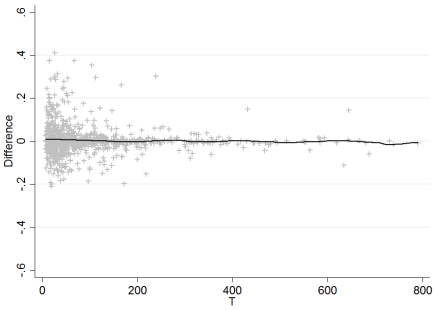
(a) Price



(b) Soda



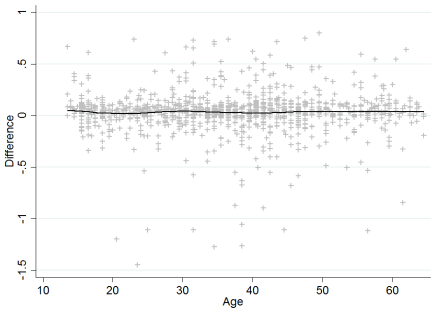
(c) Sugar



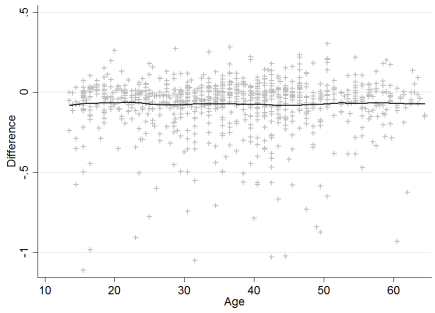
Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

Figure B.4: *Relationship between bias and age*

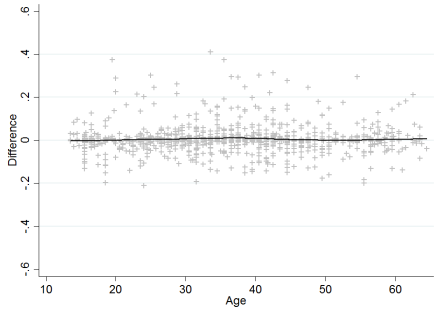
(a) Price



(b) Soda



(c) Sugar

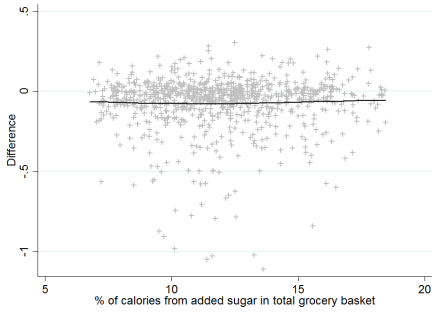


Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

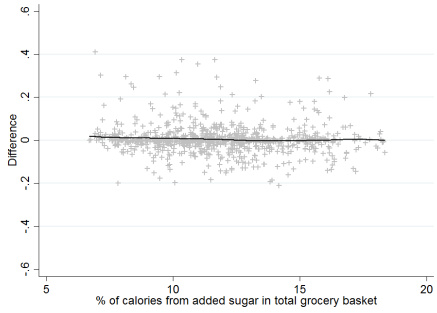
Figure B.5: *Relationship between bias and dietary sugar*

(a) Price

(b) Soda



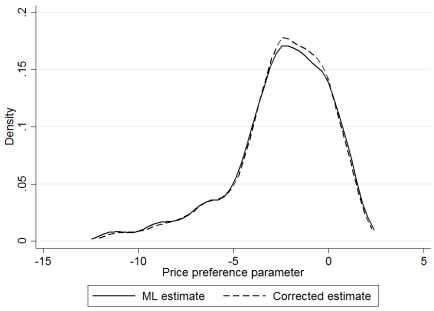
(c) Sugar



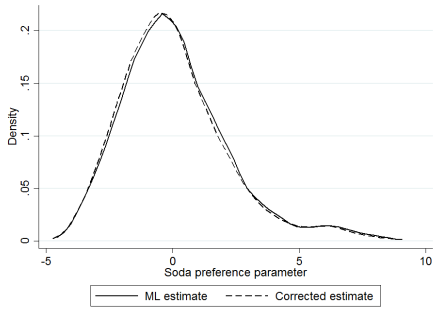
Notes: Marks represent consumer level differences. Lines are local polynomial regressions.

Figure B.6: *Preference parameter distribution*

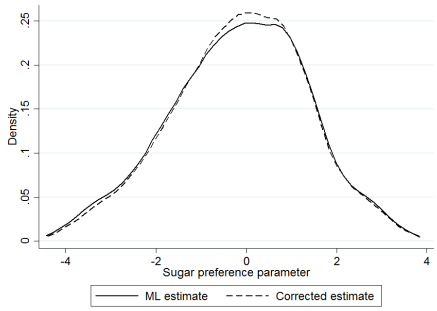
(a) Price



(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

$$p_{jt}^{cf} = \begin{cases} \tilde{p}_{jt}^{cf} + \tau l_j & \forall j \in \Omega_a \\ \tilde{p}_{jt}^{cf} & \forall j \in \Omega_n. \end{cases}$$

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 liter) using the same supply side model estimates in the first step and conducting the counterfactual simulation of pass-through of this price to consumer prices.

Table C.1 summarizes the impact of the broad soda tax on equilibrium prices and market shares (it contains analogous information to Table 4.1 in the main paper). The main difference between a tax incident on only sugary soda and one incident on all sodas is that the latter leads to prices increases for diet products

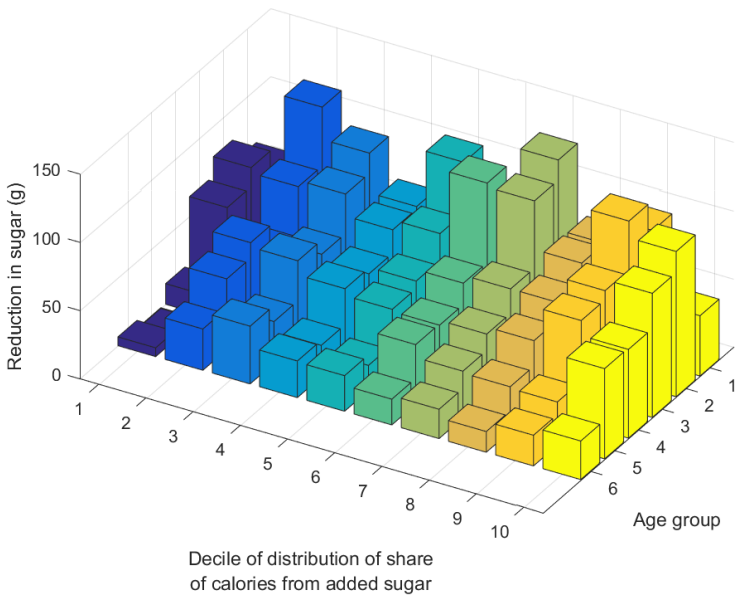
(that on average are similar to those for sugary products). The result is that the broad soda tax leads to a much smaller reduction in demand for sugary soda and a fall (rather than increase) in demand for diet sodas (relative to the sugary soda tax). Figure C.1 shows that a broad soda tax does achieve larger reductions in sugar among the young than the old, but fails to achieve relatively large reductions among those with high total dietary sugar.

Table C.1: *Effects of "broad" soda tax*

	Tax (pence)	Δ price (pence)	Δ share (p.p.)
Sugary soda	10.55	13.31	-1.83
Diet soda	11.44	17.24	-1.75
Sugary alternatives	0.00	0.00	0.73
Outside option	0.00	0.00	2.84

Notes: Numbers are means across products.

Figure C.1: *Reductions in sugar by age and total dietary sugar*



Notes: Sample includes soda purchasers and non soda purchasers. Numbers show how the mean reduction in sugar from broad soda varies by age and deciles of the distribution of share of calories from added sugar. Age groups are 1=<22, 2=22-30, 3=31-40, 4=41-50, 5=51-60, 6=60+.

D Substitution to food

The choice model we outline in Section 2 captures consumer choice between drink products $j = \{0, 1, ..., J\} = \Omega_{\mathcal{D}}$. The drink products comprise water $j = 0$, soda,

$j = \{1, ..., j'\} = \Omega_a$ and juice $j = \{j' + 1, ..., J\} = \Omega_n$ The expected utility to the consumer of purchasing a drink is:

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

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. These idiosyncratic shocks are sequentially observed.

This first stage choice probabilities are:

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

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) among all other drinks demand parameters. 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 as:

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

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 D.1 shows estimates of these parameters.

Table D.1: *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.