

Do Sugar Taxes affect the right consumers?

Valerio Serse[†]

(November 2019)

Abstract

Sugar taxes are often considered as a possible tool to tackle excessive sugar consumption. This paper estimates a dynamic multinomial Logit model of cola demand on a novel supermarket scanner dataset in order to study *preference heterogeneity* and *state dependence* in product choice. The model estimates allow evaluating the effectiveness of taxation in reducing demand for sugary colas across different consumer types. The results show that a sugar tax would be less effective among the targeted population of heavy sugar consumers. This policy, however, would be more effective among low-income households. Tax policy simulations show that a specific tax on sugar should be preferred to an ad-valorem tax on sugary colas on both corrective and equity grounds. This is because ad-valorem taxes can lead low-income households and heavy sugar consumers to substitute from expensive to cheaper sugary brands. Lastly, because households exhibit state dependence in cola choice, sugar taxes would be more effective in reducing sugar consumption in the long-run.

JEL No: D12, H31, I18, Q18.

Keywords: heterogeneity in preferences, state dependence, sugar taxes, discrete choice models.

[†] CORE, Université catholique de Louvain, FNRS-FRESH grant holder, e-mail : valerio.serse@uclouvain.be. Acknowledgements: I am very grateful to my advisor, Jean Hindriks, for helpful comments and discussions. I also would like to thank Liran Einav, Johannes Johnen, Leonardo Madio, Hassan Nosratabadi, Martin O'Connell, Frank Verboven and participants at the DSSG conference in Chicago, the ECORES summer school in LLN, the EC² conference in Rome, the RGS doctoral conference in Bochum, the WIPE workshop in Reus, the PET conference in Strasbourg, the PhD seminar at AMSE in Marseille and various seminars at the UCLouvain. The financial support from the FNRS is gratefully acknowledged.

1. Introduction

The obesity rate has substantially increased during the last decades (WHO, 2017), generating high costs in terms of healthcare and foregone productivity (Dobbs et al., 2014). Excessive sugar consumption is often blamed as one of the main drivers of this negative trend (Te Morenga et al., 2013). In this regard, some countries have already taken actions by implementing taxes aimed to discourage the excessive consumption of sugar-sweetened beverages (SSB). Assessing the effectiveness of these policies, however, requires a careful evaluation of consumers' reaction to price changes. While taxing tobacco affects mostly smokers, taxing SSB can affect various types of households, even those with low sugar consumption and a low health risk. It is therefore fundamental to understand whether sugar taxes will affect the targeted population of heavy sugar consumers and how the tax burden will fall on different population groups.

This paper studies the impact of sugar taxes on cola demand across different types of households. The aim is to understand how sugar taxes affect the probability of switching from a sugary to a diet cola (i.e., sugar-free) across households with different levels of income and sugar consumption. The analysis consists of estimating a discrete choice model of cola demand on a novel supermarket scanner dataset from a major retail chain in Belgium. The dataset includes information about the socioeconomic status and dietary habits of each household unit, which is tracked over time by the use of loyalty cards. The model estimates are then used to perform tax policy simulations, which allows studying the impact of different levels of taxation on sugary cola demand. These simulations are also used to compare the effectiveness of ad-valorem sugar taxes (i.e., taxes based on the product price) to *specific* sugar taxes¹ (i.e., taxes based on sugar content). The findings of this paper provide useful empirical insights for the optimal design of sugar taxes by focusing on their effectiveness in reducing demand among the targeted population (i.e., heavy sugar consumers) and their redistributive concerns.

The demand model accounts for both *preference heterogeneity* and *state dependence* in product choice. Preference heterogeneity means that households have some persistent differences in preferences for colas. State dependence instead means that households have a higher probability of choosing a product they have purchased in the recent past. These two behaviors have important policy implications. Accounting for preference heterogeneity in the demand model is crucial as it gives a measure of how different types of households, such as those with a high-sugar diet, would react to sugar taxes. Accounting for state dependence is also important as it implies that household demand is dynamic, and so the impact of sugar taxes on demand will tend to increase over time. State dependence and taste heterogeneity are not mutually exclusive, and there is a risk of empirically

¹ Sometimes referred as excise sugar taxes.

confound one with the other. This is because they can generate a similar sequence of choices with the consumer that keeps on choosing the same product over time. Hence, allowing simultaneously for both behaviors in the demand model is necessary for their correct identification (Heckman, 1978).

The model is estimated on a novel set of supermarket scanner data provided by a major group of Belgian retailers. The dataset is composed of 434,475 households that purchased at least five times one or more bottles of cola over two and a half years, for a total of 9.7 million shopping trips. Colas are an interesting market to analyze for at least two reasons. First, they make up the largest share of SSB, which are often considered as a possible target for public health taxes. Second, each of these products has a perfect healthier substitute, as most brands also have a diet version. Since this retail chain is a local price follower, there is substantial variation in cola prices both across stores and over time. Price data consists of the daily posted price specific to each cola in a given store. The dataset also includes information about the presence of quantity rebates and whether the product was advertised in the supermarket's weekly folder. Furthermore, there is data about which product was available in the stores on a given day. This provides complete information on the choice set faced by each household during any shopping trip.

The cola choice model accommodates random taste variation in product attributes and hence allows cross-price elasticities to depend on the similarity between products (Train, 2009). Households are allowed to have a different price sensitivity and heterogeneous preferences for sugary colas and bottle size. As the model can capture heterogeneity in price sensitivity across households, it also allows approximating the curvature of product demand. Which is an essential variable in predicting tax pass-through in imperfectly competitive markets (Weyl & Fabinger, 2013). The unobserved heterogeneity in taste across households is specified as a parametric distribution of the random taste coefficients. In order to understand how these tastes change with observed household characteristics, the means of these distributions are also allowed to vary with dietary habits and household demographics. State dependence is tested with the inclusion of the lagged choice variable in the utility specification while controlling for persistent heterogeneity in household preferences.

By exploiting the time dimension of the data, this paper studies preference heterogeneity across households and disentangle it from state dependence in cola choice. The identification of taste heterogeneity against state dependence is possible thanks to the frequent temporary price discounts and variations in the consumer choice set due to changes in product availability over time. The following example may give an idea of how these temporary market shocks are exploited for identification purposes. Imagine that there are only two brands, A and B, and that a household prefers brand A whenever it is available. One day brand A is not available, so the household chooses brand B. In the next shopping trip, brand A is back on the market. In the absence of state dependence, the

household should revert to brand A as it is again available. If it keeps choosing brand B, however, the choice would exhibit state dependence. This is because choosing B in the previous shopping trip (when A was not available) leads the household to choose B to A in the next shopping occasion (when A is again available).²

The results of this paper show that households have very heterogeneous preferences for colas and are state-dependent in cola choice. Once accounting for state dependence, the degree of unobserved taste heterogeneity across households decreases by one third. Hence, suggesting that part of the estimated heterogeneity in tastes was instead due to state dependence in cola choice.³ The model estimates show that heavy sugar consumers tend to prefer sugary to diet colas, while they are less sensitive to cola prices. This finding suggests that taxing sugary colas would be less effective in reducing demand among the targeted population and that the tax burden would mostly fall on these consumers. However, the results also show that low-income households are likely to experience larger health gains compared to richer households. This is because they have a greater preference for sugary colas and are relatively more price-sensitive than high-income households. Their higher responsiveness to sugar taxes also helps them reduce their tax burden in monetary terms.

The model estimates allow performing tax policy simulations, which are then used to compare the impact of different tax schemes on sugary cola demand. This analysis shows that specific taxes targeting sugar content should be preferred to ad-valorem taxes on sugary colas on both corrective and equity grounds. The main reason for this finding is that ad-valorem taxes would entail a larger substitution towards cheaper sugary brands. This effect is mostly concentrated among low-income households and heavy sugar consumers. Finally, because of state dependence in cola choice, households need some time before fully adjusting their demand to the implementation of a sugar tax. Hence, these policies would tend to be more effective in the long-run.

The paper is structured as follows. Section 2 reviews some of the most recent findings in the economic literature focusing on consumer demand estimation and junk food taxation. Section 3 describes the scanner data used for the estimation. Section 4 presents the cola demand model and its empirical specifications. Section 5 discusses the parameter estimates and shows the results of various tax policy simulations. In section 6, it is presented an extended version of the model, which accounts for both observed and unobserved heterogeneity in household preferences. Section 7 shows the results of various robustness checks. Section 8 concludes the paper.

² A similar reasoning can be done for temporary price discounts. Imagine a temporary price discount on brand B leading the household to switch from A to B. Once the price discount is over, the household should revert to its favorite brand A. Yet, if the household is state dependent, brand B may still be chosen against A, even though the discount is terminated.

³ This finding is robust to possible misspecifications of taste heterogeneity and autocorrelation in the error terms.

2. Background Literature

A substantial body of economic literature evaluates the impact of commodity taxes by means of consumer demand estimation. This approach allows estimating ex-ante the consumer response to a given change in tax policy. This exercise can shed a light on various aspects related to this type of policy. Such as, for instance, their distributional effect, their capacity to generate fiscal revenues or their effectiveness in reducing the consumption of unhealthy products.

It is a common practice to adopt models of consumer demand that are both consistent with economic theory and flexible enough to be estimated on commonly available purchase data. One of the most successful models in this fashion is the Almost Linear Demand System (AIDS) proposed by Deaton & Muellbauer (1980), which is suitable for the most common forms of consumer survey data. Various studies use this approach to estimate the impact of taxes on unhealthy food demand (see, for instance, Jensen & Smed, 2007; Chouinard et al., 2007; Nordström & Thunström, 2009; Allais et al., 2010; Sharma et al., 2014; Harding & Lovenheim, 2017). The main limitation of this class of models, however, is that it is based on a representative household, and hence, it does not focus on preference heterogeneity across consumers.

Understanding consumer heterogeneity plays a crucial role in the optimal design of sin good taxes. This is because policymakers may be primarily interested in the behavioral response of the targeted population. From a public health perspective, the targeted population should be those individuals overconsuming junk food products. Hence, it is important to understand how this particular group of individuals responds to price changes. Understanding preference heterogeneity is also crucial for the optimal tax design in an economic welfare framework. Some theoretical studies focused on the optimal design of sin good taxes, considering both market efficiency and redistributive concerns (see, for instance, Gruber & Kőszegi, 2004; O'Donoghue & Rabin, 2006). More recently, Griffith et al. (2018) study optimal corrective taxes with internalities in food consumption. They show that optimal corrective taxes depend on the empirical distribution of these internalities and how they correlate with the shape of demand and income distribution. In particular, they show that junk food taxes could be both efficient and progressive if low-income consumers are very sensitive to prices, have a higher marginal utility of income, consume more unhealthy food, and suffer more from internalities.

Similar results are obtained by Allcott et al. (2019), who study the optimal sin good tax with both internalities and externalities in consumption, taking into account redistributive concerns and tax revenue redistribution. They provide an optimal sin tax formula that can be directly computed using three sufficient statistics. Such as the consumer elasticities related to prices and income, a money-metric measure of consumer bias, and the progressivity of bias correction through taxation. By calibrating their model on Nielsen Homescan data, they find that the optimal soda tax in the US should

be between 1 and 2.1 cents per ounce. Their estimate of price and income elasticities is computed by regressing the natural log of SSB purchase on price and income. They address the possible simultaneity bias in demand estimation through an instrumental variable approach. However, they restrict heterogeneity in soda demand by allowing price elasticities to vary only with the income level.

A typical way to study preference heterogeneity in consumer demand estimation is to use Multinomial Logit (MNL) models of demand, such as those in Berry (1994), Berry et al. (1995), and Nevo (2000). These models are based on random utility theory and can be estimated on either aggregate or consumer-level data. Contrary to the AIDS, which can describe the allocation of expenditure across different product categories, the MNL is often adopted to study the consumer choice within a single category of products. This class of models sees the consumer as choosing across different goods the one that delivers the highest utility of consumption, which depends on prices and the consumer preference for product attributes. These models can deliver very realistic substitution patterns by allowing the taste for different product attributes to vary randomly across consumers. The resulting price elasticities of demand will depend on the similarity between products in terms of their characteristics. However, in contrast to the AIDS, the MNL does not allow any product in the demand system to be a complement to any other. Thus the choice set should be constructed carefully by including only the products that are closely substitutable with each other.

SSB are often considered as a possible target for sugar taxes given the established link between their consumption and body weight gain (Malik et al., 2013). Lately, various authors used MNL models of demand to study the potential impact of sugar taxes on the demand for these products. Bonnet & Requillart (2018) estimate a random coefficient MNL on a panel data of French households, accounting for heterogeneity in beverage consumption both within and across households. They show significant differences in the consumption of SSB across individuals. In particular, they find that SSB consumption increases with the Body Mass Index (BMI) in both adults and children. This finding suggests that taxes on SSB can be justified on public health grounds and internalities motives. However, they do not document any heterogeneity in price sensitivity across consumers. Wang (2015) proposes a MNL model of soda demand that accounts for both preference heterogeneity and product stockpiling. She shows that ignoring this behavior leads to a substantial overestimation of the price elasticity of demand for SSB. According to her estimates, this policy is unlikely to reduce soda consumption substantially, but it could be quite effective in raising fiscal revenues.

Dubois et al. (2017) circumvent the issue of soda stockpiling by studying the impact of sugar taxes on the demand for on-the-go drinks (which are less likely to be stored). They develop a nonparametric MNL model and estimate it on a panel of UK individuals that includes information about their purchases at the product level. Their model allows each individual to have a specific taste coefficient

for the price, soda, and sugary drinks, which are then correlated with their demographics and dietary habits. They find that heavy sugar consumers prefer sugary drinks and are relatively insensitive to price. While young and poor individuals, who also have a strong preference for sugary drinks, tend to be more price-sensitive. They also show that a tax on sugary sodas will achieve a larger reduction in sugar consumption than a more general soda tax since it leads to a greater substitution from sugary to diet sodas. However, heavy sugar consumers would be less likely to switch to a diet soda compared to other consumers. Their findings thus imply that a sugar tax is not well targeted at those individuals over-consuming sugary drinks.

Although these studies account for preference heterogeneity across consumers, they do not allow for other possible sources of choice inertia in soda demand. Heckman (1978) distinguished two types of brand inertia. One is the so-called *structural state dependence*, which means that the history of past purchases directly affects the current choice behavior. While the other is the *spurious state dependence*, which means instead that consumers have some persistent propensity in making certain decisions (e.g., taste heterogeneity). Importantly, these two sources of choice inertia are not mutually exclusive, and risk being empirically confounded with each other. Researchers have extensively documented persistence in product choice due to state dependence using consumer panel data (see Allenby & Lenk, 1994; Keane, 1997; Seetharaman et al., 1999; Erdem et al., 2008; Dubé et al., 2010). Some studies have also found that ignoring state dependence can overestimate the extent of preference heterogeneity across households (Allenby & Lenk, 1994; Keane, 1997). However, none of these works focused on testing state dependence in the context of SSB demand. This literature usually tests for state dependence by including the lagged choice variable in the consumer utility specification, while controlling for persistent heterogeneity in taste across consumers. Such behavior has important implications for corrective tax policies, as the reduction in demand due to a price hike tends to increase over time. As a result, static demand models can underestimate the long-run impact of sugar taxes on SSB demand.

This paper contributes to the economic literature on sugar taxes by estimating a MNL model of cola demand that accounts for both preference heterogeneity and state dependence in product choice. The dataset contains information about household demographics, such as income, educational level, age, and family composition. The 434,475 households included in the dataset represent 10% of all Belgian households. Compared to previous studies, the size of this sample permits exploring in detail how household tastes are distributed across the population. In particular, this allows understanding both observed (through household demographics) and unobserved heterogeneity in price sensitivity, taste for sugar, and package size. The information about households' dietary habits permits studying how price sensitivity and taste for sugar vary across households with different levels of sugar consumption.

This exercise allows assessing how a sugar tax can affect different types of households, thus providing useful empirical insights for an optimal sugar tax design.

In contrast with the existing literature, this paper uses posted price data that varies at the day-brand-store level. Posted prices are different from the conventional scanner price data used in the literature (e.g., Nielsen measured prices) as they are not conditional on purchase and thus less sensitive to local and cyclical shocks (Coibion et al., 2015). They are not dependent neither on measurement errors due to loyalty cards (Einav et al., 2010). The high frequency of daily price data at the store level provides an exact measure of the actual price faced by the household in every choice occasion. This feature overcomes the possible bias arising with higher levels of temporal and spatial price aggregation (Levin et al., 2017). By having information about the daily product availability in each store, the model also allows controlling for choice set heterogeneity across different shopping trips.⁴ An issue that is typically ignored in the estimation of SSB demand.⁵ The model does not consider any outside option. This approach allows focusing on the purchase probabilities of each brand conditional on buying a cola and circumvents the potential issues arising from censored demand, such as stockpiling of storable goods (Hendel & Nevo, 2006; Wang, 2015).⁶

The results show that a sugar tax would be less effective among households overconsuming sugary products, but they would be quite effective among the poor. This finding is in line with Dubois et al. (2017), which focused on the market for soda on-the-go and individual level demand in the UK. This paper, therefore, extends their findings to the market for larger bottle size and household level demand in Belgium. The tax policy simulations show that a specific tax on sugar may be preferred to an ad-valorem tax on both corrective motives and equity grounds. This is because an ad-valorem tax would imply a substitution toward cheap sugary brands, especially among low-income households. Interestingly, this result is consistent with Alvarado et al. (2019), who study the impact of the 10% ad-valorem tax on SSB implemented in Barbados in 2015. They report that while sales of the most expensive brands have declined by 7.2%, those of mid-range SSB have instead increased by 6.5%. Thus suggesting brand down-switching due to the implementation of the ad-valorem tax. This work also extends the empirical literature on state dependence by providing novel evidence about the existence of such behavior in the context of sin good markets.

⁴ See Goeree (2008) on how ignoring choice set heterogeneity can lead to biased estimates of price elasticities in the context of PC demand.

⁵ Dubois et al. (2017) allow the consumer choice set to vary across retailer type per year. Whereas this study allows for choice set heterogeneity at the day-store level.

⁶ Contrary to Dubois et al. (2017), who study the market of soda on-the-go, this paper focuses on the demand for colas in bottle (above 1 liter), which are more likely to be stored.

3. The Data

The data used in this work are provided by a major group of food retailers in Belgium. This group controls more than 300 local stores and has a market share of 33% (in 2017) in the country. The dataset contains information on all purchases made by households in the supermarkets of the group. Thanks to the use of loyalty cards, it is possible to follow the purchase history of each household unit over time. This data allows estimating both persistent heterogeneity in taste across consumers and state dependence in product choice. Having information about the full shopping basket for each household, it is also possible to compute the share of sugary-dense products on the overall food purchases. This variable is used as a proxy for the household's habits in sugar consumption.

The dataset is constructed as follows. There are three different data sources: (i) the purchase data, (ii) household demographics, and (iii) the store data. Purchase data comes from the scanning of each item at the pay desk. The purchase history of each household is collected by the use of their loyalty card in the payment transaction. Household demographics are collected at the residential level by an external company through the household's address. Store data are automatically collected daily by the retail chain. This data includes information about the market environment of the store, such as prices, the set of available products, and the presence of quantity rebates. These three sources of data are then merged to recreate the original choice set faced by the household during the day of purchase, together with the information about the products and the household characteristics.

This paper focusses on the demand for cola products, which represents two-thirds of their total SSB purchases. In the period going from August 2015 to January 2018, 434,475 households purchased at least five times one or more bottles of cola in these stores. For each of these households, there is information about their demographics either at the district of residence or at the residential address level. Sugar consumption is measured as the share of sugary-dense products on the overall food and beverage expenditures for each year.⁷ However, inferring eating habits for households making sporadic purchases in these stores could be misleading. Imagine, for instance, that a household buys only sugary-dense products in these stores and purchases other food elsewhere. This household would then have a share of sugary-dense products on the total shopping basket close to 100%. To address this issue, the model focussing on observed heterogeneity in preferences is estimated on the subsample of households spending most of their food budget in this retail chain. That is, at least 80% of their food-at-home budget.⁸ This leaves 130,514 households and around 4.2 million shopping trips where they bought a cola. Table 1 below shows the summary statistics about household demographics

⁷ The entire list of sugary-dense product included in this calculation can be found in the appendix.

⁸ The figures about the total food-at-home spending by household type are provided by Gfk Belgium.

in the entire sample. While Table 2 displays some information about the share of sugary-dense products and shopping trips in the subsample of frequent costumers.⁹

Table 1: Household demographics (entire sample)

Variable	Mean	Std. Dev.	Min.	Max.
Income	24,447.72	4,098.37	5,836	82,575
Education	0.26	0.18	0	0.85
Household members	3.23	1.28	1	9
Kids	0.53	0.50	0	1
Age > 70	0.07	0.25	0	1
70 > Age > 58	0.14	0.35	0	1
58 > Age > 46	0.31	0.46	0	1
46 > Age > 34	0.34	0.47	0	1
Age < 34	0.13	0.34	0	1

Notes: These variables are the following: **Income** - median income per capita of the district; **Education** - the percentage of higher educated people in the district; **Household members** -number of household members at the address level; **Kids** - dummy equal to 1 if there are children in the household, at the address level; **Age** - dummy by age group for the older member of the household, at the address level.

The market considered is the one for bottled colas having at least 0.5% of market share. This market is composed of a set of seven brands made by three different producers. These brands can be sold in different bottle sizes: 1 litre, 1.5 litres, or 2 litres. Just two of them (brand A and B) are present in all these three different formats. In total, there are fourteen different cola products. Two of these can be considered as discount products (Brand D.1 and D.2), while others may be perceived as high-quality brands. Six of these colas contain sugar while the others do not (diet colas). Sometimes the retailer or the producer applies quantity rebates where the unit price of a product decreases conditionally on buying a certain quantity. The dataset also includes information about which product was included in the weekly folder of suggestions and deals that is sent home to customers. There is full information about which product was on the supermarket shelves in a given day and which one was out of stock. This information allows controlling for choice set heterogeneity across stores over time. Table 3 below summarizes the characteristics of the products analyzed

Table 2: Sugar share and store visits (frequent costumers)

Variable	Mean	Std. Dev.	Min.	Max.
Sugar share	0.08	0.04	0.00	0.48
N° of store visits	32.38	30.64	5	300

⁹ Figure A.1. (in the appendix) shows the distribution of the share of sugary products across households.

Table 3: Product characteristics

Product	Sugar	Size (L)	Not in store	Price/L (μ)	Price/L (Min)	Price/L (Max)	Market Share
Brand A.1	sugary	1	2.88%	1.57	0.92	1.70	12.91%
Brand A.2	sugary	1.5	2.52%	1.18	0.57	1.39	18.46%
Brand A.3	sugary	2	51.63%	1.08	0.65	1.39	5.78%
Brand B.1	diet	1	10.74%	1.55	0.77	1.70	9.86%
Brand B.2	diet	1.5	3.47%	1.18	0.57	1.39	17.97%
Brand B.3	diet	2	65.84%	1.08	0.65	1.37	2.45%
Brand C.1	diet	1	25.08%	1.59	0.92	1.70	2.75%
Brand C.2	diet	1.5	4.96%	1.22	0.57	1.39	6.58%
Brand D.1	sugary	1.5	4.11%	0.32	0.23	0.40	5.32%
Brand D.2	diet	1.5	4.37%	0.30	0.17	0.39	2.45%
Brand E.1	diet	1	69.79%	1.26	0.89	1.35	0.74%
Brand E.2	diet	2	3.49%	0.76	0.42	1.05	12.43%
Brand F.1	sugary	1.5	44.33%	1.24	0.68	1.36	0.99%
Brand G.1	sugary	1.5	50.90%	0.77	0.45	0.98	1.31%

Notes: The third column shows the package size of each item in terms of litres of product in a bottle. The fourth column, “Not in-store,” shows the percentage of observations where the product was not available in the store. The fifth, sixth, and seventh column display the mean, the maximum and the minimum price per litre respectively.

Unlike the conventional scanner price data used in the literature (e.g., Nielsen measured prices), this study uses more detailed data on posted prices in each store. The advantage of using posted prices is that they are not conditional on purchase and thus less sensitive to local and cyclical shocks (Coibion et al., 2015). They are not dependent neither on measurement errors due to loyalty cards (Einav et al., 2010). The retailer automatically collects posted prices daily for every item sold in each store of the group. In this way, it is possible to recover the full set of cola prices faced by the household during each shopping trip. These prices vary considerably both across stores and overtime. This is because these retailers are publicly committed to act as local price followers. Hence, their prices are matching those of other local competing supermarkets.¹⁰ Figure A.2 and A.3 in the appendix show the price dispersion over stores and time, respectively.¹¹

¹⁰ Hindriks & Serse (2019) use spirit price data from the same supermarket chain and show a significant price dispersion over space due to the price-matching strategy of this group.

¹¹ Price dispersion over stores is computed as the difference between each product price and its daily average across all stores. While price dispersion over time is computed as the difference between the product price and its average in time.

4. The cola choice model

Consider a model where each household i , at any shopping occasion $t = 1, 2, \dots, T_i$, chooses one cola j from a time-varying choice set of J_t different colas. Each household maximises its utility by choosing the cola that delivers the highest utility of consumption. Let u_{ijt} represents the utility of household i conditional on buying product j on purchase occasion t and \bar{U} denote the highest utility over all the alternatives in the set J_t . Then, the household i at time t will purchase good j only if:

$$u_{ijt} > \bar{U}.$$

The conditional utility u_{ijt} can be decomposed into $u_{ijt} = V_{ijt} + \varepsilon_{ijt}$. Where V_{ijt} represents the part of utility detectable from the observable data (to the researcher) and ε_{ijt} is the random error term. Let the conditional utility u_{ijt} be given by:

$$u_{ijt} = \alpha_{j_B} + \pi_i(Size)_j + \delta_i(Sugary)_j + \theta_i(Rare)_j - \beta_i(Price)_{ijt} + \\ + \varphi_i(Rebate)_{jt} + \omega_i(Folder)_{jt} + \gamma_i I(d_{t-1})_{ijt} + \varepsilon_{ijt}. \quad (4.1)$$

The first four terms of the utility specification in 4.1 are all those characteristics that are product specific and time-invariant. α_{j_B} is the brand-specific constant capturing all those factors that are brand specific and that do not change over time. These may be generally interpreted as the brand image or appreciation. $Size_j$ is the bottle's size measured in litres. Its coefficient π_i captures the household preference for a larger bottle size. $Sugary_j$ is a dummy variable equal to one if cola j contains sugar. The coefficient δ_i measures the household taste for sugary colas, which can be either negative (it dislikes sugar) or positive (it likes sugar). As not every cola is always available, the model includes the term $Rare_j$, which is a dummy variable indicating whether cola j is rarely present in the supermarket shelves (i.e., missing more than one-third of the time). Its coefficient θ_i captures the household appreciation for less commonly available colas once they are purchasable in the store.

All those time-varying variables affecting the household utility u_{ijt} are placed from the fifth to the eighth term of eq. 4.1. $Price_{ijt}$ is the posted price of cola j faced by household i during the shopping trip t .¹² Its coefficient β_i measures the price sensitivity of household i . $Rebate_{jt}$ and $Folder_{jt}$ are dummy variables equal to one if cola j was on quantity rebate or in the weekly folder during the shopping trip t .¹³ Their associated coefficients, φ_i and ω_i , measure the household sensitivity to

¹² Prices are indexed by household since they vary across stores. Thus, two households can face different prices if they shop during same day but in two different stores.

¹³ These two variables are not indexed by household since they are applied nationally and hence do not change over stores.

rebates and the inclusion in the weekly folder, respectively. The term $I(d_{t-1})_{ijt}$ is an indicator variable that equals one if household i purchased cola j during its previous shopping trip (i.e., at time $t - 1$). Its coefficient γ_i captures the household state dependence in cola choice. If $\gamma_i > 0$, the household exhibits positive state dependence. That is, the probability of choosing cola j increases when cola j is purchased in the previous shopping trip. Conversely, if $\gamma_i < 0$, the household is variety-seeker. It likes to try different colas, so the probability of choosing cola j decreases if cola j was purchased in the previous shopping trip.

The model does not include any outside option. This allows focusing on the purchase probabilities of each brand conditional on buying a cola, thus circumventing the potential issues arising from cola stockpiling (see Hendel and Nevo, 2006; Wang, 2015). This approach does not permit variations in the overall cola expenditure overtime. Therefore, the price elasticity of demand indicates how price changes lead households to substitute from one cola to another regardless of how much cola they buy. This is still appropriate for analyzing the corrective impact of sugar taxes since leading households to switch from a sugary to a diet cola can account for a significant reduction in sugar intake. Depending on the assumptions about the distribution of these taste coefficients, we can specify two different types of MNL models: 1) the simple Logit and 2) the Mixed Logit.

4.1 Homogeneous Tastes

Let the vector of household i 's taste coefficients be denoted by $\vartheta_i = (\alpha_{j_B}, \pi_i, \delta_i, \theta_i, \beta_i, \varphi_i, \omega_i, \gamma_i)$. In the simple Logit model, taste coefficients are assumed to be fixed across households. Hence, the index i in the vector of taste coefficients ϑ_i can be dropped, thus becoming $\vartheta = (\alpha_{j_B}, \bar{\pi}, \bar{\delta}, \bar{\theta}, \bar{\beta}, \bar{\varphi}, \bar{\omega}, \bar{\gamma})$. Assuming the error term ε_{ijt} to be iid according to a type I extreme-value distribution, the probability that household i chooses good j at time t is then given by the following Logit formula (McFadden, 1974):

$$L_{ijt} = \frac{e^{\vartheta' X_{njt}}}{\sum_j e^{\vartheta' X_{njt}}}. \quad (4.1.1)$$

Where X_{njt} is the vector of product characteristics and the other explanatory variables in Eq. 4.1, which multiplies the vector of taste coefficients ϑ . The aggregated market share of good j , s_j , is the average of the choice probability L_{ijt} across each household i and shopping trip t :

$$s_j = \frac{\sum_i \sum_t L_{ijt}}{N}.$$

Where N is the total number of observations.¹⁴ Thus, by assuming the set $J = \{j_1, j_2, \dots, j_n\}$ to cover the entire market of interest, we would have that $\sum_j s_j = 1$. Own and cross-price elasticity of the Logit demand for good j can be derived by taking the derivative of s_j with respect to the price of cola z (denoted by $Price_z$):

$$\eta_{jz} = \frac{\partial s_j}{\partial (Price)_z} \frac{(Price)_z}{s_j} = \begin{cases} -\beta (Price)_j (1 - s_j) & \text{if } j = z \\ \beta (Price)_z s_z & \text{if } j \neq z \end{cases} \quad (4.1.2)$$

Although this model has the advantage of having a simple closed-form solution for the choice probabilities, it imposes very restrictive and unrealistic substitution patterns. This is because the relative choice probability for any couple of goods does not change for any variation in attributes of a third product (i.e., IIA property). Another limitation is that this model does not account for the correlation of errors for each household over time. Which is likely to arise in case of repeated purchases (i.e., panel data). The presence of dynamics in unobservable factors can induce inconsistency in the estimation of state dependence, as the lagged dependent variable is correlated with the current error term (Train, 2009). Foremost, the assumption of homogeneous taste across households can heavily restrict the demand analysis since it only focuses on the average impact of changes in product attributes on demand.

Although the restrictive substitution patterns and homogeneity in taste parameters, the Logit model has the advantage of being easy to estimate on large samples. Hence, this model can be estimated on the entire dataset, which includes 434,475 households and around 9.7 million shopping trips. The results of this estimation can give a first insight into the average household taste, price sensitivity, and state dependence by including as many observations as possible. The model is estimated through Maximum Likelihood procedure, with the likelihood function taking the following form (McFadden, 1974; Train, 2009):

$$LL(\vartheta) = \prod_i \prod_t \prod_j (L_{ijt})^{y_{ijt}}.$$

Where $y_{ijt}=1$ if household i chooses good j at time t and 0 otherwise. ϑ is the vector of fixed taste parameters to be estimated. Its estimator is the gradient vector that maximizes $LL(\vartheta)$.

4.2 Unobserved Taste Heterogeneity

While the simple Logit assumes that taste coefficients are homogeneous across households, the Mixed Logit model allows them to vary randomly in the population. This model has the advantage of accounting for the panel structure of the dataset by allowing unobserved factors that are household-

¹⁴ One observation is intended as a single shopping trip made by an household.

specific to be correlated over different shopping trip t . In the Mixed Logit, the probability that household i makes a given sequence of choices, conditional on its vector of taste coefficients ϑ_i , is given by the following product of logit formulas (McFadden, 1974; Train, 2009):

$$L_{ij}(\vartheta_i) = \prod_{t=1}^{T_i} \left(\frac{e^{\vartheta_i' X_{njt}}}{\sum_j e^{\vartheta_i' X_{njt}}} \right). \quad (4.2.1)$$

Where X_{njt} is the vector of product characteristics and the other explanatory variables that multiplies the vector of household-specific taste coefficients ϑ_i . Contrary to the simple Logit, this specification allows price sensitivity β_i , taste for sugar δ_i , and preference for package size π_i to vary randomly in the population and to be correlated with each other. In particular, β_i is assumed to be distributed as normal truncated at zero (i.e. $\beta_i \sim TN_{[0,+\infty]}[\bar{\beta}, \sigma_{\beta}^2]$) so that every household dislikes high prices. While both δ_i and π_i are assumed to be normally distributed (i.e. $\delta_i \sim N[\bar{\delta}, \sigma_{\delta}^2]$ and $\pi_i \sim N[\bar{\pi}, \sigma_{\pi}^2]$) so that they either like or dislike sugary colas and larger bottle size. The other taste coefficients are assumed to be fixed in order to ease the estimation procedure. As a result, the vector of household i 's taste coefficients can be now denoted as $\vartheta_i = (\alpha_{j_B}, \pi_i, \delta_i, \bar{\theta}, \beta_i, \bar{\varphi}, \bar{\omega}, \bar{\gamma})$. Since ϑ_i is not observable, the unconditional choice probability will be the integral of $L_{ij}(\vartheta_i)$ over all possible values of ϑ_i :

$$s_j = \int L_{ij}(\vartheta) f(\vartheta|W) d\vartheta. \quad (4.2.2)$$

Where $f(\vartheta|W)$ are the density functions of the taste coefficients. With ϑ and W being the parameters of these distributions. s_j is the average probability of choosing good j and can be interpreted as the market share of the alternative j . From eq. 4.2.2 we can derive the price elasticity of demand for the alternative j with respect to any other alternative z :

$$\eta_{jz} = \frac{\partial s_j}{\partial (Price)_z} \frac{(Price)_z}{s_j} = \begin{cases} -\frac{(Price)_j}{s_j} \int \beta_i L_{ij}(\vartheta) [1 - L_{ij}(\vartheta)] f(\vartheta|W) d\vartheta, & \text{if } j = z \\ \frac{(Price)_z}{s_j} \int \beta_i L_{ij}(\vartheta) L_{iz}(\vartheta) f(\vartheta|W) d\vartheta, & \text{if } j \neq z. \end{cases} \quad (4.2.3)$$

The upper term in the bracket represents the own-price elasticity of demand (i.e., the percentage change in the probability of purchasing alternative j given a percentage change in its price). While the lower term in the bracket is the cross-price elasticity of j with respect to z (i.e., the percentage change in the probability of purchasing alternative j given a percentage change in the price of alternative z). The product shares do not necessarily drive these price elasticities as in Eq. 4.1.2. Here, every household have different price sensitivity β_i . The price elasticity of the aggregate demand can therefore be interpreted as the mean price sensitivity across households, weighted by their specific

purchase probabilities. The correlation between $L_{ij}(\vartheta)$ and $L_{iz}(\vartheta)$ drives the substitution patterns over different values of ϑ . This means that products having similar characteristics, such as price, sugar content, and package size, will be closer substitutes to each other.

This approach can provide precious insights not only on the average “treatment effect” of a given policy but also on its heterogeneous impact across households even in the absence of demographic data. As the model can capture heterogeneity in price sensitivity across households, it also allows estimating the curvature of aggregate demand. Which is a fundamental variable in predicting tax pass-through in imperfectly competitive markets (Weyl & Fabinger, 2013). Contrary to the Logit, this specification accounts for correlation in unobserved factors over time and can provide a robust estimate of state dependence coefficient $\bar{\gamma}$ as it controls for permanent taste heterogeneity across households.¹⁵

A limitation of the Mixed Logit, however, is its heavy computational burden as compared to the simple Logit. The estimation procedure often requires a large amount of RAM, which makes such models challenging to estimate on vast datasets. For the ease of computation, this model is estimated on a random subsample of 10.000 households through Simulated Maximum Likelihood. Following Train (2009), the estimation procedure can be summarized as follows. First, it is taken a draw of ϑ from $f(\vartheta|W)$ and it is computed the following logit formula $e^{(\vartheta'_i X_{njt})} / \sum_j e^{(\vartheta'_i X_{njt})}$ for each shopping trip t . Second, it is taken the product of these logit formulas in order to get the simulated probability $L_{ij}(\vartheta^R)$ with this first draw. Third, these two steps are repeated for R times and results are averaged. This exercise gives the following average simulated probability:

$$\tilde{L}_{ij} = \frac{1}{R} \sum_{r=1}^R L_{ij}(\vartheta^R). \quad (4.2.4)$$

Which is then plugged into the log-likelihood function to obtain the simulated log-likelihood (SLL):

$$SLL = \sum_{i=1}^I \ln \left\{ \prod_{j=1}^J \tilde{L}_{ij}^{y_{ijt}} \right\}. \quad (4.2.5)$$

Where $y_{ijt} = 1$ if the household i purchase good j in shopping occasion t and zero otherwise. The SLL function is then maximized to obtain the parameters W of the density functions $f(\vartheta|W)$ of the taste coefficients.

¹⁵ Accounting for correlation in the household unobservable factors overtime through random taste coefficients does not necessarily rule out the possibility of autocorrelation in the remaining error terms (see, for instance, Keane, 1997; Dubé et al., 2010). This issue is addressed in section 7.

4.3 Identification

A key parameter to be estimated is the price sensitivity coefficient β_i . Which measures the causal impact of price changes on the probability of choosing a cola product. Similarly to Dubois et al. (2017), this parameter is estimated by exploiting two sources of price variation. The first is the cross-store variation in prices and choice set over space. The second is the price variation within the brand across different bottle sizes, which can vary over both stores and time. These two sources of price variation deliver a substantial heterogeneity in price vectors and choice sets faced by households over different shopping trips. The price vector and choice set considered for each household are those of the store they visited on a given shopping trip. As mentioned in the data section, these two vary over both stores and time daily. Hence, there is complete information about prices and choice set faced by the household during that day.

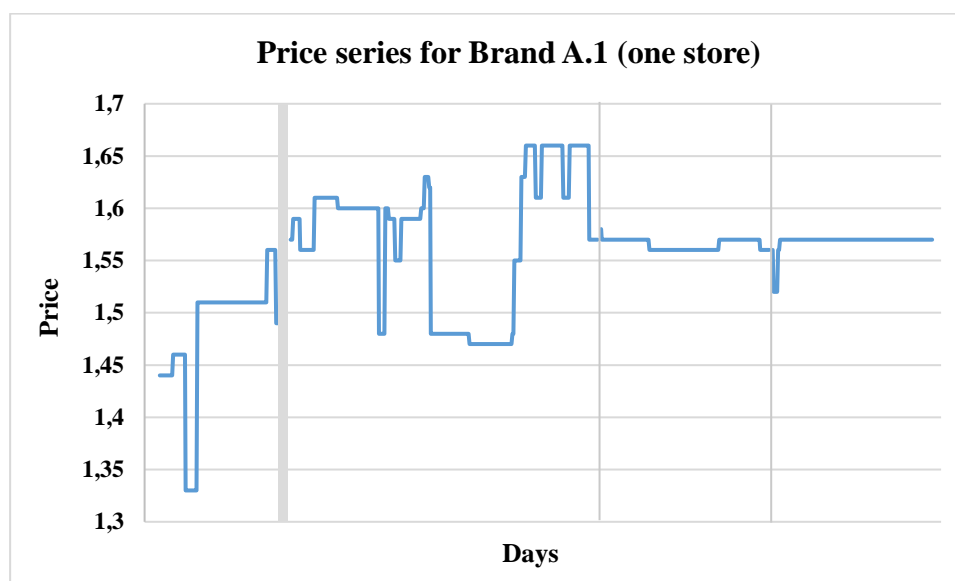
The fact that each store is committed to match the lowest price in its local market ensures enough variability in prices over choice occasions. Moreover, this price-matching strategy reduces the problems arising from households selecting the store based on the price level of their favourite brand across different chains. These households could be problematic because they would appear in the sample only when the price of their favourite cola is the lowest in this retail chain. Hence, making their choices based on prices of other retailers that are unobserved. The local price-matching strategy limits this issue because the observed vector of prices in these stores reflects those of other retail chains in the area.

A primary identifying assumption is that systematic demand shocks for different brand types do not drive cola price changes. In other words, price variations must be exogenous to household demand. This assumption seems quite reasonable in this setting, given the daily frequency of price updates and the household level demand data. However, this assumption would be invalid if stores can anticipate brand-specific demand shocks and change prices accordingly. In this study, such a scenario is presumed to be unlikely. Instead, it is assumed that variations in prices are the resultant of different cost shocks and random temporary price discounts that are induced by the retailer price-matching strategy. To give an idea of such price variation, Figure 1 displays the price evolution of one brand in a given store over time.¹⁶ As shown in the figure, the price series displays a large number of temporary price cuts.¹⁷ The assumption here is that these price cuts are not the resultants of stores anticipating temporary negative taste shocks for a given brand. They are instead assumed to occur due to random temporary price promotions induced by the retailer's price-matching strategy.

¹⁶ The grey areas in the graphs indicate that the product was not available (not on shelves) during those days in that store.

¹⁷ Each brand-store specific price series has a different shape. This is because price discounts are not necessarily coordinated across stores as each establishment follows the prices in its local market.

Figure 1: Example of price variation



Another critical assumption for the identification of the price coefficient is that price changes are not correlated with unobserved product characteristics. The extensive set of control variables such as rebates, folder information, and product availability limits this issue as they control for some characteristics of the store environment that are typically unobserved with other types of scanner data. Yet, it cannot be ruled out that omitted variables in the utility specification are correlated with prices. In a further specification, brand-specific effects are allowed to vary monthly in order to control for possible temporary unobserved shocks at the brand level that are common across all stores. This approach controls notably for advertizing campaigns, which are mostly done nationally and are not observed in the data.¹⁸

In addition to price sensitivity, this work focuses on identifying household taste heterogeneity and state dependence in product choice. The identifying assumption for taste heterogeneity is that persistent household preferences for colas can be decomposed into a vector of preferences for three product characteristics. Which are price, sugar and bottle size. These preferences are assumed to be parametrically distributed in the population and can vary with observed household characteristics. State dependence is identified as systematic deviations from the persistent household preferences that are driven by the previous shopping experience (lagged choice variable). The variation in prices and choice set overtime is fundamental in disentangling taste heterogeneity from state dependence. These temporary market shocks can in fact lead the household to switch from its favourite product to another cola. In the absence of state dependence, the household should revert to its favourite cola once these

¹⁸ Time-varying brand effects are not included in the main specification as they substantially increase the computational burden, hence making the tax simulation exercise more cumbersome. However, as shown in section 7, the price coefficient does not change much when including these controls.

shocks are over. However, if the household keeps on choosing the last product purchased, then there is evidence of dynamics in cola choice that is driven by the recent shopping history of the household. The correct identification of state dependence requires ruling out other possible sources of choice inertia that might be confounded with a real structural dynamic in product choice (Heckman, 1978). If the underlying distribution of taste heterogeneity across the population is not correctly specified, there is a risk of capturing some spurious state dependence that is due to some persistent unobserved propensity of the household in making a given choice. Moreover, even if the household heterogeneity in taste is correctly specified, it is still possible that the estimated choice dynamics are due to autocorrelated taste shocks. In this scenario, there can be a spurious state dependence just because the lagged choice variable is capturing a large random utility draw on both the previous and current shopping trip. As denoted by Dube et al. (2010), the implications of this model are remarkably different from those of a model with structural state dependence. In a model with autocorrelated errors, choice dynamics are generated by unobserved factors and hence cannot be influenced by any marketing variable. In the context of sugar taxes, this means that governments cannot exploit state dependence through a tax hike to generate a larger reduction in cola demand over time. The issues of misspecified heterogeneity and autocorrelated error terms are addressed in section 7, where various robustness checks are performed in order to rule out these alternative explanations for the observed choice dynamics.

5. Results

5.1 Parameter estimates

Table 4 shows the parameter estimates for both the Logit and Mixed Logit version of the utility specification in Eq. 4.1. Each of them is estimated with and without the state dependence variable (lagged choice). The models without the state dependence variable are labelled as *static* models. Their estimates are displayed in columns (1) and (3). While those accounting for state dependence are labelled as *dynamic* models, and their estimates are shown in columns (2) and (4). The first two columns of Table 4 show the parameter estimates of the Logit model, which assumes taste homogeneity across households. The last two columns display the parameter estimates of the Mixed Logit model, which assumes heterogeneity in taste across households instead. As the Mixed Logit model allows for price, sugar and size parameters to be randomly distributed, these columns also include estimates of their standard deviations. Which are labelled σ_β , σ_δ and σ_π respectively. The coefficients $\bar{\beta}$, $\bar{\delta}$ and $\bar{\pi}$ are the estimates of their means. All models are estimated on a random subsample of 10,000 households. The Logit models are also estimated on the entire dataset and results are compared to those obtained with the random subsample. These results can be found in table A.1

in the appendix. The comparison of the parameter estimates for the Logit models across different samples indicates very similar results. Thus suggesting that the Logit results are not sensitive to the reduction in sample size.

Table 4: Parameter Estimates

Parameter	Taste Homogeneity		Taste Heterogeneity	
	Static (1)	Dynamic (2)	Static (3)	Dynamic (4)
Price $\bar{\beta}$	0.816 (0.028)	0.450 (0.037)	4.441 (0.040)	2.531 (0.053)
Price σ_{β}	-	-	6.118 (0.031)	4.132 (0.040)
Sugary $\bar{\delta}$	0.782 (0.016)	0.492 (0.018)	1.266 (0.027)	0.782 (0.030)
Sugary σ_{δ}	-	-	3.818 (0.018)	2.428 (0.019)
Size $\bar{\pi}$	0.291 (0.021)	0.175 (0.026)	-0.459 (0.022)	-0.441 (0.024)
Size σ_{π}	-	-	5.441 (0.018)	2.977 (0.019)
Rare $\bar{\theta}$	-1.039 (0.011)	-0.500 (0.013)	-1.958 (0.012)	-0.973 (0.013)
Folder $\bar{\omega}$	-0.014 (0.007)	0.046 (0.010)	-0.093 (0.010)	-0.026 (0.011)
Rebate $\bar{\varphi}$	0.079 (0.007)	0.064 (0.009)	0.176 (0.009)	0.163 (0.010)
State Dependence $\bar{\gamma}$	-	3.033 (0.005)	-	1.854 (0.005)
Model	Logit	Logit	Mixed Logit	Mixed Logit
McFadden's R^2	0.09	0.48	0.46	0.55

Notes: standard errors are in parenthesis. All parameters are significant at the 1% level.

Firstly, I discuss the results under the assumption of taste homogeneity. Column (1) and (2) show a positive price coefficient $\bar{\beta}$ ranging from 0.82 to 0.45. Which indicates that households dislike higher prices.¹⁹ Taste for sugar $\bar{\delta}$ and bottle size $\bar{\pi}$ are also positive and below one. This means that the average household prefers sugary to diet colas and likes larger bottles. The estimate for the rarity coefficient $\bar{\theta}$ instead indicates that households dislike colas that are rarely on shelves, once they are available. This may suggest that a constant exposure on supermarket shelves tends to increase product

¹⁹ In Eq. 4.1 price enters the utility as a negative term. Hence, a positive coefficient means a lower utility if price increases.

appreciation. The coefficient for folder $\bar{\omega}$ is slightly negative in the *static* specification, while it is slightly positive in the *dynamic* one. The rebate coefficient $\bar{\varphi}$ is positive in both specifications but close to zero. This indicates that both quantity rebates and promotions in the weekly customer folder increase the product purchase probability but only to a small extent. Interestingly, the state dependence coefficient $\bar{\gamma}$ is strongly positive. This finding suggests that having purchased a cola in the previous shopping trip increases the probability of repurchasing the same cola on the next shopping occasion. Furthermore, the lagged choice variable is a significant predictor of the model. The inclusion of the lagged choice variable, in fact, increases the pseudo R^2 from 0,09 to 0,48. Hence, showing that under the assumption of homogeneous tastes, the *dynamic* logit specification is superior to the *static* one.

Accounting for unobserved taste heterogeneity, however, delivers a complete picture of households' preferences. As shown in column (3) and (4), all standard deviations of the randomly distributed taste coefficients are large and significant. This result rejects the Logit assumption of homogeneous taste coefficients, indicating that households do have heterogeneous preferences for colas. The *dynamic* Mixed Logit model fits the data better than any other specification, with a pseudo R^2 equal to 0.55. This is also confirmed by a likelihood test ratio between this model and each of the other three specifications. Thus suggesting that households exhibit both state dependence and taste heterogeneity. The comparison between column (3) and (4) shows that the omission of the state dependence variable in the Mixed Logit specification leads to an overestimation of the unobserved taste heterogeneity across households. The standard deviations of the randomly distributed taste coefficients in column (3) are indeed larger (about one third) than those displayed in column (4). Moreover, price sensitivity decreases considerably once accounting for state dependence. Hence, ignoring such behavior would overestimate the direct impact of prices on demand.

The *dynamic* model with unobserved taste heterogeneity is estimated using 200 Halton sequences. Increasing the number of Halton sequences delivers stable parameter estimates. The results of these computations are shown in table A.2 in the appendix. The price parameter of this model, which was assumed to be distributed as truncated (at zero) normal, has a mean of 2.53 and a median equal to 1.53.²⁰ The estimate of its standard deviation σ_β indicates that households are quite heterogeneous in terms of price sensitivity. The 12% of households have a price coefficient below 1 and, therefore, can be considered as relatively insensitive to prices. On the other hand, 13% percent of households have a price coefficient higher than 8, and thus can be considered as extremely sensitive to prices. This indicates that households are, on average, more price-sensitive compared to the estimates of the Logit

²⁰ The model was also estimated assuming either normal or lognormal distribution for the price coefficient. Yet, in the first case, a quite large part of household had a negative price parameter. While in case of log-normality, the estimates were not stable when increasing the number of random draws.

specification. Hence, assuming the price coefficient to be homogeneous across households leads to an underestimation of the average price sensitivity. The mean of the sugary parameter $\bar{\delta}$ is quite similar to the Logit specification, confirming that on average households like sugary colas. Yet, the estimate of its standard deviation σ_{δ} also highlights that taste for sugar is quite heterogeneous in the population. Similar results are found in preferences for bottle size, which are also quite heterogeneous across households. Most people prefer small bottles, but a large part likes colas in larger formats. Interestingly, the state dependence coefficient $\bar{\gamma}$ displayed in column (4) drops considerably compared to the one displayed in column (2). However, this remains positive and significant, decreasing from 3.03 to 1.85. This is because the Mixed Logit controls for persistent taste heterogeneity. While in the Logit specification, this can be empirically confounded with a spurious state dependence. Therefore, the Mixed Logit estimate of $\bar{\gamma}$ provides more robust evidence of state dependence in cola choice. Suggesting that part of the state dependence found with the Logit specification was instead due to unobserved taste heterogeneity across households.

5.2 Demand curves

The parameter estimates allow computing the impact of price changes on cola demand. Table 5 displays the matrix of own and cross-price elasticities of demand using the *dynamic* Mixed Logit model (computed as in eq. 4.2.3). As the model does not include any outside option, these elasticities indicate the percentage change in the average purchase probability conditional on purchasing a cola for a 1% change in price. In other words, they provide the measure in which prices lead households to substitute from one cola to another, regardless of how much cola they buy. Interestingly, the impact of a change in price on the average purchase probability tends to increase in the long-run due to state dependence in cola choice. Households need time to adjust their demand to a price change and reach the new equilibrium in their purchase probabilities. The average price elasticity of demand across all colas after the first price iteration is equal to 1.74.²¹ This is computed by simulating for each product the impact of a 1% increase in its price and keeping the household past purchase probabilities fixed at their level before the price increase. This measure can be interpreted as the short-run price elasticity of demand. Which considers the sole initial price effect and ignores the dynamics that will arise due to a change in the state dependence variable. Once the new purchase probabilities under the new price vector are reiterated in place of the old ones, household demand will continue to adjust until reaching the new equilibrium shares. At the end of this process, the long-run (or final) average price elasticity of demand across all products is equal to 2.04. This elasticity is larger than the short-run price elasticity since an initial price increase leads to a reduction in the current choice probability, which in turn reduces the purchase probability in the next shopping occasion.

²¹ This price elasticity is computed as a weighted average of all colas by using the average purchase probability as weight.

Table 5: Own and cross-price elasticities of demand (*Dynamic Mixed Logit*)

	A.1	A.2	A.3	B.1	B.2	B.3	C.1	C.2	D.1	D.2	E.1	E.2	F.1	G.1
A.1	-1.35	0.06	0.25	0.59	0.10	0.05	0.01	0.02	0.02	0.05	0.09	0.03	0.02	0.07
A.2	0.32	-1.59	0.71	0.08	0.25	0.08	0.01	0.16	0.02	0.02	0.28	0.07	0.07	0.21
A.3	0.26	0.14	-2.58	0.07	0.11	0.03	0.00	0.04	0.01	0.01	0.06	0.02	0.01	0.06
B.1	0.30	0.02	0.04	-1.59	0.13	0.03	0.01	0.03	0.03	0.08	0.02	0.04	0.04	0.02
B.2	0.36	0.17	0.38	0.90	-2.13	0.03	0.04	0.40	0.07	0.12	0.09	0.24	0.31	0.08
B.3	0.46	0.14	0.16	0.15	0.06	-1.87	0.27	0.04	0.33	0.42	0.63	0.16	0.18	0.43
C.1	0.05	0.02	0.02	0.14	0.06	0.11	-2.02	0.04	0.26	0.34	0.02	0.13	0.15	0.04
C.2	0.08	0.12	0.15	0.19	0.42	0.03	0.04	-1.63	0.05	0.06	0.03	0.14	0.16	0.05
D.1	0.14	0.03	0.04	0.40	0.17	0.28	0.61	0.11	-1.99	0.97	0.12	0.38	0.46	0.11
D.2	0.04	0.02	0.01	0.10	0.03	0.04	0.05	0.01	0.07	-3.48	0.01	0.04	0.03	0.02
E.1	1.34	0.72	1.06	0.42	0.41	0.72	0.16	0.22	0.21	0.30	-2.44	0.30	0.38	0.84
E.2	0.11	0.05	0.07	0.28	0.24	0.06	0.11	0.13	0.15	0.21	0.05	-2.29	0.23	0.06
F.1	0.32	0.14	0.20	0.91	0.72	0.15	0.35	0.40	0.44	0.63	0.24	0.59	-2.45	0.17
G.1	0.11	0.07	0.09	0.03	0.04	0.07	0.01	0.02	0.02	0.03	0.07	0.03	0.02	-2.22

Notes: own-price elasticities are in bold characters. Elasticities are computed at the sample average across households.

The Mixed Logit model allows price elasticities to depend on the distribution of taste and price sensitivity across households, with similar products being closer substitutes to each other. An easy way to capture this property is to look at the example displayed in Table 6. This table shows the percentage change in demand for different colas following a 10% increase in the price of Brand A.3, which is a bottle of sugary cola of 2 liters. As the price of this cola increases, households substitute for similar products in terms of sugar content and bottle size. A 10% increase in price for Brand A.3 leads to a reduction in its market share by 13.66%. Most of these consumers then switch other sugary and large colas (2 liters). Their market shares increase by 2.10% and 2.33%, respectively, while those for diet and small colas (1 liter) increase only to a smaller extent.

Table 6: Example of Substitution patterns for Brand A.3 (10% price increase)

	Brand A.3	Sugary	Diet	Large (2L)	Small (1L)
Δ market share	-13.66%	+2.10%	+0.77%	+2.33%	+0.28%

Notes: The table reports the change in the average purchase probability of sugary and diet colas following a 10% increase in the price of Brand A.3., which is a sugary cola of 2 liters.

From the parameter estimates, it is possible to draw the demand curve for each cola in the market. Interestingly, as households have a different price elasticity of demand, demand curves are not bounded to be linear in price. This exercise can therefore give an approximation of the curvature of demand in terms of average purchase probability. The estimates of the demand curves for each product are shown in section A.3 of the appendix. For most products, demand curves are highly convex. As price increases, price elasticity of demand decreases. They are thus indicating that as price goes up, only the less price-sensitive households keep on purchasing the same product. The demand curves for the cheapest colas (D.1 and D.2), however, are much less convex than other brands. This is probably because households purchasing these colas are the most price-sensitive. Therefore, they are much more likely to reduce their demand for higher price levels.

Importantly, these results cannot be directly extended to the shape of the aggregate demand for each product. Here the focus is on the average purchase probability conditional on buying a cola. It might be possible that for very high prices households either reduce the number of bottles purchased or decide not to purchase a cola. As a result, aggregate demand curves should be less convex than those displayed in section A.3. They might be similar only in the limit case where both the intensive margin and no purchase effects are equal to zero. The demand curves presented here can be interpreted as a lower bound in terms of price sensitivity of the aggregate demand for colas. They are thus providing the most conservative setting in order to evaluate the impact of price-based policies on the demand for sugary colas.

5.3 Tax policy simulations

The model estimates allow performing tax policy simulations to assess the impact of sugar taxes on the households' probability of purchasing sugary colas. Table 7 shows the outcome of different tax policy simulations on the average purchase probability of sugary and diet colas. The simulations assume full tax pass-through (i.e., 100%), and consider both ad-valorem and excise taxes. An ad-valorem tax on sugary colas is a tax levied as a percentage of their market price. Under this tax scheme, cheap colas will be taxed less in absolute value than more expensive brands. This type of taxation was recently implemented in various countries, although sometimes it affects all sodas regardless of their sugar content. For instance, in 2015, the Barbados implemented a 10% ad-valorem tax on SSB. While in Chile, since 2014, drinks with high sugar content are taxed by 18%. The main advantage of ad-valorem taxes is their ease of implementation, as the amount of tax levied varies only as a function of prices. However, their shortcoming is that they are not related to the amount of sugar contained in each product. Hence, they may be less effective both in reducing sugar intake among consumers and in incentivizing firms in reformulating their products. An alternative to this tax scheme is the implementation of excise taxes on sugar. These are taxes levied on the amount of sugar

contained in each product so that prices increase proportionally to their sugar content. This type of tax was recently implemented in few American localities, such as Berkley, in 2015, and Philadelphia in 2017. In the UK, since April 2018, sodas with more than 8 grams of sugar per 100 millilitres of products are taxed by £0.24 per litre. Another interesting example is Ecuador, where on SSB with less than 25 grams of sugar per liter is levied an ad-valorem tax of 10%. However, SSB with higher sugar content are instead taxed by an equivalent of \$0.18 for every 100 grams of sugar.

Table 7: Sugar tax simulations

Δ market share	Timing	Ad-valorem tax		Excise tax	
		20%	30%	0,20€ (/100gr)	0,30€ (/100gr)
Sugary colas	Δ short run	-11.70%	-16.76%	-13.38%	-19.42%
	Δ long run	-13.13%	-18.69%	-14.47%	-21.29%
Diet colas	Δ short run	+11.37%	+16.28%	+13%	+18.87%
	Δ long run	+12.76%	+18.16%	+14.73%	+20.69%

Notes: The table reports the change in the average purchase probability of sugary and diet colas under different sugar tax schemes. The simulations are performed at the sample average across households.

The results of the tax policy simulations displayed in Table 7 are expressed in terms of the percentage change in the market share (i.e., average purchase probability) for the two different groups of cola. As sugary colas are the only type of product being taxed, their market share decreases considerably after any tax is implemented. The market share for diet colas instead increases, as households then substitute towards healthier varieties. Table 7 also shows the different impacts of these policies over time. Because of state dependence, the immediate impact of a price change on household demand tends to be smaller than in the long-run. Although the tax mostly affects demand already in the short-run, this finding suggests that the public health benefits of a sugar tax tend to increase over time. The difference between the short-run versus the long-run demand response to the tax, however, differs across tax schemes. This difference tends to increase with the initial tax change. The more the current purchase probability decreases due to a tax hike, the more the future purchase probability decreases due to a change in the state dependence variable. As demand curves are quite convex, this difference is not strictly increasing in the tax rate. At a certain tax level, the difference between short-run and long-run demand response will not increase with a higher tax. This occurs when the tax reaches the inelastic part of the demand so that any further tax increase does not marginally affect market shares. A typical concern with sugar taxes is that they may incentivise households to substitute for cheaper sugary brands. In order to study this possibility, Figures 2.a and 2.b below compare the effectiveness

of ad-valorem versus excise taxes in reducing cola demand across different product types. The y-axes indicate the different levels of taxation, while the x-axes represent the market shares. The blue curve is the demand for the cheapest sugary cola in the market (brand D.1) as a function of the tax. The red curve is instead the demand for all other sugary colas. Figure 2.a shows that the demand for the cheapest sugary cola is inelastic to ad-valorem taxes, while Figure 2.b shows that this is quite elastic to any level of excise tax. The demand for more expensive sugary colas instead declines similarly under both tax schemes.

Figure 2.a: ad-valorem tax

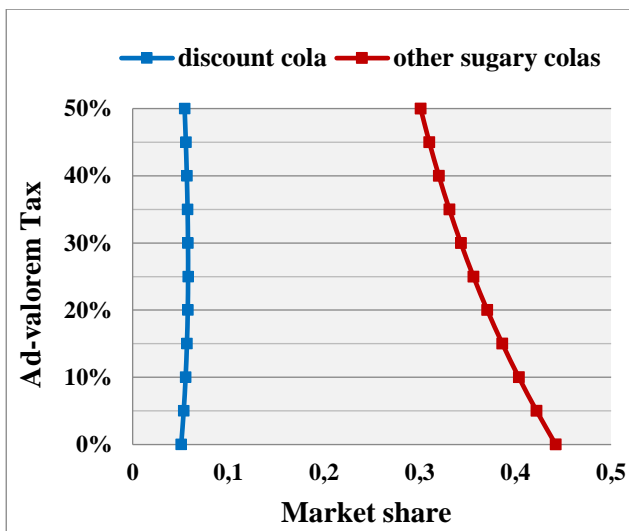
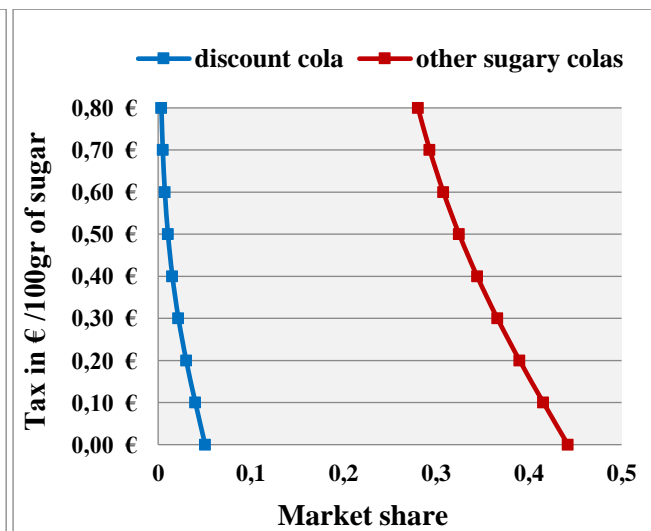


Figure 2.b: excise tax



Notes: Figures 2.a and 2.b show the impact of ad-valorem and excise taxes on the demand for sugary colas. The y-axes indicate the different levels of taxation. The x-axes indicate the market shares (i.e., average purchase probability). The blue curve represents the cheapest sugary cola in the market (brand D.1). The red curve represents all other sugary colas. The simulations are performed at the sample average across households.

This difference in demand response is originated from the measure in which these two types of taxes affect the price of cheap brands. Ad-valorem taxes tend to increase the price of cheap colas by a smaller amount than specific taxes on sugar, thus leading to the following two effects. First, the least price-sensitive households purchasing the cheap brand are not concerned by such a small price increase and hence keep on purchasing this cheap product. Second, those households preferring expensive brands, having a strong taste for sugar, and high price-sensitivity are likely to substitute their preferred cola with another sugary cola but of a cheaper brand. The sum of these two effects makes the demand for the cheapest sugary cola inelastic to any level of ad-valorem tax, as shown in Figure 2.a above. When a tax levied on sugar content, these two effects are reduced since the price increase for cheaper colas tends to be larger than with an ad-valorem tax. Furthermore, as all sugary colas have similar sugar content, a specific tax on sugar implies a similar price hike across sugary colas in absolute terms. This feature increases the relative price between cheaper and expensive sugary colas, hence disincentivizing the substitution towards cheaper brands. As a result, these

findings suggest that specific taxes on sugar should be preferred to ad-valorem taxes to avoid a substitution towards cheaper sugary brands.

The simulations above assume that the pass-through of these taxes to prices is equal to 100%. That is, any tax would increase cola prices by the same amount. Bonnet & Requillart (2013), in line with previous findings, find that while excise taxes are usually over-shifted to soft drink prices, ad-valorem taxes are instead often under-shifted. This result reinforces the effect of above, thus supporting the view that specific taxes on sugar should be preferred on public health grounds. Furthermore, Ally et al. (2014) show that alcohol taxes are typically under-shifted to the price of cheaper alcoholic beverages. Suggesting that the substitution towards cheaper sugary products with ad-valorem taxes may be larger than estimated. Interestingly, a possible explanation of the lower tax shifting for cheaper products can be found in the shape of their demand curves. In the context of imperfectly competitive markets, tax pass-through should increase with the convexity of the demand curve (Weyl & Fabinger, 2013). The estimated demand curves for the two cheapest colas (D.1 and D.2) display a much lower degree of convexity compared to more expensive products (see in the appendix). This finding implies that ad-valorem taxes could be shifted more to the price of expensive brands than to the price of cheaper colas.

5.4 Model validation

A possible concern is that the estimates of taste distributions over-fit the subsample in which the model is estimated, thus providing results that are not generalizable to the rest of households. The comparison of the Logit models in Table A.1 suggests that the Logit parameter estimates do not change much with the reduction in sample size. However, this cannot be confirmed for the Mixed Logit, as it is practically unfeasible to estimate this model on the entire sample. One way to validate the results of this model is to compare its in-sample predictive performance against its out-of-sample one. Although cola choice prediction is not the ultimate scope of this study, testing the predictive performance of the model can serve as an indicator of its relevance in explaining household choices. Comparing the predictive performance within and outside the estimation sample can also indicate whether the model is still relevant in explaining the choices of households that were not considered in the estimation procedure. Furthermore, as demand curves represent the outcome of predicting the average purchase probability under different price levels, this exercise can give an idea of how reliable these counterfactual simulations are.

The predictive performance of the model is computed in terms of the average percentage error in predicting actual market shares. More formally, the average of the correct predicted shares with the Mixed Logit estimates is given by the following formula:

$$1 - \left[\frac{1}{J} \sum_j \left(\frac{|\tilde{s}_j - s_j|}{s_j} \right) \right].$$

Where s_j and \tilde{s}_j are the actual market share and the predicted market share of cola j , respectively.²² While J is the total number of cola products. The expression in square brackets is equivalent to the percentage prediction error of the model on average. The model is fitted to four distinct samples, where one is the original estimation sample, and the other three are other randomly selected samples of 10,000 households. Fitting the model to its estimation sample, correctly predicts the 79.29% of market shares on average. Interestingly, fitting the model to the other random samples delivers very similar results, with an average of 78.86% correct share predictions. This result indicates that the model predicts reasonably well actual market shares, even when considering different households and under different market scenarios. Thus suggesting that the parameter estimates do not over-fit the estimation sample.

6. Observed taste heterogeneity

The results shown in the previous section suggest that households have very heterogeneous preferences for colas. Thus implying that their behavioral response to sugar taxes can be highly heterogeneous. Accounting for observed taste heterogeneity, however, is fundamental to understand the impact of these policies among those households overconsuming sugary products. Similarly, policymakers may also be interested in understanding how the tax burden and health benefits are distributed across families with different socioeconomic statuses. In order to address these questions, the utility specification outlined in equation 4.1 is extended by allowing price sensitivity, taste for sugar and bottle size to vary with observable households' characteristics. Formally, their respective coefficients β_i , δ_i and π_i are replaced in the utility specification 4.1 by:

$$\beta_i = \bar{\beta} + \sum_H \beta_H (Demo_H)_i + \beta_S (Sugar_{share})_{it} + \zeta_i \sigma_\beta ;$$

$$\delta_i = \bar{\delta} + \sum_H \delta_H (Demo_H)_i + \delta_S (Sugar_{share})_{it} + \eta_i \sigma_\delta ;$$

$$\pi_i = \bar{\pi} + \sum_H \pi_H (Demo_H)_i + \eta_i \sigma_\pi .$$

²² Here market shares are considered as the share of shopping occasions in which households choose a given cola. Thus, this is equivalent to $s_j = \frac{\sum_i \sum_t y_{ijt}}{N}$. Where N is the total number of shopping trips in the sample and y_{ijt} is a variable equal to 1 if household i bought cola j during shopping trip t .

Where the variables $Demo_H$ are the household demographics described in the data section. Which include income, number of household members, a dummy variable indicating whether they have kids, educational level, and age of the head of household. The coefficients β_H , δ_H , and π_H indicate how price sensitivity, taste for sugar and preference for bottle's size change with the observed household demographics. In addition to these household characteristics, price sensitivity β_i and taste for sugar δ_i are allowed to vary with the variable $Sugar_{share}$. Which is the share of sugary dense products in the overall food shopping basket of household i . Their associated coefficients β_S and δ_S indicate how price sensitivity and taste for sugar change with the level of consumption of sugary products. These two coefficients are crucial in order to evaluate the effectiveness of sugar taxes on public health and internality grounds. If $\delta_S > 0$, heavy sugar consumers tend to prefer sugary to diet colas, and therefore, there is a public health rationale in taxing these products. Conversely, if $\delta_S \leq 0$, taxing sugary colas may not lead to significant public health benefits, as these products are not particularly appealing to those overconsuming sugar. Sign and magnitude of β_S indicate how effective this policy will be in reducing the demand for sugary colas among the targeted population of heavy sugar consumers. With $\beta_S > 0$, price sensitivity is higher among those overconsuming sugar. Hence, a sugar tax would be quite effective in reducing their demand for sugary colas and so their sugar intake. The larger β_S , the higher the effectiveness of this policy in correcting the household behavior. The variables $\bar{\beta}$, $\bar{\delta}$ and $\bar{\pi}$ are the intercepts of the taste coefficients and therefore represent the part of preferences that do not change with observable demographics. While the terms ζ_i and η_i are random variables, which are distributed as follows: $\zeta_i \sim TN_{[-\mu_\beta, +\infty]}(0,1)$; $\eta_i \sim N(0,1)$. With μ_β representing the mean of β_i .²³

This model is estimated on a random subsample of households spending at least 80% of their food-at-home budget in this retail chain. This approach allows circumventing the issue of not capturing sugar eating habits of those households purchasing food in different supermarket chains. The total number of households corresponding to this characteristic is 130,514. The Mixed Logit with observed group heterogeneity is estimated on a random subsample of 10,000 households, for a total of 358,225 shopping trips. The result of this estimation is shown in Table 8 below. Table A.4 in the appendix shows the parameter estimates of a simple Logit version of this model, which is instead estimated on the entire sample of assiduous costumers. Such a model does not allow for unobserved taste heterogeneity and hence it assumes that $\sigma_\beta = \sigma_\delta = \sigma_\pi = 0$.

The results displayed in Table 8 suggest that price sensitivity, taste for sugar, and preference for bottle's size do indeed vary with observable household characteristics. All coefficients of the interaction variables between taste parameters and household demographics are significant at the 1%

²³ This ensures that $\beta_i \geq 0 \forall i$.

level. The sign of these coefficients indicates how demographics are correlated with taste parameters. As many demographics are included, this has to be interpreted as a partial correlation between the household characteristic and taste coefficients. That is, it is a measure of how taste parameters vary with a given demographic variable while holding all other household characteristics as fixed.

Table 8: Parameter estimates of *Dynamic Mixed Logit* with observed group heterogeneity

Parameter	Estimate	Demographics	Estimate		
			Price	Sugary	Size
Price $\bar{\beta}$	33.268 (0.793)	log(Income)	-3.242 (0.079)	-1.080 (0.035)	0.463 (0.057)
Price σ_{β}	4.253 (0.032)	Education	-0.193 (0.077)	-0.330 (0.031)	-0.786 (0.052)
Sugary $\bar{\delta}$	10.222 (0.351)	Sugar share	-8.222 (0.312)	6.273 (0.152)	
Sugary σ_{δ}	2.347 (0.014)	Household's size	0.254 (0.012)	0.119 (0.006)	0.263 (0.009)
Size $\bar{\pi}$	-4.911 (0.572)	Kids	-0.343 (0.038)	0.006 (0.017)	-0.275 (0.027)
Size σ_{π}	2.835 (0.014)	Age>70	-0.517 (0.080)	0.384 (0.034)	-1.534 (0.054)
Rare $\bar{\theta}$	-0.972 (0.011)	70>Age>58	0.659 (0.066)	0.173 (0.026)	-1.327 (0.044)
Folder $\bar{\omega}$	-0.036 (0.019)	58>Age>46	1.091 (0.055)	0.475 (0.021)	-0.935 (0.035)
Rebate $\bar{\varphi}$	0.173 (0.018)	46>Age>34	0.586 (0.053)	0.101 (0.020)	-0.529 (0.034)
State dependence $\bar{\gamma}$	1.930 (0.004)				
Brand fixed effects		yes			
N° of Observations		358,225			
McFadden's R^2		0.57			

Notes: standard errors are in parenthesis. The first and second columns show intercepts and standard deviations of random taste parameters. Together with the estimates of the fixed taste parameters. From the third to the last column, the table displays the estimates of the interactions between price, sugary and size coefficients with observed demographics. Age coefficients are computed with respect to families with a head of household below 34 years old. All parameters are significant at the 1% level except for the interaction between kids and the taste for sugar.

The parameters displayed in Table 8 indicate that taste coefficients do vary with household demographics. At the same time, their estimated standard deviations also suggest that a substantial part of such heterogeneity is still not explained by differences in observed household characteristics. This finding suggests that a Mixed Logit model should be preferred to a simple Logit, as household

preferences also vary randomly across the population. A possible concern for the Mixed Logit model with observed household demographics is that the reduction in sample size can undermine the correct identification of observed taste heterogeneity across households. However, this does not seem to be the case for these estimates. Most of the interaction coefficients are significant at the 1% level. Furthermore, the simple Logit version of the model, which was estimated on the entire sample of assiduous costumers, shows very similar results in terms of observed taste heterogeneity across households.

Price sensitivity is decreasing with income and educational level, and it is lower among households having kids, while it increases with family size. These results are somehow expected, as poorer, larger, and low-educated households tend to be more resource-constrained and thus more sensitive to prices. For households having similar characteristics, old families (with the head of household above 70 years old) tend to be less price-sensitive, followed by the youngest (with the head of household below 34 years old). Interestingly, the estimates suggest that price sensitivity decreases with the household share of sugary-dense products in the overall shopping basket. This means that households consuming more sugar are less sensitive to changes in cola price compared to healthier households.

Taste for sugary colas also tends to vary with observable household characteristics. Poorer, low-educated, and larger households prefer sugary to diet colas. The relationship between age and preference for sugary colas is instead ambiguous. Oldest families and those with the head of household between 46 and 58 years old are more likely to prefer a sugary cola compared to those in other age groups. Not surprisingly, households that usually consume more sugar have stronger preferences for sugary colas. The higher the total consumption of sugar (net of that one consumed in colas), the higher the probability of consuming a sugary cola. This result suggests that households perceive sugar in colas as a sort of complement to the overall sugar consumption. A household in the lower quartile of the sugar consumption distribution (i.e., the share of sugary-dense products) has a probability of 41.25% of choosing a sugary cola. In contrast, a household in the top quartile of the sugar consumption distribution has a 50.90% probability of choosing a sugary cola.

Table 8 also indicates that higher-income households prefer larger bottle sizes, while highly educated families tend to prefer smaller bottles. Larger families are more likely to go for large bottles, although this effect is smaller if the household has kids. Contrary to both price and sugary parameters, the relationship between age and size coefficients is somehow linear. The older the family, the higher the preference for smaller bottle sizes.

6.1 The impact of taxation across different household types

The estimates of the *dynamic* Mixed Logit model with observed group heterogeneity allows evaluating the impact of sugar taxes on cola demand across households with different demographics and sugar eating habits. The results of this model highlight two important empirical facts. First, heavy sugar consumers tend to prefer sugary to diet colas. This indicates that there is a public health rationale in taxing these products, as their consumption is more prevalent among those households that have higher health risks. Second, these households are not more sensitive to price changes than others. Thus suggesting that sugar taxes would not be more effective in reducing cola demand among the targeted population of heavy sugar consumers.

However, the model estimates also suggest that poorer households are likely to benefit the most in terms of public health. This because they have both a greater preference for sugary colas and higher price sensitivity compared to richer households. Therefore, a sugar tax would be quite effective in inducing these households to switch from sugary to diet colas. The fact that poor households are very price-sensitive also helps to reduce their tax burden in monetary terms. Which tends to be larger than the one of richer households as they are more likely to purchase sugary colas. In a broader economic welfare framework, these findings can suggest that a sugar tax on colas may be both efficient and progressive (see Griffith et al., 2018; Allcott et al., 2019). A necessary condition, however, is that sugar externalities exist and decrease with the level of income.

Table 9 below shows the potential impact of an excise tax of € 0.30 per 100 grams of sugar across different household groups. Lower-income households, who generally prefer sugary colas, are more likely to switch to a diet cola since they are more price-sensitive than higher-income households. Conversely, high-income households, which generally tend to dislike sugary colas, reduce their demand by less than poorer households. In particular, with the implementation of this tax, the probability of choosing a sugary cola decreases by -21.32% for low-income households against the -17.25% for high-income households. Moreover, this tax is less effective in reducing sugary cola demand among heavy-sugar consumers than among healthier households. This because habitual consumers of sugar are less sensitive to cola prices, even though they have a higher demand for sugary products. The estimates in Table 9 shows that the implementation of the tax reduces the demand for sugary colas by -16.41% among heavy sugar consumers, while the reduction amounts to -22.24% among the group of healthiest households. The impact of this tax, however, is quite similar across households with different educational levels. In terms of tax incidence, heavy-sugar consumers and poor households are the groups bearing this tax the most. This effect, however, is smaller for poor households, as they are more sensitive to prices and so they switch more quickly to untaxed diet products.

Table 9: Impact of €0.30 excise tax per 100gr sugar across household types

Household types	Demand for sugary colas (average purchase probability)		
	Before	After	$\Delta\%$
Low Income (Q1)	48.75%	38.08%	-21.32%
High Income (Q4)	42.66%	35.30%	-17.25%
Low Sugar Consumption (Q1)	41.25%	32.08%	-22.24%
High Sugar Consumption (Q4)	50.90%	42.55%	-16.41%
Low Education (Q1)	46.41%	37.51%	-19.18%
High Education (Q4)	43.98%	35.44%	-19.42%

Notes: in parenthesis, there are the quartiles of the distribution for each household type. Q1 stands for the first quartile, while Q4 is the last quartile. The column “before” and “after” indicate the market share of sugary colas before and after the tax simulation.

In section 5.3, it was shown that excise taxes should be preferred to ad-valorem taxes because they prevent households from substituting to cheaper sugary colas. Accounting for observed taste heterogeneity, however, allows investigating the demographic composition of these households. In particular, it is interesting to understand whether these households are part of the targeted population of heavy sugar consumers. This finding would undermine the effectiveness of ad-valorem taxes on both public health and internalities motives. It is also crucial to check whether these are low-income households, which would then suggest that ad-valorem taxes tend to be more regressive than excise taxes.

These questions are addressed by simulating the impact of a 30% ad-valorem tax on the purchase probability of two distinct categories of consumers. That are: (i) low-income households with a high sugar consumption (i.e., those in both the first and last quartile of the income and sugar consumption distribution, respectively); (ii) high-income households with a low sugar consumption (i.e., those in both the last and first quartile of the income and sugar consumption distribution, respectively). Table 10 shows the results of this simulation by presenting the change in the purchase probability of both the cheapest and the other more expensive sugary colas for these two categories of households. The simulation of a 30% ad-valorem tax highlights important differences in the behavioral response across different household types. Low-income households with high sugar consumption are much more likely to substitute for a cheaper sugary cola in response to the ad-valorem tax. In particular, their demand for the cheapest sugary cola increases by 15.43%, while this increase amounts to only 4.11% for high-income households with low sugar consumption.

Table 10: Impact of 30% ad-valorem tax across household types

Household types	Cheapest sugary cola			Other sugary colas			All sugary colas		
	Before	After	$\Delta\%$	Before	After	$\Delta\%$	Before	After	$\Delta\%$
Low Income & High Sugar Consumption	4,66%	5,38%	+15,43%	49,45%	40,43%	-18,23%	54,10%	45,81%	-15,34%
High Income & Low Sugar Consumption	2,76%	2,88%	+4,11%	35,81%	29,01%	-18,98%	38,57%	31,89%	-17,33%

Notes: The column “before” and “after” indicate the market share of each type of cola before and after the tax simulation. The cheapest sugary cola is brand D.1, while the other sugary colas are all the remaining sugary brands. The column “All sugary colas” refers to the overall market share for sugary colas, which is the sum of the cheapest and the other sugary brands.

This substitution effect makes the ad-valorem tax less effective in reducing demand for sugary colas, especially among the targeted population of heavy sugar consumers and low-income households. On public health grounds, this finding reinforces the hypothesis that excise taxes should be preferred to ad-valorem taxes in order to reduce sugar consumption among the targeted population. Furthermore, since the consumption of cheap sugary colas is more prevalent among the poor, this also suggests that ad-valorem taxes can be more regressive than excise taxes. With an excise tax, the demand for cheap sugary colas would decline rapidly since the incentive to substitute for a diet cola is higher than with an ad-valorem tax (see Figure 2.b). Although the demand for cheaper sugary colas is more prevalent among the poor, this larger substitution towards diet colas reduces their tax burden in monetary terms and makes this tax also more effective in reducing their sugar intake.

7. Robustness Checks

The empirical evidence presented so far suggests that households have heterogeneous tastes and positive state dependence in product choice. In the previous sections, it was also discussed the importance of disentangling these two behaviors and their implications in terms of sugar tax design. This section of the paper presents the results of various robustness checks that were performed in order to confirm these findings and rule out other possible explanations for the observed inertia in cola choice. The two main hypotheses to be rejected are that the estimated state dependence is the result of either a misspecification of taste heterogeneity distribution or autocorrelation in the error terms.

If the distribution of taste heterogeneity is not correctly specified, the coefficient on the lagged choice variable may still be positive because it captures some persistent propensity of the household to make a given choice. An example of this phenomenon can be seen in table 4. When accounting for taste heterogeneity through random coefficients, the state dependence parameter dropped considerably compared to the logit specification (see column 2 against column 4). Hence, it should be checked

whether accounting for a more flexible distribution of taste heterogeneity across the population would still deliver a positive state dependence coefficient. Following Dubé et al. (2010), this is tested by randomly reshuffling the order of the households' shopping trips and then checking whether the lagged choice coefficient is still positive and significant. The rationale behind this test is the following. If the estimated state dependence is due to some misspecification of taste heterogeneity, then changing the order of households' shopping trips should still deliver a positive state dependence coefficient. In contrast, if taste heterogeneity is correctly specified, only the recent shopping experience should influence the current choice. Hence, changing the order of the shopping trips should deliver a coefficient on the lagged choice variable that is close or equal to zero.

Table 11: Robustness Checks for Misspecified Heterogeneity

Random Reshuffling		
	<i>Unobserved Heterogeneity</i>	<i>Unobserved Heterogeneity with HH Demographics</i>
<i>actual</i>	2.121 (0.007)	2.167 (0.006)
<i>reshuffled</i>	0.339 (0.008)	0.252 (0.007)

Notes: standard errors are in parenthesis. All coefficients are significant at the 1% level.

The result of this analysis is displayed in Table 11 above. The row *actual* shows the coefficient on the lagged choice variable when the reshuffled lagged choice is equal to the actual lagged choice. The row *reshuffled* shows the lagged choice coefficient when the reshuffled lagged choice is not equal to the actual lagged choice.²⁴ The test is performed under the two different specifications of household taste heterogeneity. The first column shows the results of this test under unobserved taste heterogeneity. The second column shows the results of this test under both observed and unobserved taste heterogeneity.

The reported coefficients suggest that the hypothesis that state dependence is due to a misspecification of taste heterogeneity should be rejected. The order of shopping trips does matter in predicting the current household choice correctly. The coefficient of the actual lagged choice is positive and significant, while that of the reshuffled lagged choice is close to zero. The fact that the latter is positive and significant, however, does suggest the existence of some misspecification in the distribution of

²⁴ By randomly reshuffling the order of the shopping trips, around 50% of reshuffled lagged choices are still equal to the actual past choice. Although here it is presented the result of one random reshuffling only, this exercise was performed many times with different reshuffled sequences and always delivered similar results.

taste heterogeneity. However, this seems too small to explain the observed state dependence in cola choice. Accounting for a more flexible distribution of taste through household demographics reduces this misspecification sensibly. The coefficient on the reshuffled choice gets, in fact, closer to zero.

A model with true state dependence and model with autocorrelated error terms have somewhat different policy implications. If unobservable factors drive choice dynamics, governments cannot exploit state dependence to enhance the impact of sugar taxes in the long-run. Three different tests are performed in order to rule out this hypothesis. Their results are presented in Table 12, and each of them is discussed in turn below.

Table 12: Robustness Checks for Autocorrelation

Test for Autocorrelation					
<i>Lagged Prices</i>		<i>State Dependence with price Discount</i>		<i>Cola not Available in Past Choice Set</i>	
<i>price_{t-1}</i>	0.694*** (0.041)	<i>SD</i>	1.927*** (0.007)	<i>NA_{t-1}</i>	-0.413*** (0.032)
		<i>SD * D_{t-1}</i>	-0.020 (0.011)		

Notes: standard errors are in parenthesis. The *** indicates statistical significance at the 1% level. All coefficients are obtained by taking as baseline model the dynamic Mixed Logit without household demographics. Similar results are obtained by including household demographics.

The first test follows Chamberlain (1985). It is estimated a model where the lagged choice variable is substituted by lagged prices (i.e., the price vector faced by the household in the previous shopping trip). If choice inertia is due to autocorrelated error terms, then past prices should not influence the current choice probabilities. Conversely, in case of state dependence, lagged prices should negatively affect current choice probabilities as they can influence the household's state variable. This means that the higher is the lagged price of cola, the lower should be its current purchase probability.

The result of this test is shown in the first column of Table 12 above, where the lagged price coefficient is shown in the first column next to the row labelled *price_{t-1}*. The hypothesis of autocorrelated error terms is rejected as the lagged price coefficient is positive and significant (remember that price enters the utility specification as a negative term). Higher past prices reduce the current purchase probabilities, hence proving evidence of positive state dependence. The main issue with this test, however, is that lagged prices might proxy for households' price expectations (Chamberlain, 1985; Dubé et al., 2010). This means that lagged prices can potentially influence current household choices even in the absence of state dependence.

Following the suggestion of Dubé et al. (2010), it is performed a further test for autocorrelation in the error terms by checking at the impact of temporary price cuts on the measured state dependence. They show that in a model with autocorrelated error terms the estimated coefficient for state dependence should be smaller when a price discount initiated the past state. The idea behind this test is that the past error term is larger when the past choice is made at a regular price rather than at a price discount. Hence, if the lagged choice variable is capturing some autocorrelation in the error terms, then the state dependence coefficient must be larger when the past choice was made at a regular price. This is because the lagged choice proxies for both a large past and current random utility draw. This test is performed by adding to the utility specification an interaction term between the lagged choice variable and a dummy variable indicating whether the product was under discount in the previous shopping trip. A product was considered under discount if the price in the previous shopping trip was at least 5% lower than in the current shopping trip.

The result of this test is displayed in the second column of Table 12. The variable SD indicates the state dependence coefficient on the lagged choice variable. While the interaction term $SD * D_{t-1}$ measures how state dependence varies if the product was under discount in the previous shopping trip.²⁵ The hypothesis of autocorrelated error terms is also rejected by this test. The coefficient of the interaction variable does not capture any lower state dependence when a previous price discount initiated the past state.

The nature of the dataset allows proposing a further test for autocorrelated error terms. The lagged choice variable is substituted with a dummy variable indicating whether a product was not available in the previous shopping occasion. In a model with autocorrelated error terms, a product that was not available in the previous shopping trip should not have any impact on the current choice probability. In a model with true state dependence, however, this dummy variable can affect current choice probabilities as a change in the household's previous choice set can influence its state variable. In particular, the coefficient of this dummy variable should be negative and significant since the absence of a product in the previous choice set prevents this product from being purchased in the past, and so reduces its current choice probability. The third column of Table 12 shows the result of this test. The variable NA_{t-1} represents the coefficient of the dummy variable that indicates whether a product was not available in the previous shopping occasion. This coefficient is negative and significant. Hence, suggesting the existence of a true state dependence driving the households' choice dynamics.

The model is also re-estimated by including a set of time-varying brand intercepts to control for some possible unobservable factors that might be correlated with prices. This check controls notably for

²⁵ D_{t-1} is a dummy variable indicating whether the product was under discount in the previous shopping trip.

advertising campaigns, which are done nationally and are unobserved in the data. The results of this estimation can be found in Table A.5 in the appendix. As shown in this table, taste coefficients are very similar to the one displayed in Table 4 (column 4). Thus suggesting that time-varying unobservable factors, such as advertising, do not affect the results.

8. Conclusions

Sugar taxes are often considered as a possible tool to tackle excessive sugar consumption. As the issue of excessive sugar intake does not affect all consumers in the same way, accounting for heterogeneity in preferences is crucial to assess the impact of these policies among those overconsuming sugary products. This work estimates a dynamic multinomial Logit model of cola demand on a novel set of supermarket scanner data in order to study preference heterogeneity across households. A particular focus is given to disentangling persistent taste heterogeneity from other sources of choice inertia, such as state dependence in product choice. The model estimates allow evaluating the effectiveness of taxation in reducing demand for sugary colas across different household types.

The results of this analysis show that households have very heterogeneous preferences for sugary colas and have different price sensitivity. Heavy sugar consumers tend to prefer sugary to diet colas but are less sensitive to cola prices. This finding suggests that although taxing sugary colas can be justified on public health grounds, this policy would not have a greater impact on the targeted population. Poor households, however, are likely to experience larger health gains since they would be more reactive to sugar taxes. Their higher tax responsiveness also helps them reduce their tax burden in monetary terms.

The estimated demand model allows performing tax policy simulations that account for very flexible substitution patterns. These simulations show that specific taxes targeting sugar content should be preferred to ad-valorem taxes on sugary colas on both corrective and equity grounds. The main reason for this finding is that ad-valorem taxes would entail a larger substitution towards cheaper sugary brands and that this effect is mostly concentrated among low-income households and heavy sugar consumers.

Finally, this work finds robust evidence of state dependence in cola choice. This means that households need some time before adjusting their demand to price changes. Ignoring this behavior can lead to an overestimation of the degree of preference heterogeneity in the population. Because of state dependence, the impact of sugar taxes on household demand would tend to increase over time. The tax policy simulations show that sugar taxes would mostly affect demand already in the short-run, but their impact would also have a moderate increase with time.

References

- Allais, O., Bertail, P., & Nichèle, V. (2010). The effects of a fat tax on French households' purchases: a nutritional approach. *American Journal of Agricultural Economics*, 92(1), 228-245.
- Allcott, H., Lockwood, B. B., & Taubinsky, D. (2019). Regressive Sin Taxes, with an Application to the Optimal Soda Tax. *The Quarterly Journal of Economics*, 134(3), 1557–1626.
- Allenby, G. & Lenk, P. (1994). Modeling Household Purchase Behavior with Logistic Normal Regression. *Journal of the American Statistical Association*, 89(428), 1-14.
- Ally, A. K., Meng, Y., Chakraborty, R., Dobson, P. W., Seaton, J. S., Holmes, J., ... & Meier, P. S. (2014). Alcohol tax pass-through across the product and price range: do retailers treat cheap alcohol differently?. *Addiction*, 109(12), 1994-2002.
- Alvarado, M., Unwin, N., Sharp, S. J., Hambleton, I., Murphy, M. M., Samuels, T. A., Suhrcke, M. & Adams, J. (2019). Assessing the impact of the Barbados sugar-sweetened beverage tax on beverage sales: an observational study. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1), 13.
- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242-262.
- Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841-890.
- Bonnet, C., & Réquillart, V. (2013). Tax incidence with strategic firms in the soft drink market. *Journal of Public Economics*, 106, 77-88.
- Bonnet, C., & Réquillart, V. (2016). *The effects of taxation on the individual consumption of sugar-sweetened beverages* (No. 16-638). Toulouse School of Economics (TSE).
- Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In J. Heckman & B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data* (Econometric Society Monographs, 3-38). Cambridge University Press.
- Chouinard, H. H., Davis, D. E., LaFrance, J. T., & Perloff, J. M. (2007). Fat taxes: big money for small change. *In Forum for Health Economics & Policy*, 10(2).
- Coibion, O., Gorodnichenko, Y., & Hong, G. H. (2015). The cyclicality of sales, regular and effective prices: Business cycle and policy implications. *American Economic Review*, 105(3), 993-1029.
- Deaton, A., & Muellbauer, J. (1980). An almost ideal demand system. *The American economic review*, 70(3), 312-326.
- Dobbs, R., Sawers, C., Thompson, F., Manyika, J., Woetzel, J. R., Child, P., ... & Spatharou, A. (2014). *Overcoming obesity: an initial economic analysis*. McKinsey global institute.

- Dubé, J. P., Hitsch, G. J., & Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*, 41(3), 417-445.
- Dubois, P., Griffith, R., & O'Connell, M. (2017). *How well targeted are soda taxes?* (No. 17-868). Toulouse School of Economics (TSE).
- Einav, L., Leibtag, E., & Nevo, A. (2010). Recording discrepancies in Nielsen Homescan data: Are they present and do they matter?. *QME*, 8(2), 207-239.
- Erdem, T., Keane, M. P., & Sun, B. (2008). A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science*, 27(6), 1111-1125.
- Goeree, M. S. (2008). Limited information and advertising in the US personal computer industry. *Econometrica*, 76(5), 1017-1074.
- Griffith, R., O'Connell, M. & Smith, K. (2018). Corrective taxation and externalities of food consumption. *CESifo Economic Studies*, 64, 1–14.
- Gruber, J., & Köszegi, B. (2004). Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. *Journal of Public Economics*, 88(9-10), 1959-1987.
- Harding, M., & Lovenheim, M. (2017). The effect of prices on nutrition: comparing the impact of product-and nutrient-specific taxes. *Journal of health economics*, 53, 53-71.
- Heckman, J. (1978). Simple statistical models for discrete panel data developed and applied to tests of the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales de l'INSEE* 30/31, 227– 69.
- Hendel, I., & Nevo, A. (2006). Sales and consumer inventory. *The RAND Journal of Economics*, 37(3), 543-561.
- Hindriks, J., & Serse, V. (2019). Heterogeneity in the tax pass-through to spirit retail prices: Evidence from Belgium. *Journal of Public Economics*, 176, 142-160.
- Jensen, J. D., & Smed, S. (2007). Cost-effective design of economic instruments in nutrition policy. *International Journal of Behavioral Nutrition and Physical Activity*, 4(1), 10.
- Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics*, 15(3), 310-327.
- Levin, L., Lewis, M. S., & Wolak, F. A. (2017). High frequency evidence on the demand for gasoline. *American Economic Journal: Economic Policy*, 9(3), 314-47.
- Malik, V. S., Pan, A., Willett, W. C., & Hu, F. B. (2013). Sugar-sweetened beverages and weight gain in children and adults: a systematic review and meta-analysis. *The American journal of clinical nutrition*, 98(4), 1084-1102.
- McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior. P. Zarembka (Ed.), *Frontiers in Econometrics*, Academic Press, New York, 105–142.

- Nevo, A. (2000). A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand. *Journal of Economics & Management Strategy*, 9(4), 513-548.
- Nordström, J., & Thunström, L. (2009). The impact of tax reforms designed to encourage healthier grain consumption. *Journal of health economics*, 28(3), 622-634.
- O'Donoghue, T., & Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, 90(10-11), 1825-1849.
- Seetharaman, P. B., Ainslie, A., & Chintagunta, P. K. (1999). Investigating household state dependence effects across categories. *Journal of Marketing Research*, 36(4), 488-500.
- Sharma, A., Hauck, K., Hollingsworth, B., & Siciliani, L. (2014). The effects of taxing sugar-sweetened beverages across different income groups. *Health economics*, 23(9), 1159-1184.
- Te Morenga, L., Mallard, S., & Mann, J. (2013). Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. *Bmj*, 346, e7492.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Wang, E. Y. (2015). The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity. *The RAND Journal of Economics*, 46(2), 409-441.
- Weyl, E. G., & Fabinger, M. (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy*, 121(3), 528-583.
- World Health Organization. (2017). *Obesity and Overweight Fact Sheet 2018*. Available at < <http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight> > (Accessed on 6 May 2018)

Appendices

1.A: List of food categories used to compute the stock of past sugar consumption

- Biscuits
- Bonbons
- Cakes
- Candies
- Candy bars
- Canned fruits
- Cereals
- Chewing gums
- Choco pasta
- Chocolate bars
- Chocolate beverages
- Chocolate milk
- Chocolate snacks
- Donuts
- Energy drinks
- Fruit juices
- Honey
- Ice creams
- Jams
- Muffins
- Other desserts
- Pastry
- Pudding
- Soya milks
- Sugar sweetened beverages (Cola excluded)
- Syrups
- Table sugar
- Wafers
- Waffles
- Yogurt

2.A: Figures and Tables

Figure A.1: Distribution of sugar share

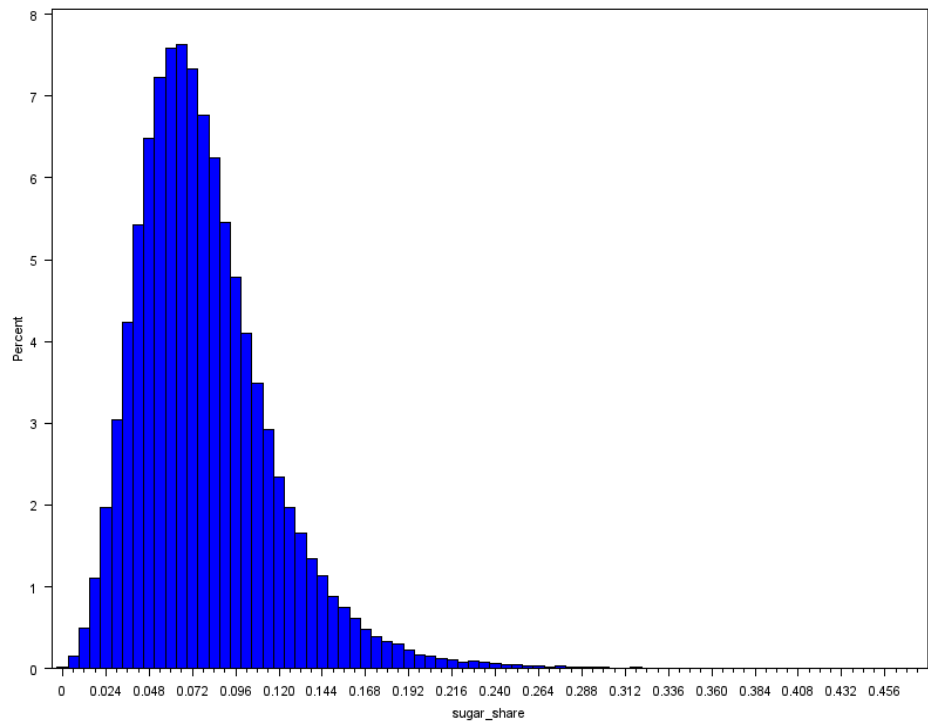


Figure A.2: Spatial price dispersion

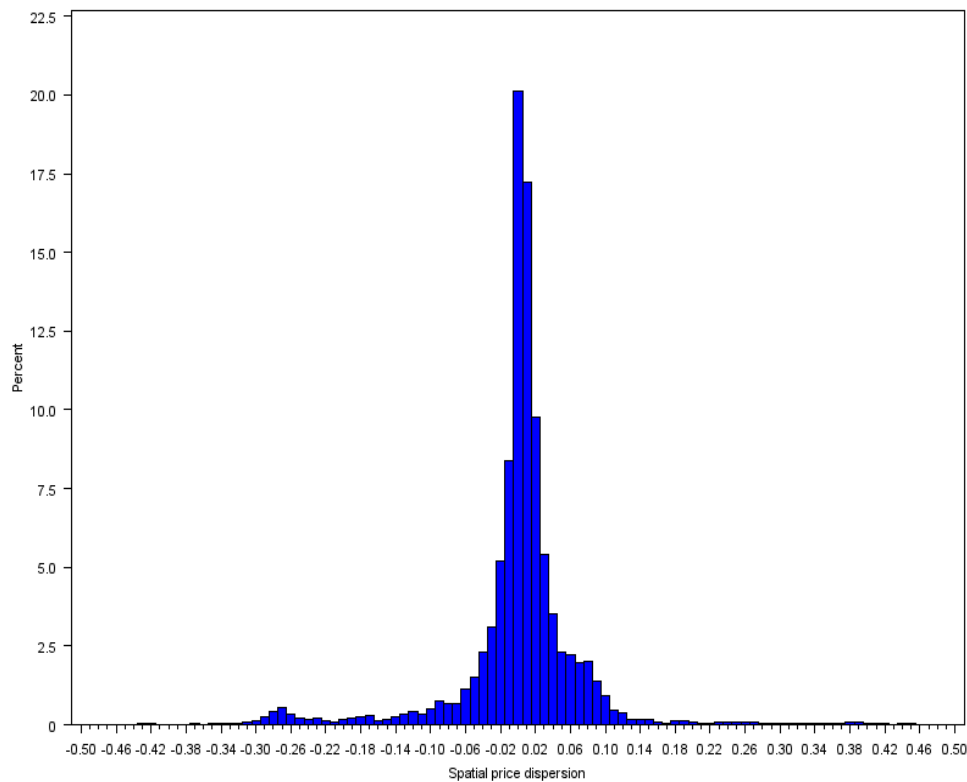


Figure A.3: Temporal price dispersion

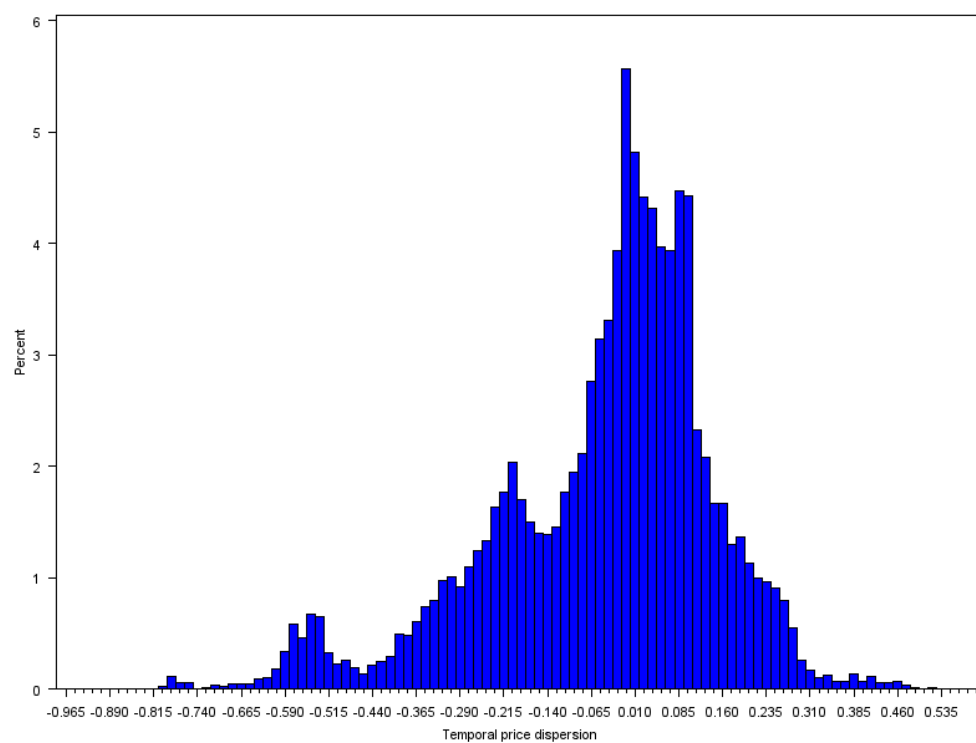


Table A.1: Logit estimates (full sample)

Parameter	Random Subsample (10.000 HHs)		Full Sample (434,475 HHs)	
	<i>Static</i> (1)	<i>Dynamic</i> (2)	<i>Static</i> (3)	<i>Dynamic</i> (4)
Price $\bar{\beta}$	0,816 (0,028)	0,450 (0,037)	0,810 (0,008)	0,448 (0,006)
Sugary $\bar{\delta}$	0,782 (0,016)	0,492 (0,018)	0,789 (0,006)	0,510 (0,003)
Size $\bar{\pi}$	0,291 (0,021)	0,175 (0,026)	0,322 (0,005)	0,231 (0,004)
Rare $\bar{\theta}$	-1,039 (0,011)	-0,500 (0,013)	-1,050 (0,03)	-0,536 (0,002)
Folder $\bar{\omega}$	-0,014 (0,007)	0,046 (0,010)	-0,010 (0,002)	0,035 (0,002)
Rebate $\bar{\varphi}$	0,079 (0,007)	0,064 (0,009)	0,078 (0,002)	0,067 (0,001)
State Dependence $\bar{\gamma}$	-	3,033 (0,005)	-	3,039 (0,001)
Observations	221.554	221.554	9.706.325	9.706.325
Model	Logit	Logit	Logit	Logit
McFadden's R^2	0,09	0,48	0,10	0,48

Notes: standard errors are in parenthesis. All parameters are significant at the 1% level. Each column displays the parameter estimates of the logit model on different samples, using both the static and the dynamic specification.

Table A.2: Mixed Logit estimates by the number of Halton sequences

Parameter	Mixed Logit (200 Hs)	Mixed Logit (500 Hs)	Mixed Logit (1000 Hs)
Price $\bar{\beta}$	2,531 (0,053)	2,464 (0,057)	2,500 (0,058)
Price σ_{β}^2	4,132 (0,040)	4,225 (0,041)	4,509 (0,044)
Sugary $\bar{\delta}$	0,781 (0,030)	0,782 (0,030)	0,802 (0,030)
Sugary σ_{δ}^2	2,428 (0,019)	2,467 (0,020)	2,445 (0,019)
Size $\bar{\pi}$	-0,441 (0,024)	-0,372 (0,025)	-0,342 (0,025)
Size σ_{π}^2	2,977 (0,019)	2,806 (0,017)	2,793 (0,017)
Rare $\bar{\theta}$	-0,973 (0,013)	-0,974 (0,013)	-0,984 (0,014)
Folder $\bar{\omega}$	-0,026 (0,011)	-0,026 (0,011)	-0,027 (0,011)
Rebate $\bar{\varphi}$	0,163 (0,010)	0,164 (0,010)	0,165 (0,010)
State dependence $\bar{\gamma}$	1,854 (0,005)	1,839 (0,005)	1,834 (0,005)
Log-Likelihood	-246460	-246060	-245920

Notes: standard errors are in parenthesis. All parameters are significant at the 1% level. Each column displays the parameter estimates using 200, 500 and 1000 Halton sequences (Hs).

Table A.3: Own and cross-price elasticities of demand (Logit)

	A.1	A.2	A.3	B.1	B.2	B.3	C.1	C.2	D.1	D.2	E.1	E.2	F.1	G.1
A.1	-0,37	0,04	0,05	0,06	0,04	0,06	0,08	0,06	0,06	0,07	0,08	0,05	0,08	0,07
A.2	0,05	-0,23	0,05	0,05	0,05	0,06	0,07	0,06	0,06	0,07	0,07	0,05	0,08	0,07
A.3	0,02	0,01	-0,30	0,02	0,02	0,03	0,03	0,02	0,02	0,02	0,03	0,02	0,03	0,02
B.1	0,05	0,03	0,04	-0,40	0,04	0,05	0,06	0,05	0,05	0,06	0,07	0,04	0,07	0,06
B.2	0,04	0,05	0,05	0,05	-0,24	0,06	0,06	0,06	0,05	0,07	0,07	0,04	0,07	0,07
B.3	0,01	0,01	0,02	0,01	0,01	-0,33	0,01	0,01	0,01	0,01	0,00	0,01	0,01	0,01
C.1	0,02	0,01	0,02	0,02	0,01	0,02	-0,59	0,02	0,02	0,02	0,03	0,02	0,03	0,02
C.2	0,02	0,02	0,03	0,02	0,02	0,03	0,03	-0,37	0,03	0,03	0,04	0,02	0,04	0,03
D.1	0,01	0,00	0,01	0,01	0,01	0,01	0,01	0,01	-0,10	0,01	0,01	0,01	0,01	0,01
D.2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-0,11	0,00	0,00	0,00	0,00
E.1	0,00	0,00	0,01	0,01	0,00	0,00	0,01	0,01	0,01	0,01	-0,50	0,00	0,01	0,01
E.2	0,02	0,02	0,02	0,02	0,02	0,03	0,03	0,03	0,03	0,03	0,03	-0,18	0,04	0,04
F.1	0,01	0,00	0,01	0,01	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	-0,52	0,01
G.1	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,00	0,01	-0,29

Notes: own-price elasticities are in bold characters. Elasticities are computed at the sample average.

Table A.4: Parameter estimates of Logit model with observed group heterogeneity

Parameter	Estimate	Demographics	Estimate		
			Price	Sugary	Size
Price $\bar{\beta}$	14,185 (0,143)	Income	-1,379 (0,014)	-0,517 (0,008)	0,295 (0,015)
Sugary $\bar{\delta}$	5,280 (0,080)	Education	-0,148 (0,013)	-0,094 (0,007)	-0,048 (0,014)
Size $\bar{\pi}$	-2,837 (0,151)	Sugar share	-0,695 (0,046)	2,257 (0,033)	
Rare $\bar{\theta}$	-0,547 (0,003)	Household's size	0,049 (0,002)	0,062 (0,001)	0,059 (0,002)
Folder $\bar{\omega}$	0,029 (0,002)	Kids	0,018 (0,007)	0,015 (0,004)	0,055 (0,008)
Rebate $\bar{\varphi}$	0,069 (0,002)	Age>70	0,116 (0,013)	0,011 (0,007)	-0,247 (0,014)
State dependence $\bar{\gamma}$	3,069 (0,001)	70>Age>58	0,067 (0,011)	-0,032 (0,005)	-0,278 (0,011)
		58>Age>46	0,014 (0,009)	0,001 (0,005)	-0,159 (0,009)
		46>Age>34	0,007 (0,008)	-0,080 (0,005)	-0,005 (0,009)
Brand fixed effects			yes		
N° of Observations			4.207.299		
McFadden's R^2			0,50		

Notes: *standard errors are in parenthesis.*

Table A.5: Robustness check (time-varying brand intercepts)

Parameter	Dynamic Mixed Logit (1)	Dynamic Mixed Logit (2)
Price $\bar{\beta}$	2,531 (0,053)	2,829 (0,053)
Price σ_{β}	4,132 (0,040)	4,048 (0,037)
Sugary $\bar{\delta}$	0,782 (0,030)	0,818 (0,030)
Sugary σ_{δ}	2,428 (0,019)	2,436 (0,019)
Size $\bar{\pi}$	-0,441 (0,024)	-0,670 (0,027)
Size σ_{π}	2,977 (0,019)	2,832 (0,018)
Rare $\bar{\theta}$	-0,973 (0,013)	-0,912 (0,015)
Folder $\bar{\omega}$	-0,026 (0,011)	-0,058 (0,012)
Rebate $\bar{\varphi}$	0,163 (0,010)	0,183 (0,011)
State Dependence $\bar{\gamma}$	1,854 (0,005)	1,844 (0,005)
Time-varying brand effect	no	yes
McFadden's R^2	0,55	0,57

Notes: standard errors are in parenthesis. All parameters are significant at the 1% level.

3.A Demand Curves

