Who is Most Affected by Soda Taxes? Evidence from Purchases At-Home and Away-From-Home

Xirong Lin* Linqi Zhang†

August 7, 2023

Abstract

Using a novel dataset that includes at-home and away-from-home food purchases, we study who is affected by soda taxes. We nonparametrically estimate a random coefficient nested logit model to exploit the rich heterogeneity in preferences and price elasticities across households, including SNAP participants and non-SNAP-participant poor. By simulating its impacts, we find that soda taxes are less effective away-from-home while more effective at-home, especially by targeting the total sugar intake of the poor, those with high total dietary sugar, and households without children. Our results suggest that ignoring either segment can lead to biased policy implications.

JEL: D12, H22, H25, H71, L66

Keywords: soda tax, discrete choice demand, preference heterogeneity

*Department of Economics, Shanghai University of Finance and Economics, (e-mail: xironglin0614@gmail.com).

†Department of Economics, Boston College, (e-mail: linqi.zhang@bc.edu).

We thank Joanna Venator, Tomoki Fujii and Charles Murry. All errors are our own.

1 Introduction

It is well known that sugary drinks can have negative effects on health, including a correlation with diabetes, heart disease, and childhood obesity (Currie (2009), Currie *et al.* (2010), Gortmaker *et al.* (2009), Griffith *et al.* (2020), Cutler *et al.* (2003)). Experts suggest that in the developed world sugar consumption is far above the recommended level. Therefore, the resulting individual and social costs of the internalities (that is, the ignored future costs of current consumption) and externalities (costs that are borne by others) of sugar consumption have attracted policymakers' concern. Many countries have implemented taxes on sugar-sweetened beverages (SSB) in order to discourage soft drink consumption.[1] Soft drinks are also a main contributor to the sugar consumption of vulnerable individuals: the young, high sugar consumers, and the poor. Whether soda taxes can be effective in reducing sugar consumption and improving welfare depends crucially on how demand responses vary across different demographic groups and may also vary across consumption locations (at-home versus away-from-home).

In this paper, we assess whether soda taxes are effective at lowering the sugar consumption of households by taking into account all channels (at-home, on-the-go, and restaurants) of households' SSB consumption. We account for heterogeneity in household taste for SSB and price sensitivity in these different channels in order to comprehensively and precisely estimate the demand responses of households who are targeted by the policy (i.e., households with children, high sugar consumers, SNAP (Supplemental Nutrition Assistance Program) participants, and non-SNAP-participant poor). We reveal new evidence on the important but understudied away-from-home segment (on-the-go and restaurants) of the market. With novel data, we can estimate a model of consumer choice in both at-home and away-from-home segments. We uncover household preferences in each segment, and we then simulate the impacts of a soda tax, allowing for different pass-through rates in each segment.

Extending the existing literature, we make two main advances in this paper. First, we document descriptive patterns of household SSB purchases at-home, on-the-go, and in restaurants. We use a unique dataset, USDA's National Household Food Acquisition and Purchase Survey (FoodAPS), which contains detailed information about U.S. household food purchases at-home, on-the-go, and in restaurants in one week. As a scanner type dataset, similar to other datasets like Nielsen Consumer Panel Data, all product information is at bar-code (UPC) level. For purchases in restaurants, receipts and recall books are used to record purchase information. FoodAPS is the first and only dataset that include both at-home and away-from-home food purchase information at the level of bar-codes.[2]

Most papers on SSB taxes look at only one of the segments, at-home (Allcott *et al.* (2019a), Bollinger & Sexton (2018)), on-the-go (Dubois *et al.* (2020)), or in restaurants (Moran *et al.* (2019)). In fact, the latter two channels have rarely been studied although they actually constitute

[1]See Allcott *et al.* (2019b) for a summary of countries that implement sugar-sweetened beverage taxes.

[2]However, FoodAPS was filed between April 2012 and January 2013, which was before the enactment of any SSB taxes in the U.S.. Hence, we cannot exploit any variation in tax policy by time or region.

a large fraction of household expenditures. For example, Americans drink 52% of SSB calories at-home and 48% of SSB calories away-from-home.[3] Different from in other countries, U.S. households purchase a considerable amount of SSB from the nation's largest chain restaurants, particularly when combination meals or kids' menu items are ordered (Moran *et al.* (2019)).[4] These facts imply that any SSB tax analysis missing one of the segments will be an incomplete documentation of the impact of soda taxes on household sugar intake.

Second, we estimate a discrete choice demand model by adapting the nonparametric framework of Fox *et al.* (2011) and Fox *et al.* (2016). Prior work that uses a similar approach includes Nevo *et al.* (2016) and Blundell *et al.* (2020). The estimation technique is nonparametric in the sense that it estimates the distribution of random coefficients over a fixed grid of potential values, rather than assuming that the random coefficients are drawn from a known distribution. This is important given the recent empirical finding in Dubois *et al.* (2020) that preferences vary with demographics in ways that would be difficult to capture by specifying a priori the distribution of random coefficients.[5] Our model allows us to use the rich demographic information in FoodAPS, including SNAP participation and eligibility and household income and composition, in order to reveal the diverse preferences and elasticities across household types.

We find that preferences and elasticity vary with demographics in terms of SNAP participation, income, the existence of children, and household dietary sugar. Consistent with previous literature, low-income households have the strongest preferences for SSB. But among the poor, SNAP eligible nonparticipants have weaker preferences than SNAP poor and are the most elastic to price among all groups. Elder households and those without children have weaker preferences for SSB and are more sensitive to price. Lastly, SSB preferences exhibit an increasing relationship with dietary sugar while price elasticity exhibits a decreasing relationship with dietary sugar.

In terms of heterogeneity among segments, we find that consumers have diverse preferences and elasticity at-home versus away-from-home. For example, low-income households have strong preferences for SSB at-home while high-income households prefer more SSB in restaurants. Those with higher dietary sugar from SSB obtain much more SSB at-home than away-from-home and vice versa. In terms of price responsiveness, overall the average elasticity across all groups at-home is larger than that away-from-home. For heterogeneity along the line of SNAP participation and income, elasticities vary widely across groups both at-home and away-from-home. But for heterogeneity in the dimension of total dietary sugar from SSB, there is minimal variation in elasticities away-from-home but large differences in elasticities at-home.

Our findings suggest that on average, the current taxes of the form and size implemented in the U.S. lead to reductions of around 18.12 percent, 5.75 percent, and 14.53 percent in the total sugar intakes from SSBs at-home, away-from-home, and in total. We find that soda taxes

[3] See Kit *et al.* (2013).

[4] For example, in the U.K. data from Dubois *et al.* (2020), on-the-go purchases are three times as large as that of restaurant purchases, while the U.S. sample shows the opposite pattern.

[5] Dubois *et al.* (2020) overcome this problem by estimating an individual-level demand model using U.K. households SSB purchases on-the-go. We cannot follow the same strategy because it is not possible to have individual-level consumption information at-home and in restaurants. Hence, we choose the framework of Fox *et al.* (2011) and Fox *et al.* (2016) as a middle ground. Throughout the paper, we compare our model, findings, and implications to Dubois *et al.* (2020).

are less effective away-from-home and there is little variation in responses across households. One reason is that households (mainly the high-income households) who have strong preferences for SSB away-from-home are also less price sensitive and hence have small reductions in sugar intake from SSB. In contrast, we find substantial variation in demand responses at-home across households. Soda taxes are relatively effective at targeting the total sugar intake of the poor, those with high sugar consumption, and households without children.[6] Lastly, our results suggest that ignoring any segment will lead to biased policy implications. For example, Dubois *et al.* (2020) find that the soda tax is less successful at targeting those with high total dietary sugar for the on-the-go segment. However, we find that the total (at-home and away-from-home) reduction in sugar intake from SSB is largest for those households if we account for all segments (the reduction almost doubles that of the high-income households).

One major debate about SSB taxes is the concern that they are regressive, i.e., the poor spend a disproportionately large fraction of expenditures on SSB and they end up bearing the largest share of the tax burden. We use compensating variation (the amount of money that an individual needs to reach her pre-tax utility level after the imposition of an excise tax) as our welfare measure for soda taxes and compare it across household groups by income and SNAP participation. Unlike the previous literature which often finds a larger compensating variation for low-income households, we find that even though low-income households obtain more added sugar from SSB, their compensating variation is not much higher than that of high-income households. That is because soda taxes are based on volume rather than the amount of sugar in each drink. Given the fact that, high-income households obtain more soft drinks away-from-home, this was not accounted for by the previous literature. Meanwhile, both groups can obtain similar amounts of soft drinks while the sugar amount in each drink is much higher in low-income households. These findings suggest that household preference heterogeneity in each purchase segment (at-home or away-from-home) and their preference for the specific drink types (in terms of the amount of sugar) together determine the welfare cost of a soda tax. Ignoring either segment can lead to biased policy implications of soda taxes.

Literature Review This paper belongs to a burgeoning literature on the effect of soda taxes.[7] One strand of this literature exploits a specific SSB tax implementation or reform, used as natural experiments, in order to estimate the effects of those reforms on household SSB expenditures (Seiler *et al.* (2021), Rojas & Wang (2017), Bollinger & Sexton (2018)). Most of these studies use retailer-side scanner data in order to study the overall impact of an SSB tax on prices and consumption. The findings speak to the effect of the specific reform and there are different results for different places. For example, Seiler *et al.* (2021) find that a soda tax in Philadelphia is passed through at an average rate of 97% and demand decreases by 46%. However, accounting for cross-border shopping reduces the demand response by 20%. Rojas & Wang (2017) compare SSB tax

[6]Following previous literature like Dubois *et al.* (2020), we measure the effectiveness of SSB tax in terms of the level reduction, rather than percentage reduction in sugar from SSB. For example, those with high total dietary sugar have the strongest preferences for SSB while the lowest price elasticity. However, the soda tax is effective at targeting these households because their level-reduction in sugar from SSB is still the largest across groups even though their percentage reduction in sugar from SSB is the lowest.

[7]See Allcott *et al.* (2019a) and Cawley *et al.* (2019) for overviews of the theory and evidence of an SSB tax.

pass-through rates and volume sales in Washington DC and Berkeley CA. They find that both retail price and volume sales in Washington react sharply while in Berkeley retail price reacts only marginally with no effect on the volume of sales. Yet, these empirical studies do not provide any mechanisms that explain the conflicting findings or allow further re-valuation of alternative reforms in order to derive the most effective policy.

In contrast, our paper belongs to the other strand of the literature that include papers with structural models of household demand for SSB and that simulates counterfactual exercises of alternative tax policies. These include Dubois *et al.* (2020), Allcott *et al.* (2019a), Wang (2015), Bonnet *et al.* (2012), Harding & Lovenheim (2017), and Chernozhukov *et al.* (2019). We complement these papers by exploiting a novel dataset with consumer SSB consumption in all segments as well as rich demographic information on the policy targeted groups. Our nonparametric demand model estimates and the counterfactual findings have meaningful implications for the effectiveness of soda taxes. Similar to Ramsey (1927), Diamond & Mirrlees (1971a,b), and Miravete *et al.* (2020), we find that preference heterogeneity among consumers and variation in demand elasticities by segments are important aspects in the design of optimal tax schemes.

Our paper also complements the literature on food and beverage consumption with consumer side scanner data at-home (Aguiar & Hurst (2007), Lin (2023), and Dubois *et al.* (2014)), away-from-home (Dubois *et al.* (2020), O'Connell & Smith (2020), and Griffith *et al.* (2022), Saksena *et al.* (2018), Moran *et al.* (2019)), or both (Okrent & Alston (2012)).

The rest of this paper is structured as follows. In Section 2 we describe the data and the soft drinks market. We present a detailed descriptive analysis of the consumption patterns across diverse demographic groups and retail segments. In Section 3 we present the demand model, the identification, the nonparametric estimation, and the empirical estimates of the model parameters. In section 4 we simulate the counterfactual tax exercise and discuss its implications on effectiveness, targeting, and regressivity of an SSB tax. Section 5 concludes.

2 The Nonalcoholic Drinks Market

We model household behavior in the nonalcoholic drinks market. Nonalcoholic drinks include soft drinks (e.g., carbonated drinks, commonly referred to as soda, fruit drinks, sport and energy drinks, and sweetened coffee and tea), alternative sugary drinks (fruit and vegetable juice, unsweetened coffee and tea, flavored milk), and bottled water. soda taxes are typically imposed on soft drinks that contain added sugar. Diet drinks and drinks with natural sugar like fruit juice are normally exempted from soda taxes .

We focus on household purchases at-home (grocery store, supermarket, pharmacy, club store, dollar store, gas station/market), on-the-go (convenience store, vending machine, and retail store), and in the restaurants (coffee shop and care, fast-food outlet, restaurants, drinking places, other store and farmers market). Drinks market at-home have been extensively studied in the literature, while there are not many studies on drinking away-from-home. But it is very important to study this segment because near half of the nonalcoholic drinks are purchased away-from-home in the U.S. (Table 2). We document the purchase behavior in the three segments separately in this section. However, due to the small fraction of the on-the-go segment out of total SSB expenditures (10%), in the demand estimation, we aggregate drinks on-the-go and restaurants into one segment, that is, drinks away-from-home.

We start by documenting household nonalcoholic drinks purchases in the FoodAPS dataset, e.g., the prices paid, expenditures and expenditure shares, products bought, and places types. We then relate the purchase behavior to household demographic characteristics such as household income, age, the existence of children, and overall dietary sugar intake from SSB.

2.1 FoodAPS

The data that we use is the Food Acquisition and Purchase Survey (FoodAPS), which was fielded from April 2012 through January 2013.[8] Mathematica Policy Research conducted the survey under contract to USDA Economic Research Services (ERS). The FoodAPS data collect information on foods that all household members acquired over a 7-day period, for a sample of 4,826 US households. These include all foods and drinks brought into the home as well as those got outside the home, including those made in bars and restaurants. To our knowledge, the FoodAPS away-from-home survey is unique.

The nationally representative sample of 4,826 households includes four distinct sub-groups, based on household income and participation in the Supplemental Nutrition Assistance Program (SNAP): (1) SNAP participants, (2) Households with income below the poverty line but do not participate in SNAP, (3) Households with income at or above 100 percent and less than 185 percent of the poverty guideline but do not participate in SNAP, and (4) Households with income equal to or greater than 185 percent of the poverty guideline and do not participate in SNAP.

Households in the FoodAPS data use scanners to scan all grocery purchases brought into the home. In both the food at-home and food away-from-home surveys we know what products (at the bar code, UPC level) were purchased, the product attributes, and the transaction price. We also observe information on the household and individual attributes, such as household size and composition, demographic characteristics, income, and participation in food assistance programs. For more information regarding the collection of the data, please refer to the Appendix Section 1.

We measure the SSB consumption as the sum of all purchases of SSB in ounces during the data collection week. Similarly, we measure the SSB expenditure as the total spending on SSB during a seven-day period. We further construct the fraction of SSB expenditures spent at-home, in the restaurants, and on-the-go respectively, given by the SSB expenditures in each segment divided by the total SSB expenditures. The deciles of income are constructed using the household

[8] Soda taxes in the U.S. were introduced between 2014 and 2017. Hence, the FoodAPS sample is ideal for analyzing the demand with simulated soda taxes, as no soda taxes were in place during the sample period.

average monthly income.[9]

In our estimation, we separately estimate the at-home and away-from-home demand for soft drinks. We use information on the food at-home purchases of 4,412 households and food away-from-home purchase of 3,977 households. For either segment, we define a choice occasion as a trip in which a household makes a purchase of any good (including SSB drinks, non-SSB drinks or foods). When purchasing drinks for consumption at home, households choose a single item 51 percent of the time, whereas for consumption outside the home, 83 percent of the time the households choose a single item. On the remaining trips, the household chooses more than one type of drink product. In these cases, we treat each purchase in the multi-purchase transaction as a separate choice occasion. We observe households on an average of 9.7 choice occasions in our estimation sample. In total, the sample contains 46,921 choice occasions. For over 95 percent of households we observe more than five choice occasions.

Table 1 describes the distribution of place types where households purchase SSB. In total, the sample contains 3,556 households who make any SSB purchases during the data collection week. Among them, 514 households buy SSB on-the-go, 1,837 households buy SSB at-home, and 2,115 households make SSB purchases in restaurants. We observe more than 50 percent of households purchasing SSB only in one of the three segments. For households who make SSB purchases in only two of the three segments, most of them make it at-home and in restaurants. Only 11 percent of the households purchase SSB in all three segments during the week.

Table 1: Place Types

	At-home	Restaurant	On-the-go	At-home + Restaurant	At-home + On-the-go	Restaurant + On-the-go	At-home + Restaurant + On-the-go
Number of households	745	989	124	914	178	212	394
Percent of sample	20.95	27.81	3.49	25.7	5.01	5.96	11.08

Notes: The table shows different place type combinations at which households make any SSB purchases during the data collection week. Columns 1 to 3 show the number of households who purchase SSB in only one of the three place types. Columns 4 to 6 show the number of households who purchase SSB in two out of the three place types. The last column is the number of households who purchase SSB in all three place types.

2.2 Places, Prices, and Products

2.2.1 Places

Consumers visit different retailers when they shop at-home, on-the-go, or in restaurants. This implies that the prices and choice sets that they are faced with in a choice occasion will also differ. In Table 2, we describe the types of retailers and the expenditure share of drinks at each type. In total, the away-from-home segment accounts for 46% of the total SSB spending, while that number is 54% for the at-home segment. Previous literature that ignore the away-from-home drinks spending miss a large fraction of household total spending on drinks. Under the away-from-home segment, convenience store, vending machine, and retail store together can be

[9]We did the same exercise using household *equivalized* income to account for household size and composition. It does not lead to significant difference.

classified as the on-the-go segment. They account for only 18% of spending in the away-from-home segment. This suggests that previous literature like Dubois *et al.* (2020) and O'Connell & Smith (2020) who only study SSB at-home or on-the-go overlook 38% of household total SSB spending that happens in restaurants and cafes.

Table 2: Expenditure share (%)

Away-from-home	46%	At-home	54%
Convenience store	10%	Combination grocery/other	3%
Retail store	6%	Dollar store	4%
Coffee shop and cafe	15%	Convenience store	4%
Fast-food outlet	34%	Gas station/market	2%
Restaurants	31%	Grocery store, large	1%
Drinking place	1%	Grocery store, medium	1%
Vending machine	2%	Pharmacy	2%
Other store and farmers market	0%	Super store	44%
		Supermarket	35%
		Club stores	4%

Notes: Numbers are % of drinks spending, in at-home and away-from-home segments, by retailer.

We do not model household choices over which place to shop in. We assume that the decision is driven by factors such as the proximity of nearby super stores or restaurants and overall preferences for different segments and place types (for which we control in demand). We assume away from the possibility that consumers choose places in order to search for a temporarily low price for a specific drink. The assumption is reasonable because we find that consumers tend to choose nearby places to shop in FoodAPS. Previous evidence also shows that fixed shopping costs lead consumers to undertake their grocery shopping in one or a small number of stores. Similar assumption is also made in O'Connell & Smith (2020).

2.2.2 Products

In Table 3, we describe the products of soft drinks available in the U.S. market, their percentage of transactions, and mean prices.[10] The way we construct products is the following. We aggregate millions of UPCs (bar code) into ten product categories defined by the FoodAPS: soft drinks, fruit drinks, sport and energy drinks, sweetened coffee and tea, diet drinks, fruit and vegetable juice, unsweetened coffee and tea, flavored milk, flavored water, and water. By comparing the definition of drinks in the FoodAPS and the definition of taxable non-alcoholic drinks in the city and county websites in the U.S., we conclue that the current SSB tax is placed on the first four categories: soft drinks, fruit drinks, sport and energy drinks, and sweetened coffee and tea.[11]

We further classify whether a product is purchased in one of the two segments: at-home or away-from-home (on-the-go, or in restaurants). We also allow products to differ by packaging formats (regular, large, or multi-pack).[12] As a result, a product in the demand estimation is

[10] Mean prices are calculated as the average transaction prices across choice occasions.

[11] Information regarding the taxable non-alcoholic beverages can be found in the city and county websites. See, e.g., `https://www.seattle.gov/license-and-tax-administration/business-license-tax/other-seattle-taxes/sweetened-beverage-tax.`.

[12] Regular size is defined as a container size smaller than 32 ounces. Large size is defined as a container size larger than 32 ounces.

defined as either a category-package-segment, or a category-segment combination[13].

From Table 3, we find that households in our sample on average purchase an SSB product in 34 percent of choice occasions. Among them, at-home purchases of SSB products account for around 16 percent of choice occasions, while away-from-home purchases of SSB products account for around 18 percent of choice occasions. The most frequently purchased product is the regular soft drinks away-from-home, which account for 9.4 percent of choice occasions. When consumers do not purchase any of the drinks during a choice occasion, we assume them to choose the "outside options": either the household purchases other food (e.g. meat or snacks) if it was a trip to a food store, or the household purchases a meal without ordering drinks if it was a trip to an eating place.

Table 3: Products

Product	Percentage	Price (dollar)	Product	Percentage	Price (dollar)
AH Regular Soft Drinks	1.387	1.457	AFH Regular Soft Drinks	9.409	1.505
AH Large-Bottle Soft Drinks	3.335	1.337	AFH Large-Bottle Soft Drinks	1.176	1.684
AH Pack Soft Drinks	3.097	3.507	AFH Fruit Drinks	1.703	2.055
AH Regular Fruit Drinks	1.473	1.294	AFH Large-Bottle Fruit Drinks	0.102	2.201
AH Large-Bottle Fruit Drinks	2.630	2.141	AFH Regular Sport and Energy Drinks	0.663	1.974
AH Pack Fruit Drinks	1.174	2.611	AFH Large-Bottle Sport and Energy Drinks	0.051	3.069
AH Regular Sport and Energy Drinks	1.552	1.171	AFH Regular Sweetened Coffee and Tea	4.288	2.097
AH Pack Sport and Energy Drinks	0.460	6.406	AFH Large-Bottle Sweetened Coffee and Tea	0.258	1.682
AH Regular Sweetened Coffee and Tea	0.431	1.624	AFH Regular Diet Drinks	2.421	1.506
AH Large-Bottle Sweetened Coffee and Tea	0.810	2.485	AFH Large-Bottle Diet Drinks	0.318	1.528
AH Pack Sweetened Coffee and Tea	0.104	5.000	AFH Fruit and Vegetable Juice	0.825	1.703
AH Regular Diet Drinks	0.603	1.359	AFH Unsweetened Coffee and Tea	4.245	1.666
AH Large-Bottle Diet Drinks	1.147	1.486	AFH Flavored Milk	1.023	2.095
AH Pack Diet Drinks	0.961	3.970	AFH Flavored and Enhanced Water	0.151	1.728
AH Fruit and Vegetable Juice	3.331	3.064	AFH Water	2.231	1.245
AH Unsweetened Coffee and Tea	0.311	1.997			
AH Flavored Milk	0.654	2.471	AH Outside Options	22.292	0.000
AH Flavored and Enhanced Water	1.485	1.439	AFH Outside Options	21.300	0.000
AH Water	2.596	2.506	Total Number of Choice Occasions	46921	46921

Notes: At-home segment is abbreviated as AH. Away-from-home segment is abbreviated as AFH. Regular size is defined as a container size smaller than 32 ounces. Large size is defined as a container size larger than 32 ounces. Multi-pack is defined as a pack with more than one unit of bottles/cans. AH (AFH) outside options refer to any foods or drinks except for nonalcoholic bevarages that are obtained from AH (AFH) segment. The second column shows the percentage of choice occasions where the indicated product is purchased, based on transactions made by the 4,683 households in the estimation sample. Prices are averaged across choice occasions.

Unlike most of the previous literature where a product is defined as a brand-size combination (e.g., Dubé (2005), Marshall (2015), Dubois *et al.* (2020)), we abstract away from the substitution across brands because brand information is not available in the public file of FoodAPS. Also in the away-from-home segment, most items have missing UPCs. Instead, we aim to measure how consumers substitute between regular coke and diet coke rather than between regular Cola and regular Pepsi.

2.2.3 Prices

For each transaction, we observe the type of store the consumer shops and the total expenditures on each UPC. We calculate the transaction price as the expenditure on a UPC per unit purchased.

[13] As in Dubois *et al.* (2020), for fruit juice, unsweetened coffee and tea, flavored milk and water, we aggregate across different sizes. In total, these non-SSB beverage categories account for less than 16 percent of the market.

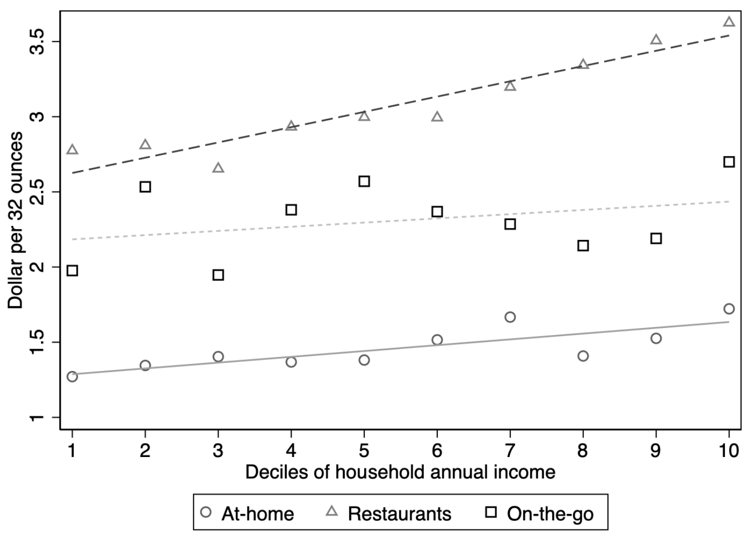
In Figure 1, we show the average prices of all SSB purchased across all choice occasions in each segment by household income, age of the primary respondent, and total added sugar from SSB. First, panel (a) shows that overall, SSB prices are the highest in restaurants, second highest on-the-go, and the lowest at-home. There is a positive relationship between prices and income in all three segments. The slope is the steepest in the restaurants segment. In other words, richer households buy much more expensive SSB in restaurants. Second, panel (b) shows that the age of the primary respondent is positively correlated with SSB prices in restaurants while negatively correlated with SSB prices on-the-go. Together with panel (a), this implies that richer and older households pay much higher price for SSB products in restaurants. Lastly, SSB price is negatively correlated with weekly added sugar from SSB in all three segments. In other words, those with higher dietary sugar buy cheaper SSB products, no matter at-home, on-the-go, or in restaurants.

In the demand estimation, we construct price index p_j for each product type j defined in Table 3. We average the transaction prices of all UPCs belonging to a particular product type in a given month in a specific retailer across all consumers. More specifically, each product type contains $\mathcal{K}_j$ UPCs indexed by $k = 1, ..., \mathcal{K}_j$. We denote the price paid by consumer i for each UPC at time t in a retailer r as $p_{kj}(i, r(\tau), t(\tau))$ and the price index for each product type j as $p_{jr(\tau)t(\tau)}$. We will discuss the notation in more detail in Section 3. The resulting price index for each product type is a product-month-retailer tuple. For example, assume the total number of transactions under product type j happened at time t in a retailer r is $\mathcal{T}_{jr(\tau)t(\tau)}$, we compute the product price index as

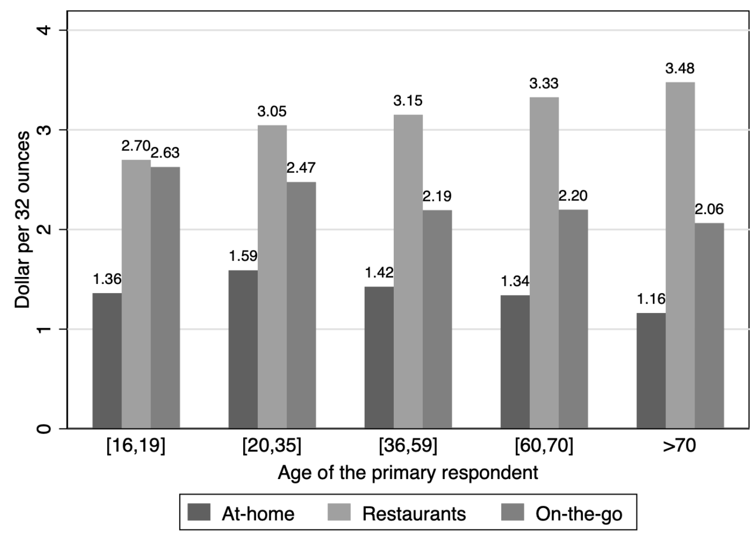
$$p_{jr(\tau)t(\tau)} = \frac{1}{\mathcal{T}_{jrt}} \sum_{\tau=1}^{\mathcal{T}_{jrt}} \sum_{k=1}^{\mathcal{K}_j} p_{kj}(i, r(\tau), t(\tau))$$

If there are no transactions happened for a product-month-retailer tuple, we impute the price index using the average price of the same product in the same retailer in last month. If a product-retailer type involves no transactions in all time periods, we treat the price as a missing value. For example, purchases of large-bottle fruit drinks are never observed in drinking places. In the later section where we introduce the demand model, we will assume that products with missing prices are not included in consumers' choice set. In the previous example, this assumption implies that if a consumer visits a drinking place, large-bottle fruit drinks are not on the menu. We present details of data construction and how missing values are dealt with in Appendix Section 2.

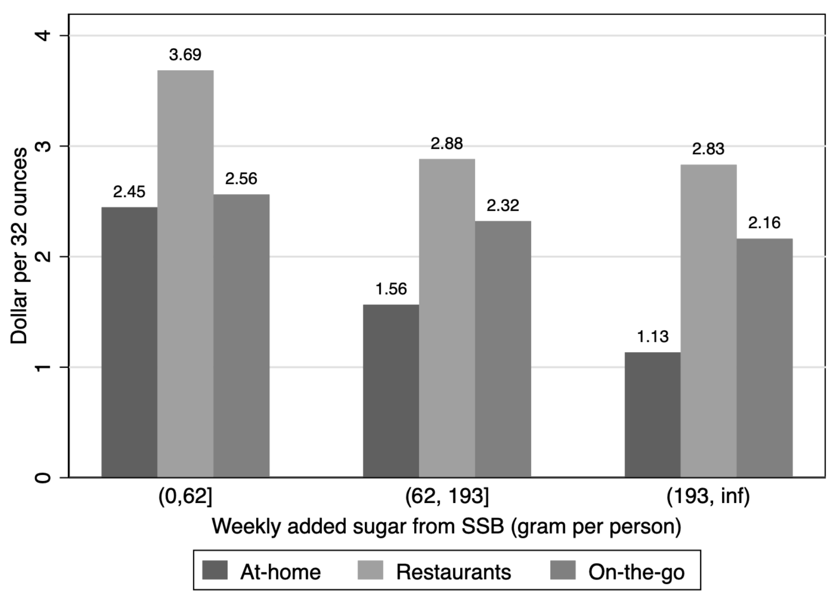
A typical problem faced by researchers in discrete choice demand estimation is that the prices of products not chosen by the consumers are not observed. Using the mean prices to proxy for these unobserved prices may induce measurement error problem. These errors are "Berkson" errors (Berkson (1950)). Schennach (2013) proposes a solution to fix the measurement error in prices in continuous demand models. Methods that can be applied to the discrete choice framework is an interesting research question that is worth future research.



(a) By deciles of household annual income



(b) By age of the primary respondent



(c) By total added sugar from SSB

Figure 1: Average price of SSB

Notes: The figure shows how purchase price varies across households with or without children, age groups, and total added sugar from SSB. In plot (a), the household annual income is equivalized by the OECD-modified equivalence scale. In plot (b), age groups are classified according to the same cutoffs as in the FoodAPS dataset. Plot (c) is restricted to households who have positive amount of added sugar from SSB. The cutoff levels are the terciles of the total sugar from SSB.

2.3 Demographics

2.3.1 Income

In Figure 2, we show that income is negatively correlated with household shares of SSB expenditures on food at-home while positively correlated with household shares of SSB expenditures in restaurants. This pattern mainly reflects two commonly found evidence in the empirical literature. First, poor people are more likely to have a less healthy diet, like high dietary sugar. Second, income and restaurant expenditures are positively correlated. Figure 2 also shows that household SSB expenditure shares on-the-go do not vary much across the deciles of household income.

The above findings have significant implications on how current literature on SSB consumption could go wrong by using only data on food at-home. For papers that only use food at-home SSB expenditures, they will miss the important asymmetry of income and SSB expenditures. First, they will mistakenly conclude that lower-income households purchase more SSB than higher-income households. Second, their elasticities estimates will also miss the heterogeneity in household demand responses by household income and segments. For papers that only use on-the-go SSB expenditures, like Dubois *et al.* (2020), they will not find any income effect in SSB demand, which is certainly not true from the evidence of Figure 2. Moreover, in the US, SSB consumption on-the-go constitutes very small shares, only 10%, of households overall SSB expenditures.

Another important implication from Figure 2 is related to regressivity of an SSB tax. Empirical evidence suggests that poor households spend more on SSB than richer households.[14] This implies that an SSB tax will fall disproportionately more on the poor. However, we see that high-income households actually spend more on SSB in restaurants than poor households. So the overall regressivity of an SSB taxes is unknown when we account for both food at-home and food away-from-home.

[14] For example, Allcott *et al.* (2019b) find that in the Nielsen Consumer Panel dataset, households with annual income below \$10,000 purchase about 101 liters of SSB per adult each year, while households with income above \$100,000 purchase only half that amount.

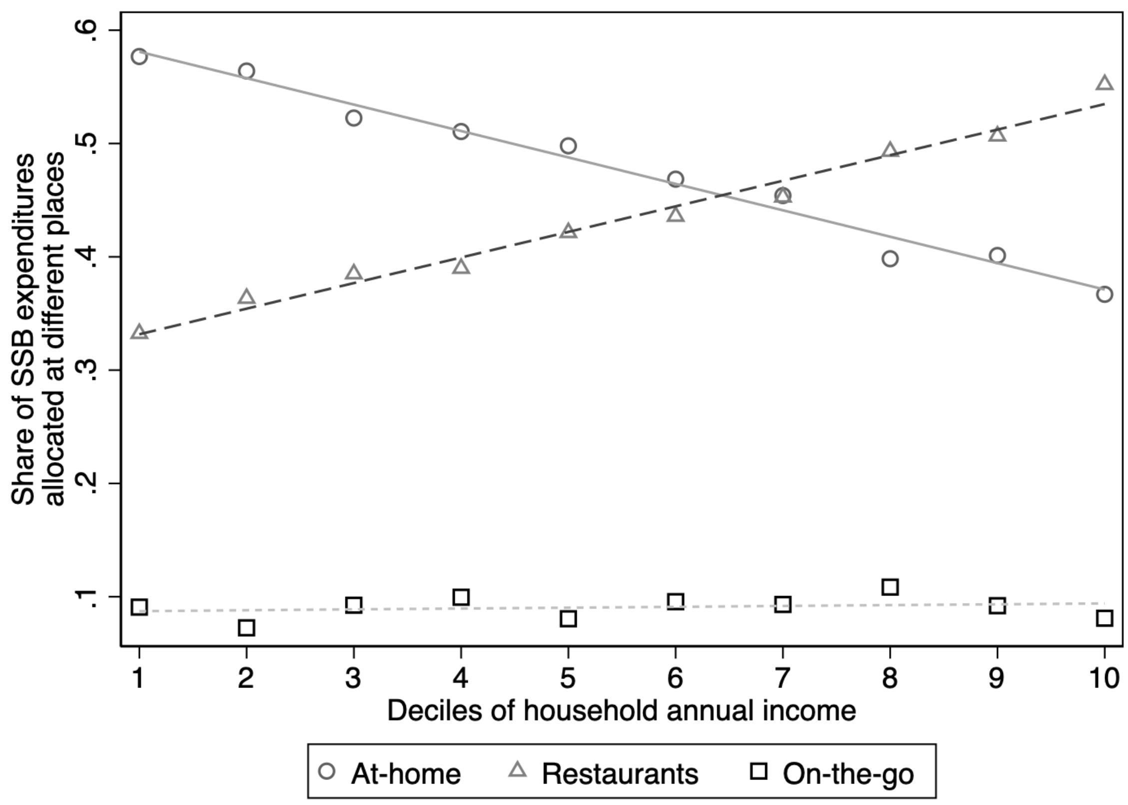


Figure 2: Average share of SSB expenditures with respect to income

2.3.2 SNAP Poor and Non-SNAP Poor

FoodAPS dataset has direct information on SNAP participation and income. They classify households into four types: SNAP participants, SNAP non-SNAP-participant poor (income < 100% of the Federal Poverty Guideline), medium (income >= 100% and income < 185% of FPG) and higher (income >= 185% of FPG) income households.[15] Table 4 reports the descriptive statistics for each group. Compared to SNAP participants, the non-SNAP poor have much lower household income, slightly older primary respondent, and fewer children.

First, in terms of the SSB consumption, we find that SNAP households have the largest volume consumption, SSB expenditures, and total grams of added sugar from SSB. In contrast, non-SNAP-participant poor have the lowest value for all the previous three variables. In other words, SNAP participants eat the least healthy while SNAP non-SNAP-participant poor eat the most healthy, even compared to higher income households. This finding contradicts with previous evidence that low-income households in general eat less healthy than higher income households. However, this evidence is possibly driven by the fact that a large fraction of low income households are SNAP participants and SNAP benefits cover soft drink purchases. Previous literature find that SNAP households' shopping cart consist of lots of soda.[16]

Second, non-SNAP rich households turn out to have the second highest SSB consumption in volume and expenditures, as well as total added sugar from SSB. This finding is also very

[15] As reported by the Census, in 2019, 17% of those who were eligible for SNAP benefits did not participate in the program (see https://www.census.gov/library/stories/2021/02/demographic-snapshot-not-everyone-eligible-for-food-assistance-program-receives-benefits.html). The finding that non-SNAP poor households have lower income than SNAP participants is consistent with a previous study that documents SNAP patterns using CPS data (Gundersen (2021)).

[16] See O'Connor (2017) https://www.nytimes.com/2017/01/13/well/eat/food-stamp-snap-soda.html

different from previous analyses that only look at consumption at-home. They normally find rich households to spend less on SSB. Rich households turn out to eat healthy at-home but unhealthy away-from-home. This finding highlights the importance of accounting for away-from-home consumption in order to evaluate households' overall diet quality and sugar intake.

Table 4: Descriptive Statistics by SNAP Status

	SNAP		Non-SNAP Poor		Non-SNAP Medium		Non-SNAP Rich	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Equivalized household annual income	12725.10	11461.10	6863.13	3208.24	15261.88	2634.37	43886.84	29479.16
Age of the primary respondent	42.92	15.51	44.95	16.96	48.10	19.25	47.33	16.59
Number of children	1.36	1.47	0.91	1.36	0.89	1.32	0.67	1.03
Household size (adult equivalents)	3.23	1.77	2.78	1.72	2.72	1.70	2.67	1.37
Total grocery expenditure	114.29	109.26	85.36	81.43	89.57	78.74	124.07	109.95
Total SNAP EBT amount reported	60.85	89.36	0.00	0.00	0.00	0.00	0.00	0.00
SSB volume consumption (ounce)	65.39	100.17	45.94	76.33	49.95	75.91	51.38	88.30
SSB expenditure (dollar)	2.68	3.64	1.98	2.51	2.18	2.77	2.64	3.41
Added sugar from SSB (gram)	184.54	282.57	129.65	214.08	140.07	212.77	141.35	254.04
Observations	1536		318		820		2009	

Notes: Grocery expenditure is the total dollar amount a household spent during a one week period in stores for at-home consumption. SNAP EBT amount reported is the total SNAP EBT amount reported by respondent or corrected by value observed on receipt when SNAP EBT payment is used for acquisition. SSB volume consumption, SSB expenditure and added sugar from SSB are measured at a per adult equivalent level.

Figure 3 provides a breakdown of the SSB consumption by places. It shows that SNAP households consume the highest ounces of SSB per adult equivalent per week at-home. Non-SNAP poor have similar SSB consumption at-home as higher income households. In terms of SSB consumed in restaurants, higher income households have higher ounces purchased. These findings suggest that the main difference in SSB consumption between SNAP and non-SNAP poor is driven by at-home purchases. Again, this can be driven by the fact that SNAP benefits can be spent on soft drinks.

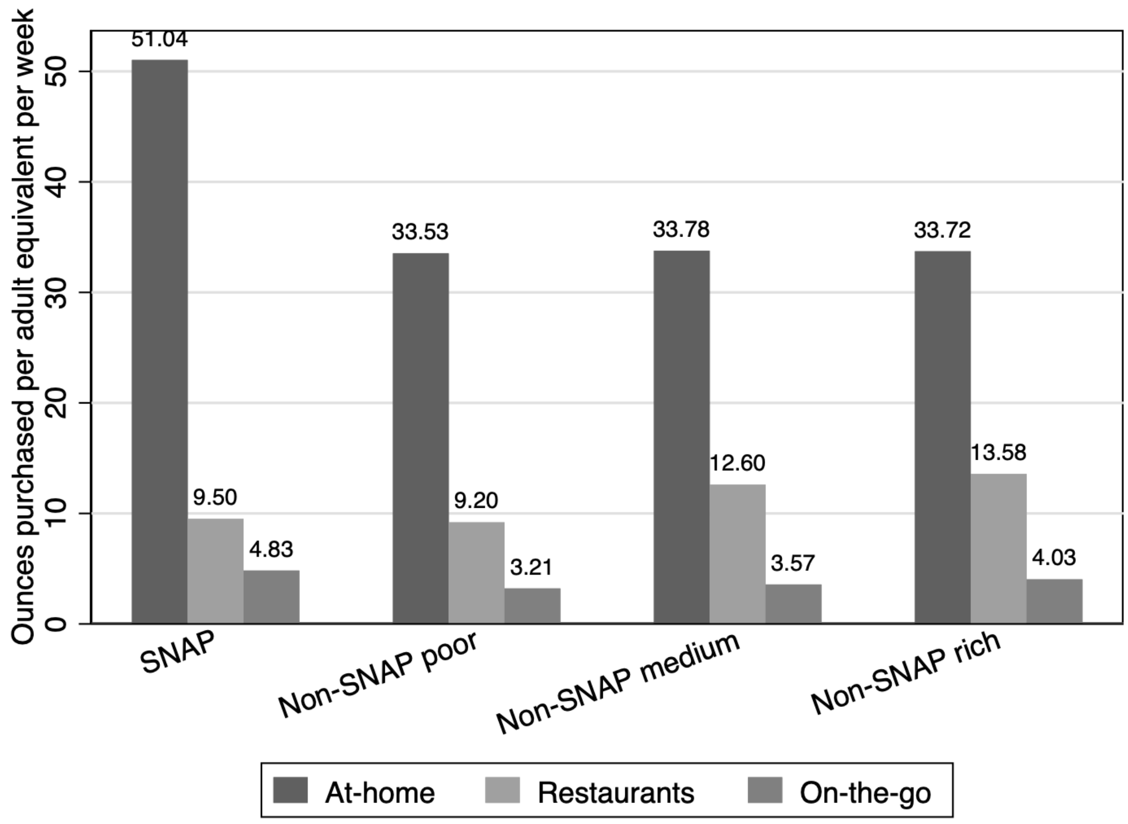


Figure 3: SSB volume consumption with respect to targeted group

2.3.3 Age, Existence of Children, and Overall Added Sugar

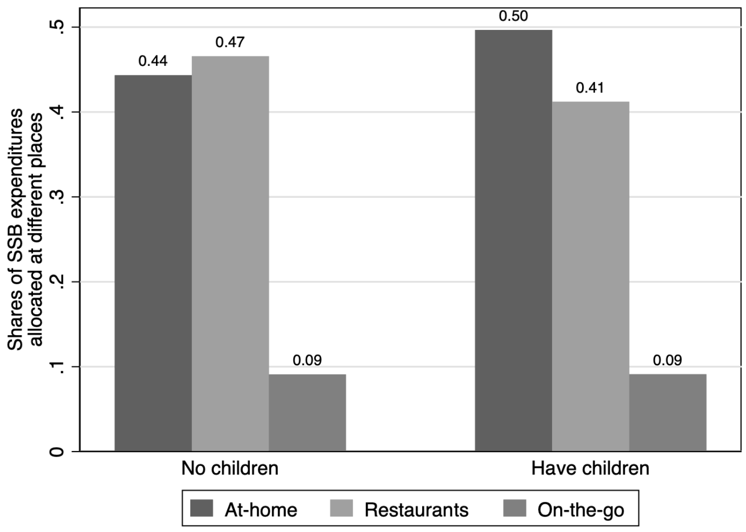
Figure 4 shows the average share of SSB expenditures at-home, on-the-go, and in restaurants by whether households have children, the age of the primary respondent, and overall added sugar from SSB.

First, compared with households without children, households with children spend slightly larger shares (50%) of SSB at-home and slightly smaller shares (41%) of SSB in restaurants. Both groups spend 9% of SSB expenditure on-the-go. Second, there is an increasing trend of the age of the primary respondent with respect to SSB shares at-home, while a decreasing trend of that with respect to SSB shares on-the-go. There is barely any significant relationship between SSB shares in restaurants and primary respondent's age. Lastly, on average, households with a higher sugar intake are more likely to spend SSB budget shares at home while less likely to spend SSB shares in restaurants. Combined with previous evidence that lower-income households consume more at home than in restaurants, the finding here simply reflects that poor households spend larger expenditure shares at home, eat less healthy diet, tend to have higher sugar intake, and are more likely to purchase SSB at home. There is little variation in SSB purchases on-the-go across groups of overall added sugar intake.

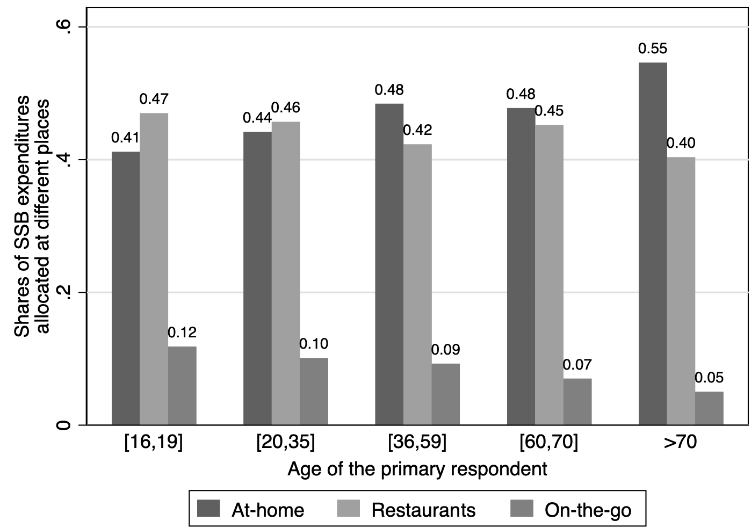
2.3.4 Demographics and SSB Consumption in Volume

Figure 5 plots the average purchased volume of SSB per adult equivalent per week with respect to household income and the age of the primary respondent.

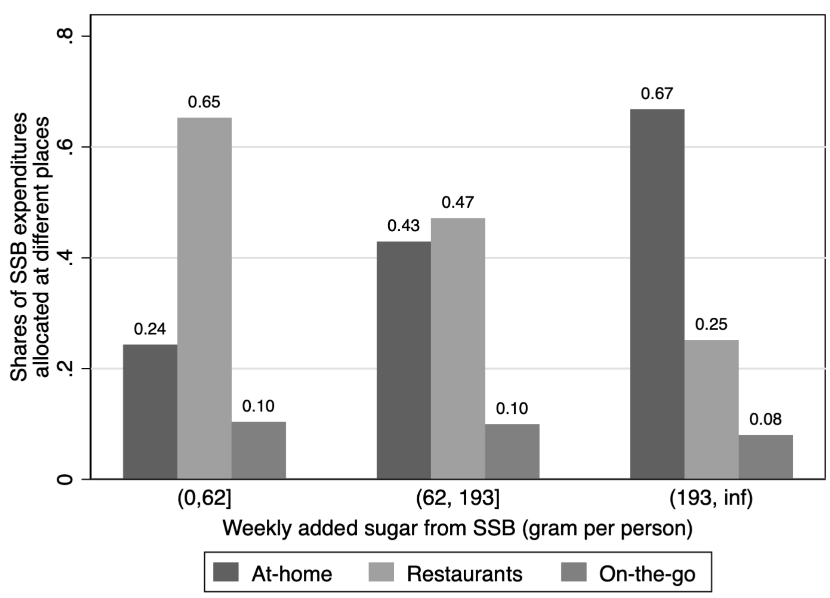
Similar to Figure 2, income and SSB volume consumption are still negatively correlated.



(a) By households with or without children



(b) By age of the primary respondent



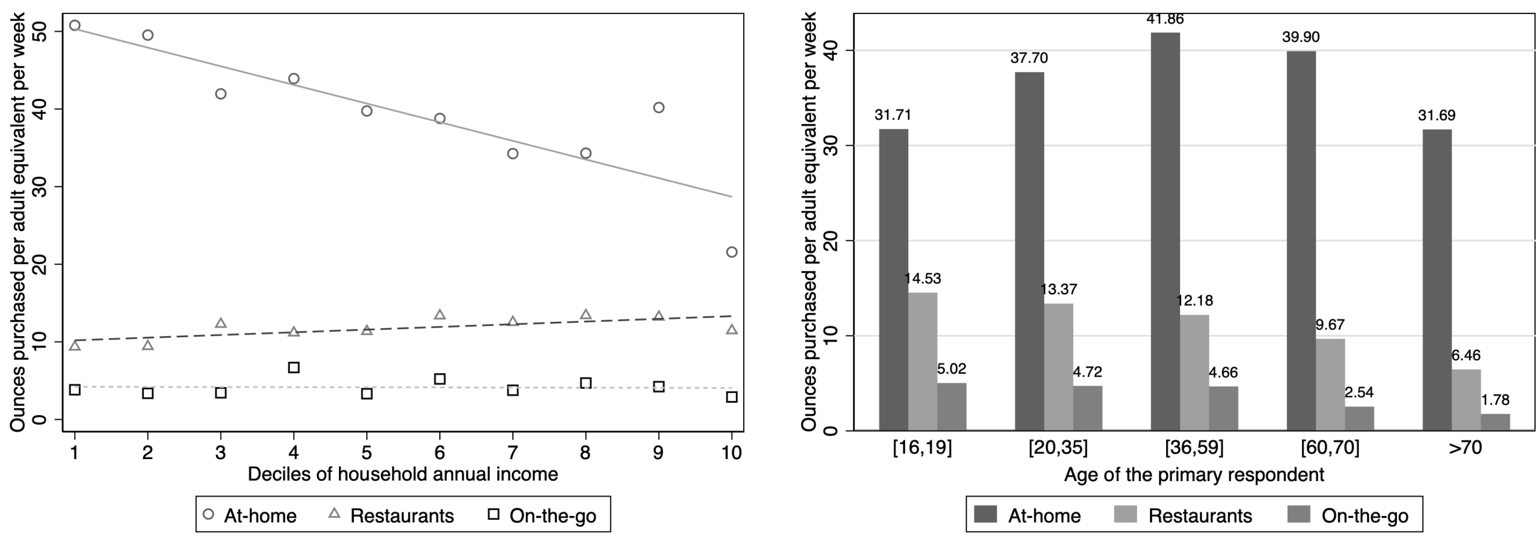
(c) By total added sugar from SSB

Figure 4: Average share of SSB expenditures allocated at-home, on-the-go, and in restaurants

Notes: The figure shows how average share of SSB expenditures allocated at different places varies across households with or without children, age groups, and total added sugar from SSB. In plot (b), age groups are classified according to the same cutoffs as in the FoodAPS dataset. Plot (c) is restricted to households who have positive amount of added sugar from SSB. The cutoff levels are the terciles of the total sugar from SSB.

However, different from Figure 2, there is much smaller in magnitude the increasing trend between income and SSB volume consumption in restaurants. It implies that the steeper positive trend between income and SSB budget shares in restaurants in Figure 2 is mainly due to the price effect. That is, richer households buy slightly more SSB products in restaurants but much more expensive products there.

Panel (b) in Figure 5 turns out to be very different from Figure 4. First, across all age groups, the volume consumption of SSB in restaurants is much lower than that at-home. Combined with the similar budget shares of SSB at-home and in restaurants in Figure 5, this finding implies that prices should be much higher for SSB products in restaurants than at-home. Second, there is an inverted U-shape relationship between the age of the primary respondent and SSB volume purchased at-home. In other words, middle age families buy the most SSB at-home. There is a decreasing trend between age and SSB volume purchased in restaurants. This is consistent with the evidence found in Martin *et al.* (2020).



(a) By deciles of household annual income

(b) By age of the primary respondent

Figure 5: SSB volume consumption at-home, on-the-go, and in restaurants

Notes: The figure shows weekly SSB purchases (in oz.) per adult equivalent at different places varies by distribution of household equivalized income and age groups. In plot (a), the household annual income is equivalized by the OECD-modified equivalence scale. In plot (b), age groups are classified according to the same cutoffs as in the FoodAPS dataset.

2.4 Composition

Table 5 shows the average share of non-alcoholic beverage expenditures allocated at each type of products. For SSB purchased at-home, the top three purchased products, measured by shares of total beverage expenditures, are soft drinks (26.7%), fruit and vegetable juice (18.1%), and fruit drinks (17.7%). For SSB purchased away-from-home, the top three purchased products are soft drinks (42.6%), sweetened coffee and tea (18.9%), and unsweetened coffee and tea (12.7%). In total, SSB (the first four types) accounts for 54.4% of beverage expenditures in at-home market segment and it accounts for 72.4% in away-from-home purchases.

Tables 6-9 show the share of beverage expenditures in different products at-home or away-from-home by demographic groups. We find that different demographic groups have very different basket of goods under at-home and away-from-home categories. For example, Table 6 shows that households with no children purchase more diet drinks at-home and unsweetened coffee and tea away-from-home than households with children. This suggests that households with children have stronger taste for sugary drinks. Table 7 shows that demand differ by age, and different patterns for at-home and away-from-home products. For at-home segment, younger households purchase more soft drinks and fruit drinks at-home while older households purchase more fruit and vegetable juice. For away-from-home segment, all households like soft drinks but older households don't purchase much fruit drinks as younger households and tend to drink more unsweetened coffee and tea. Table 8 shows that households with different level of sugary diet also have different baskets. The higher the total added sugar from SSB a household has, the more soft drinks and fruit drinks this household purchases. Instead, low sugary diet households purchase more diet drinks and water at-home. They also purchase more sweetened and unsweetened coffee and tea away-from-home than high sugary diet households.

Table 5: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products

	At-home	Away-from-home
Soft drinks	0.267	0.426
Fruit drinks	0.177	0.083
Sport and energy drinks	0.054	0.026
Sweetened coffee and tea	0.046	0.189
Diet drinks	0.098	0.065
Fruit and vegetable juice	0.181	0.022
Unsweetened coffee and tea	0.010	0.127
Flavored milk	0.029	0.036
Flavored and enhanced water	0.034	0.004
Water	0.105	0.022

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across all households.

Table 6: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products, by Households with or without Children

	At-home		Away-from-home	
	No Children	Have Children	No Children	Have Children
Soft drinks	0.241	0.290	0.386	0.465
Fruit drinks	0.149	0.203	0.072	0.093
Sport and energy drinks	0.044	0.064	0.022	0.029
Sweetened coffee and tea	0.052	0.040	0.197	0.181
Diet drinks	0.119	0.078	0.082	0.048
Fruit and vegetable juice	0.208	0.155	0.020	0.025
Unsweetened coffee and tea	0.013	0.007	0.166	0.091
Flavored milk	0.026	0.032	0.033	0.039
Flavored and enhanced water	0.042	0.026	0.003	0.006
Water	0.105	0.105	0.020	0.024

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across households with or without children.

Table 7: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products, by Age of the Primary Respondent

	At-home					Away-from-home				
Age group	[16,19]	[20,35]	[36,59]	[60,70]	>70	[16,19]	[20,35]	[36,59]	[60,70]	>70
Soft drinks	0.333	0.274	0.282	0.228	0.190	0.400	0.477	0.419	0.360	0.313
Fruit drinks	0.307	0.197	0.178	0.121	0.168	0.145	0.086	0.089	0.057	0.053
Sport and energy drinks	0.024	0.064	0.056	0.037	0.030	0.045	0.038	0.024	0.007	0.001
Sweetened coffee and tea	0.008	0.048	0.043	0.060	0.035	0.220	0.180	0.193	0.194	0.184
Diet drinks	0.029	0.077	0.099	0.135	0.129	0.035	0.042	0.070	0.094	0.102
Fruit and vegetable juice	0.171	0.172	0.158	0.242	0.271	0.044	0.020	0.024	0.020	0.021
Unsweetened coffee and tea	0.003	0.012	0.009	0.013	0.002	0.085	0.095	0.112	0.220	0.270
Flavored milk	0.000	0.030	0.029	0.025	0.037	0.013	0.032	0.041	0.028	0.039
Flavored and enhanced water	0.042	0.025	0.038	0.040	0.028	0.000	0.004	0.006	0.002	0.001
Water	0.083	0.102	0.109	0.099	0.110	0.014	0.025	0.022	0.018	0.016

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across households of different age groups.

Table 8: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products, by Total Added Sugar from SSB

	At-home			Away-from-home		
Total added sugar level	Low	Medium	High	Low	Medium	High
Soft drinks	0.168	0.268	0.409	0.436	0.472	0.492
Fruit drinks	0.155	0.208	0.214	0.089	0.094	0.088
Sport and energy drinks	0.052	0.075	0.057	0.020	0.032	0.032
Sweetened coffee and tea	0.041	0.044	0.064	0.233	0.195	0.189
Diet drinks	0.142	0.074	0.053	0.052	0.047	0.048
Fruit and vegetable juice	0.210	0.172	0.087	0.020	0.017	0.017
Unsweetened coffee and tea	0.011	0.011	0.008	0.101	0.089	0.073
Flavored milk	0.046	0.022	0.018	0.030	0.030	0.036
Flavored and enhanced water	0.048	0.025	0.022	0.003	0.003	0.006
Water	0.128	0.101	0.067	0.015	0.019	0.018

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across households of different levels of total added sugar from SSB.

Table 9: Average Share of Non-alcoholic Beverage Expenditures Allocated at Different Products, by Household Annual Income

	At-home			Away-from-home		
Income level	Low	Medium	High	Low	Medium	High
Soft drinks	0.339	0.276	0.192	0.492	0.439	0.369
Fruit drinks	0.167	0.182	0.179	0.081	0.080	0.087
Sport and energy drinks	0.048	0.054	0.060	0.025	0.029	0.023
Sweetened coffee and tea	0.043	0.047	0.046	0.152	0.189	0.212
Diet drinks	0.074	0.095	0.122	0.055	0.064	0.072
Fruit and vegetable juice	0.161	0.182	0.196	0.033	0.019	0.019
Unsweetened coffee and tea	0.010	0.009	0.010	0.108	0.113	0.155
Flavored milk	0.027	0.023	0.040	0.029	0.040	0.036
Flavored and enhanced water	0.020	0.031	0.048	0.005	0.005	0.003
Water	0.109	0.101	0.107	0.020	0.022	0.023

Notes: The table shows how share of non-alcoholic beverage expenditures are allocated at ten product types for at-home and away-from-home, respectively. The expenditure shares are averaged across households of different income groups.

2.5 Summary

To summarize, prices of SSB are the highest in restaurants, middle on-the-go, and the lowest at-home. Higher income households purchase more expensive soda drink in all three segments and buy drinks in restaurants more frequently. SNAP participants buy more SSB compared to nonparticipant poor, whose spending on SSB is even lower than higher-income households. Heavy sugar consumers tend to be lower income households and purchase less expensive soda drinks in all three segments. Once we account for drinks away-from-home, the regressivity concern of soda taxes becomes less serious because higher income households will be taxed more in restaurants while lower income households will be taxed more at-home. Who bares the most tax burden is ambiguous unless we have a demand model that accounts for all three segments.

Previous literature normally focus on the poor, children, and heavy sugar consumers as the main targets of soda taxes . For these three types of households, restaurants purchases seem to be less important for them compared to at-home purchases. However, these households still buy certain amount of soda drinks away-from-home. We may underestimate the effect of taxes on their total SSB demand if we ignore away-from-home purchases. On the other hand, they can have different price elasticities in the at-home and away-from-home segment. For example, if they are more price sensitive in the at-home segment, then we may overestimate the effectiveness of taxes because they may still buy a certain part of SSB from the away-from-home segment.

The other main critical benefit of including restaurants data in the analysis is the finding that higher income households will also be largely affected by an SSB taxes because they purchase more SSB, and potentially more expensive SSB, in restaurants. This finding suggests that the regressivity concern of an SSB taxes may be less serious than we expect. We will calculate the regressivity in the next chapters by predicting households' counterfactual SSB expenditure shares in terms of total expenditures on food and drinks in the three segments when we simulate tax incidences.

3 Model and Estimated Coefficients

In this section we estimate a structural model of non-alcoholic beverage demand. We employ a random coefficients nested logit model to create more flexible substitution patterns.[17] We estimate the random coefficient flexibly following Fox *et al.* (2011) and Fox *et al.* (2016), for two reasons: First, Dubois *et al.* (2020) finds that unobserved preferences might not be fully captured by specifying a priori the distribution of random coefficients. Thus, we relax the often-made assumption of normal distribution in BLP applications. Second, the method of Fox *et al.* (2011) and Fox *et al.* (2016) has the advantage of achieving a flexible distribution while maintaining computational tractability. We then use the model to evaluate counterfactual tax policies that could reduce SSB consumption.

3.1 A Model of Non-Alcoholic Beverage Demand

We index consumers by $i \in \{1, \cdots, N\}$. Each consumer visits a retailer $r \in \{1, \cdots, R\}$ at time t and makes a transaction or incurs choice occasion $\tau \in \{1, \cdots, \mathcal{T}\}$. Let $r(\tau)$ and $t(\tau)$ denote the specific retailer and time that the consumer visits the retailer. Notice that the retailer here can be grocery stores, convenience stores, vending machines, or restaurants.

We index the non-alcoholic beverage products by $j \in \{1, \cdots, J\}$, as those defined in Table 3. The construction of the product-level price is discussed in the prior section.

[17]See Grigolon & Verboven (2014) for a comprehensive comparison between the random coefficients nested logit, random coefficients logit and nested logit models.

When making a decision, the choice set facing consumers contains purchasing options that are available to the consumers on each specific trip. This means that when a consumer visits a grocery store, she only considers drinks available at the store. Similarly, on a trip to the food at-home place, the choice sets facing the consumer only include food at-home drink products. We denote the choice set by $\Omega_{r(\tau)}$.

We allow for the possibility that a consumer instead chooses either other non-beverage products like meat or snacks in a store, or purchases a meal without ordering any drinks in a restaurant. We refer to these as "outside options". We indicate outside options by $j = 0$, and the choice set $\Omega_{r(\tau)}$ includes the outside option. [18]

We partition the choice set $\Omega_{r(\tau)}$ to two disjoint subsets denoted by C_0 and C_1. They are also called nests. C_0 is the nest of outside options. C_1 is the nest of all available products in the choice set. The indirect utility of a consumer i on choice occasion τ from product j in the nest C_g where $g \in \{0, 1\}$ is given by

$$U_{ij\tau} = V_{ij\tau} + \varepsilon_{ij\tau} \tag{1}$$

where

$$V_{ij\tau} = \alpha_i p_{jr(\tau)t(\tau)} + \eta_i s_j + \mathbf{x}'_{ij}\beta, \tag{2}$$

and the utility obtained from choosing the outside options is

$$U_{i0\tau} = \varepsilon_{i0\tau}. \tag{3}$$

The term $p_{jr(\tau)t(\tau)}$ denotes the price of product j, which varies over time t and across retailer types r. The variable s_j is an indicator variable of an SSB product. $\mathbf{x}_{ij}$ is a vector of observed product characteristics (including the constant term) and their interactions with household demographics. Specifically, $\mathbf{x}_{ij}$ include package size (measured in ounces), an indicator variable of products in pack, indicator variables of drink categories, time fixed effects, and retailer-drink category fixed effects.[19] $\varepsilon_{ij\tau}$ is an error term following the generalized extreme value distribution, with cumulative distribution of the following form

$$\exp\left(-\sum_{g=0}^{1}\left(\sum_{j\in C_g} e^{-\varepsilon_{ij\tau}/\lambda}\right)^{\lambda}\right),$$

which gives rise to the nested logit structure. For this distribution of $\varepsilon_{ij\tau}$, the idiosyncratic error terms are correlated within a given nest. For any two products belonging to different nests, the error terms are uncorrelated. The nest parameter λ measures dissimilarity among products within a nest. A value of $\lambda = 1$ indicates that $\varepsilon_{ij\tau}$ are uncorrelated within nests and the model degenerates to the standard logit model. As λ decreases, the correlation within nests rises. The nested logit assumption implies that products in the same nest are closer substitutes than

[18] Our definition of outside options is most close to that of Marshall (2015), assuming the outside option is chosen when a shopping trip is observed with no purchase of any inside goods.

[19] Retailers are defined in Table 2. Drink categories are defined in Table 5.

products in different nests. In our context, the two nests are the outside options and all products in the choice set, respectively. This modeling assumption implies that all available products are considered closer substitutes than the outside option.

Allowing for preference heterogeneity is essential in capturing realistic demand features. The demand model here is flexible in that it incorporates preference heterogeneity through two aspects. The first is through the idiosyncratic error component $\varepsilon_{ij\tau}$, as we previously discussed. The second is through the taste heterogeneity for product attributes. Specifically, we let taste parameters like α_i and η_i vary by household observed and unobserved characteristics. We define the marginal (dis)utility of price and taste for SSB as the following[20]

$$\alpha_i = \alpha_0 + \mathbf{v}_i'\alpha_1 + \mu_i \quad \eta_i = \mathbf{v}_i'\eta_1,$$
$$\text{where } \mu_i \sim F(\mu).$$

$\mathbf{v}_i$ denotes observed household demographics, and μ_i is a random coefficient and captures unobserved preference related to prices. By allowing α_i and η_i to depend on household characteristics, we allow different consumers to have different price sensitivity when making purchases for at-home and away-from-home consumption and different tastes for SSB products.

The observed household demographics $\mathbf{v}_i$ include a joint variable of household income and SNAP participation, the age of the primary respondent, whether the household has children, and the household's overall sugar intake.[21] The unobserved household characteristics, or random coefficient μ_i can include individual household information that affects the purchasing decision, yet unobservable to econometricians.[22] Prior empirical work on random coefficients logit model usually make parametric assumptions on the distribution of the random coefficient $F(\mu)$, e.g., a normal distribution. We will relax this assumption and estimate $F(\mu)$ using a fixed grid approach, following Fox *et al.* (2011) and Fox *et al.* (2016), to allow for more flexible unobserved taste heterogeneity.[23]

Conditional on the unobservable μ_i, the joint probability of a consumer choosing a product j is

$$P_{ij\tau}(\mu_i) = P_{ij\tau|C_g}(\mu_i) P_{iC_g\tau}(\mu_i),$$

where $P_{ij\tau|C_g}(\mu_i)$ is the probability of choosing a product j conditional on a product in the nest C_g being chosen. $P_{iC_g\tau}(\mu_i)$ is the marginal probability of choosing a product in the nest C_g, conditional on the unobservable μ_i. As shown in McFadden (1978), the joint probability of

[20] We do not put a constant term η_0 in the η_i function because $\mathbf{x}_{ij}$ contains indicator variables of drink categories, which leads to the mean level of the taste coefficient for SSB η_0 unidentified due to perfect collinearity problem.

[21] The correlation between "SNAP households" indicator and "households with children" indicator is only -0.2 and hence we include both indicators in the model. We also find that the average age of children is similar across sugar intake groups and hence we do not include it in the model.

[22] Following Bonnet & Dubois (2010) and Eizenberg & Salvo (2015), among others, we only have a random coefficient on price.

[23] Birchall & Verboven (2022) and Miravete *et al.* (2022) highlight that the functional form of the indirect utility function implicitly imposes restrictions on demand curvature. Miravete *et al.* (2022) specifically suggest that incorporating a flexibly distributed price random coefficient not only offers greater flexibility in substitution patterns, but also allowing for a wider range of estimable demand curvature.

choosing product $j \in C_g$ (conditional on μ_i) takes the nested logit formula

$$P_{ij\tau}(\mu_i) = \frac{e^{V_{ij\tau}(\mu_i)/\lambda} \left(\sum_{k \in C_g} e^{V_{ik\tau}(\mu_i)/\lambda}\right)^{\lambda - 1}}{\sum_{l=0}^{1} \left(\sum_{k \in C_l} e^{V_{ik\tau}(\mu_i)/\lambda}\right)^{\lambda}}. \tag{4}$$

The unconditional probability of consumer i choosing product $j \in C_g$ in a choice occasion τ is

$$P_{ij\tau} = \int \frac{e^{V_{ij\tau}(\mu)/\lambda} \left(\sum_{k \in C_g} e^{V_{ik\tau}(\mu)/\lambda}\right)^{\lambda - 1}}{\sum_{l=0}^{1} \left(\sum_{k \in C_l} e^{V_{ik\tau}(\mu)/\lambda}\right)^{\lambda}} dF(\mu). \tag{5}$$

3.2 Identification

Our identification is similar to Dubois *et al.* (2020) except that we only know retailer types as those in Table 2 rather than specific retailers. We do not know brand information of products and hence we don't have the price variation across brands.

The main identification challenge is to isolate the causal effect of price on demand for at-home and away-from-home products. That is, the parameter vector α. We rely on two sources of variation to identify the price effects. First, conditional on time and retailer-drink type effects, we exploit the variation in prices of the same product in different retailer types across time. The identification assumption is that consumers do not choose retailers when they make consumption choices for a specific product. Instead, retail choices are more driven by convenience factors like distance to school and workplace. Second, we utilize price variation for the same product in the same retailer at the same time but across different containers and sizes.

To address potential endogeneity bias arising from the correlation between the unobserved error term $\varepsilon_{ij\tau}$ and prices, we control for time and retailer-drink category fixed effects. These fixed effects capture demand shocks that vary at these levels, which could have driven price fluctuations. By including these fixed effects, we capture most of the unobservable factors, leaving only time-varying shocks specific to the product-retailer combination. The timing assumption we adopt is that such demand shocks are realized after firms make pricing decisions. Therefore, the inclusion of a rich set of fixed effects in our demand model should mitigate any bias resulting from unobservable factors that might be correlated with prices.

3.3 Estimation

Following recent literature, we estimate the at-home and away-from-home segment separately and obtain distinct preferences parameters for these two segments.[24] We refer to the method of Fox *et al.* (2011) and Fox *et al.* (2016) to estimate the random coefficients. The method has the

[24] The other papers that estimate different food segments separately include O'Connell & Smith (2020) and Dubois *et al.* (2020), both studying the at-home and on-the-go segments. They follow Browning & Meghir (1991) to test for non-separabilities between segments and find no evidence of demand dependence between the two segments. Following them, we also conduct a separability test (see Section 3 in the Appendix). The results support separability between at-home and away-from-home soft drinks consumption. Meanwhile, we have also tried to estimate the two segments jointly. The computation is burdensome due to the large choice set included.

advantage of achieving a flexible distribution, while maintaining computational tractability.

Consider a fixed grid $\mathcal{M}_R = (\mu^1, \cdots, \mu^R)$, where R represents the number of grid points. One can interpret R as the number of discrete household types.[25] We assume each μ_i is a draw from the set of values $(\mu^1, \cdots, \mu^R)$ and each grid point μ^r occurs with probability γ^r, for $r = 1, \cdots, R$. Given the choice of $\mathcal{M}_R$, we estimate the weights $\gamma = (\gamma^1, \cdots, \gamma^R)$ on the grid points. We impose the constraints $0 \leq \gamma^r \leq 1, \forall r$, and $\sum_{r=1}^{R} \gamma^r = 1$, such that μ has a well-defined distribution.

For each household type r, we can rewrite choice probability (4) by replacing μ_i with μ^r:

$$P_{ij\tau}(\mu^r) = \frac{e^{V_{ij\tau}(\mu^r)/\lambda} \left(\sum_{k \in C_g} e^{V_{ik\tau}(\mu^r)/\lambda} \right)^{\lambda - 1}}{\sum_{l=0}^{1} \left(\sum_{k \in C_l} e^{V_{ik\tau}(\mu^r)/\lambda} \right)^{\lambda}}. \tag{6}$$

Let $\alpha = (\alpha_0, \alpha_1)$ and $\eta = \eta_1$. Denote by $\theta = (\alpha, \beta, \eta, \gamma)$ the vector of preference parameters. Using choice probabilities defined above, we calculate the likelihood function defined by

$$\mathcal{L}(\theta) = \sum_i \sum_\tau \sum_{j \in \Omega_{r(\tau)}} d_{ij\tau} \log \left(\sum_{r=1}^{R} \gamma^r P_{ij\tau}(\mu^r) \right), \tag{7}$$

where θ is the vector of parameters to be estimated and $d_{ij\tau}$ is an indicator variable equal to one if consumer i chose product j on choice occasion τ and zero otherwise. The parameters are estimated using maximum likelihood estimator.

3.4 Estimated Coefficients and Elasticities

In Table 10 and 11, we summarize the estimated at-home and away-from-home preference parameters obtained by maximizing the likelihood function (equation (7)).

We interact the price and the taste for SSB coefficient with four demographic variables. They are income and SNAP participation joint variable (SNAP participants, nonparticipants low income households, medium, and high income households), the age of the primary respondent, an indicator variable for households with children, and household total added sugar intake (from SSB) groups (zero sugar households, low, medium, and high sugar consuming households). So the baseline group in the coefficient of α and η is SNAP participants without children with zero sugar intake.

We start from consumers' disutility of price α. For the at-home segment (Table 10), the estimated price parameter is -0.285 for the baseline households. Non-SNAP low income and medium income households are even more price sensitive than the baseline households, as suggested by the interaction coefficients -0.074 and -0.038. In contrast, the positive interaction coefficient 0.072 of high income households implies that they are less price sensitive compared to the baseline group.

For the away-from-home segment (Table 11), the baseline group has a negative marginal disutility of price (-0.515). Nonparticipants low income households are slightly even more price

[25] In our estimation, we varied the choice of R and the results did not change qualitatively. The range of $\mathcal{M}_R$ is chosen to be of a similar magnitude to the price coefficients estimated from an initial regression of plain logit model without random coefficients.

sensitive than participants (-0.031). Medium and high income households are both less price sensitive than the baseline group (0.079 and 0.13).

The effect of age on the marginal disutility of price is -0.009 at home and 0.041 away-from-home. The interpretation is that a one year increase in the age of the primary respondent is associated with an increase of disutility of price by 0.009 at-home and a decrease of that by 0.041 away-from-home. Households with children are less price sensitive both at-home (0.056) and away-from-home (0.026) than the baseline group.

Lastly, there is a decreasing relationship between price sensitivity and household total sugar intake from SSB at-home. The coefficients of the interaction between price and household sugar intake is 0.044 for low sugar consumers, 0.138 for medium sugar consumers, and 0.257 for high sugar consumers. In other words, The difference in the marginal disutility of price between high and low sugar households is 0.21 (0.257 minus 0.044). In contrast, the difference becomes very small across low and high sugar taking households away-from-home (ranging from 0.13 to 0.147 for low and higher sugar households).

We next look at consumers' taste for SSB, η. First, we find that there is a decreasing relationship between the taste for SSB and household income, both at-home and away-from-home. This is suggested by the negative coefficients of the interaction between SSB and household income and SNAP participation status (-0.017 to -0.481 at-home and -0.093 to -0.164 away-from-home). Comparing SNAP participants and nonparticipant poor, we find that the latter has weaker taste for SSB, both at-home (-0.017) and away-from-home (-0.093). SNAP participants turn out to have the strongest taste for SSB no matter at-home or away-from-home. This is consistent with the empirical evidence that SNAP participants buy a lot of SSB because SNAP benefits are allowed to be spent on SSB.[26]

Households with older primary respondent have a distaste for SSB both at-home (-0.032) and away-from-home (-0.099) compared to the baseline group. In contrast, households with children have a very strong taste for SSB at-home (0.284) and a relatively weaker but also a positive taste for SSB away-from-home (0.089) compared to the baseline group.

In terms of the relationship between the taste for SSB and household total sugar intake from SSB, we find that for the at-home segment, there exists a positive linear relationship between these two variables. The difference in the marginal utility from SSB between high and low sugar households is as large as 1.57. For the away-from-home segment, we find that across all sugar intake groups of households, they all have very large positive marginal utility from SSB (coefficients ranging from 4.889 to 5.548) compared to the at-home segment. High sugar consuming households still have slightly larger marginal utility of SSB compared to low sugar consumers.

We then look at the right panel of Table 10 and 11. Multi-pack has a negative impact on households marginal utility (-0.341) at-home. Package size has a positive effect (0.006) on households marginal utility at-home but a large negative effect away-from-home (-0.037). This

[26]The article "In the Shopping Cart of a Food Stamp Household: Lots of Soda" can be found on https://www.nytimes.com/2017/01/13/well/eat/food-stamp-snap-soda.html

finding is intuitive as we expect that consumers prefer small convenient package size on-the-go or in restaurants compared to in the grocery stores.

The nesting parameter λ is 1 in the at-home segment (Table 10). This implies that the nests are not significant and the demand model degenerates to a random coefficients logit model without nest structure. The interpretation is that when a specific drink product in the at-home segment (e.g., grocery store) becomes unavailable, the probabilities of choosing another drink product and choosing other grocery item such as meat and dairy would increase by the same proportion. The nesting parameter λ is 0.832 in the away-from-home segment (Table 11). The value implies that when consumers visit the away-from-home segment (e.g., restaurants), they are more likely to switch between drink products than to switch from drink products to a dish.

The price coefficients we obtained in both segments are quite small in magnitude compared to prior work.[27] However, this can be explained by two facts. The first is that the nesting parameter rescales all demand parameters. Grigolon & Verboven (2014) find that to make the price coefficient in a nested logit model comparable to the ones from standard logit model, the price parameter should be rescaled by α/λ. Thus, with a nesting parameter less than 1, the rescaled price coefficient should be larger in magnitude than those reported in Table 10 and 11. The second main reason is due to our definition of "products". Instead of choosing among specific brands or narrowly defined products, we define "products" in our analysis as drink categories such as regular soft drinks, large-bottle diet soft drinks, a pack of juice drink, etc. Furthermore, the magnitude of estimated coefficients varies depending on the unit of the variables. Therefore, it would be more meaningful to consider price elasticities rather than parameter values, which we will discuss next.

[27]Using supermarket data in a developing country, Marshall (2015) reports an average marginal (dis)utility of price of -6.15. Dubois *et al.* (2020) uses data in UK covering on-the-go purchases to study soda demand. Their estimate of mean level of price preference is -3.15,

Table 10: Random Coefficients Nested Logit Demand Estimates, At-Home

	Estimate	SE		Estimate	SE
α			β		
Price	-0.285	0.011	Constant	-2.791	0.048
Price × non-SNAP poor	-0.074	0.003	Multi-pack	-0.341	0.024
Price × non-SNAP med	-0.038	0.002	Package size	0.006	0.000
Price × non-SNAP rich	0.072	0.04			
Price × age	-0.009	0.002			
Price × child	0.056	0.003	λ		
Price × sugar low	0.044	0.008	Nesting parameter	1.000	0.008
Price × sugar med	0.137	0.008			
Price × sugar high	0.257	0.008	μ	Weight	
			-0.1	0.74	
η			-0.078	0.24	
SSB × non-SNAP poor	-0.017	0.029	-0.056	0.01	
SSB × non-SNAP med	-0.226	0.023			
SSB × non-SNAP rich	-0.481	0.027	Drink category FE	Yes	
SSB × age	-0.032	0.008	Retailer-drink category FE	Yes	
SSB × child	0.284	0.027	Time FE	Yes	
SSB × sugar low	1.014	0.059			
SSB × sugar med	1.778	0.052			
SSB × sugar high	2.584	0.051	Number of choice occasions	23384	

Notes: We estimate demand on a sample of 4,412 households on 23,384 At-home choice occasions. Consumers choose from the products in At-home segments including the outside options. The reference group is SNAP households that consumed zero added sugar from SSB within a week. The coefficients of interaction between price and other demographic groups represent the change relative to the baseline level. Non-SNAP income group indicators are constructed by ERS, using household income measures and adjusted by poverty guidelines. The level of sugary diet is constructed based on weekly total added sugar from SSB a household has. We include a random coefficient for price. We report three μ^r with the highest estimated weights.

Table 11: Random Coefficients Nested Logit Demand Estimates, Away-From-Home

	Estimate	SE		Estimate	SE
α			β		
Price	-0.515	0.031	Constant	-0.903	0.044
Price × non-SNAP poor	-0.031	0.007	Package size	-0.037	0.001
Price × non-SNAP med	0.079	0.006			
Price × non-SNAP rich	0.130	0.007			
Price × age	0.041	0.004			
Price × child	0.026	0.008	λ		
Price × sugar low	0.130	0.008	Nesting parameter	0.832	0.019
Price × sugar med	0.134	0.006			
Price × sugar high	0.147	0.008	μ	Weight	
			-0.1	0.79	
η			-0.011	0.10	
SSB × non-SNAP poor	-0.093	0.047	-0.078	0.05	
SSB × non-SNAP med	-0.020	0.035			
SSB × non-SNAP rich	-0.164	0.026	Drink category FE	Yes	
SSB × age	-0.099	0.008	Retailer-drink category FE	Yes	
SSB × child	0.089	0.024	Time FE	Yes	
SSB × sugar low	4.889	0.196			
SSB × sugar med	5.336	0.201			
SSB × sugar high	5.548	0.204	Number of choice occasions	23539	

Notes: We estimate demand on a sample of 3,977 households on 23,539 away-from-home choice occasions. Consumers choose from the products in away-from-home segments including the outside options. The reference group is SNAP households that consumed zero added sugar from SSB within a week. The coefficients of interaction between price and other demographic groups represent the change relative to the baseline level. Non-SNAP income group indicators are constructed by ERS, using household income measures and adjusted by poverty guidelines. The level of sugary diet is constructed based on weekly total added sugar from SSB a household has. We include a random coefficient for price. We report three μ^r with the highest estimated weights.

We report aggregate price elasticities for the at-home and away-from-home segment in Table 12 by demographic groups.[28] Specifically, we simulate a one percent increase in the price of all SSBs in any segment and then re-estimate the demand model to predict SSB (own demand effect) and non-SSB purchases (cross demand effect). The elasticities are then calculated as the change in quantity demand divided by the price change.

We find substantial differences in elasticities across demographic groups and between the two segments. In Panel A of Table 12, SNAP participants are more elastic away-from-home while all other three non-SNAP poor, medium, and higher income households are more elastic at-home. In Panel B, households without children are more elastic at-home and are more elastic than household with children. In Panel C, households across all levels of sugar intake from SSB are more elastic at-home. The low sugar intake households are the most elastic while the high sugar intake households are the least elastic. This pattern is consistent with findings from Dubois *et al.* (2020), who find that the soda taxes is less effective in targeting households with high total dietary sugar.

On one hand, our estimated elasticities for SSB at-home are of a similar magnitude to the existing literature on SSB demand with highly aggregated level of products. For example, Lopez

[28] We also show the product-level price elasticity for the at-home and away-from-home segment in Tables A2 to A5 in Appendix Section 5.

& Fantuzzi (2012) estimate an own price elasticity of -0.58 for all caloric carbonated soft drinks (these are SSBs in our paper but excluding juice, energy drink, and sweetened tea and coffee). Andreyeva *et al.* (2010) collect own price elasticities for soft drink categories from 14 studies. The mean own price elasticities across the 14 studies is -0.79, with a 95% confidence interval of $[-0.33, -1.24]$.[29]

On the other hand, the demand for an aggregate-level category defined as in Table 12 here would in general be less elastic than the demand estimated from more disaggregated brand-level individual products. For example, Bonnet & Réquillart (2013) report a brand specific own price elasticities to be between -2.13 and -3.95. Dubé (2005) estimates a brand-level own price elasticities ranging between -2 and -4. The reason is that compared to a broadly defined category, brand-level products have more competition and substitution across each other.

Table 12: Aggregate-Level Price Elasticity

	Effect of 1 percent increase in the price of SSB on			
	Own demand effect	Cross demand for non-SSB	Own demand effect	Cross demand for non-SSB
	At-Home		Away-From-Home	
	Panel A: SNAP status and income level			
SNAP	-0.267	0.130	-0.335	0.197
Non-SNAP poor	-0.473	0.161	-0.396	0.202
Non-SNAP medium	-0.374	0.126	-0.233	0.136
Non-SNAP rich	-0.233	0.063	-0.194	0.089
	Panel B: Households with or without children			
No children	-0.348	0.093	-0.248	0.098
Have children	-0.244	0.111	-0.243	0.150
	Panel C: High vs low sugar consumers			
Low	-0.620	0.091	-0.278	0.116
Medium	-0.393	0.137	-0.249	0.171
High	-0.163	0.119	-0.206	0.178

Notes: We simulate the effect of a 1 percent price increase for all SSB belonging to the at-home market segment and all SSB belonging to the away-from-home segment, respectively. In column 1, we report the change in demand for those products. In column 2 we report the effect on demand for non-SSB products. Elasticities are computed separately by demographic groups and are averaged across markets.

4 Effects of a soda taxes on Sugar Intake

We use our demand estimates to simulate the introduction of a tax levied on SSB. We consider a tax rate of 1 cent per ounce. This is similar to the level of tax under the U.S. Soft Drinks Industry Levy.[30]

Let Ω_{SSB} denote the set of SSB products, r a soda taxes rate and q_j the volume in ounce. We

[29] Their definition of soft drink categories are slightly different from ours. Their narrowest definition is carbonated soft drinks, and the broadest definition is non-alcoholic beverages.

[30] As of 2022, seven cities and counties in the United States have introduced an SSB taxes. The level of current excise taxes on SSB ranges from 1 to 2 cents per ounce, with five out of the seven cities being 1 cent per ounce.

assume the post-tax prices, p_j^{post} are given by

$$p_j^{post} = \begin{cases} p_j^{pre} + rq_j & \forall j \in \Omega_{SSB} \\ p_j^{pre} & \forall j \notin \Omega_{SSB} \end{cases}$$

We study the impact of the tax on household at-home and away-from-home sugar consumption. Our main results assume 100 percent pass-through of soda taxes for both segments, given the evidence of almost 100 percent pass-through.[31] We also try setting the pass-through rate of soda taxes for the away-from-home segment to be 70 percent given the empirical finding in Cawley *et al.* (2021) using restaurants data in Boulder Colorado. They are reported in the Appendix. The counterfactual results are quite similar in both settings.

4.1 The Effectiveness of an SSB Tax

Our tax simulations suggest that consumers who purchase SSB will, on average, lower the amount of sugar they purchase from SSB at-home by 14.39g per week, away-from-home by 2.56 per week, and in total by 15.72g.[32] The average percentage reduction is 18.12 percent at-home and 5.75 percent away-from-home. The distribution of reductions in sugar in total is right skewed with the seventy-fifth, ninetieth, and ninety-fifth percentiles being 20.48g, 34.80g, and 47.61g.

An important aspect about the effectiveness of an SSB tax is whether it successfully achieves the reductions in sugar amongest the targeted groups of consumers: low-income households, in particular, SNAP participants and SNAP-eligible nonparticipants, households with children, and those with high total weekly dietary sugar. In Figure 6, we show how the effect of tax vary across these demographic characteristics. Panel (a) to (f) show how the mean reduction in sugar and the percentage reduction in sugar varies across SNAP status and income level, households with or without children, and total weekly dietary sugar, separately for the at-home and away-from-home segment.

Panel (a), (c), and (e) show that the tax on sugary soft drinks achieves relatively large reductions in total sugar (at-home and away-from-home) among low-income households, households without children, and households with high weekly added sugar intake from SSB.

However, if we look at the at-home and away-from-home segments separately, we see very diverse patterns between the two segments. The reduction in sugar away-from-home is much smaller than that at-home, and it is small for all demographic groups. In other words, there is not much variation in the reduction in sugar away-from-home for all groups because the reduction

[31] The literature that estimate SSB tax pass-through rate includes Cawley & Frisvold (2017), Grogger (2017), Berardi *et al.* (2016), Bergman & Hansen (2019), and Falbe *et al.* (2015). They tend to find that taxes are fully shifted to consumers, or even overshifted. The most recent papers like Cawley *et al.* (2021) find a pass-through rates of 71.1% on taxed drinks in Boulder, Colorado using hand-collected retail store data and 74.2% using restaurant data. Marinello *et al.* (2021) find similar price increases (82%) of bottled regular soda and diet soda in fast-food restaurants in Oakland, CA.

[32] Note that the total reduction in level (15.72g) is lower than the sum of the reduction at-home (14.39g) and the reduction away-from-home (2.56g). This is because the at-home, away-from-home, and the total segment is calculated based on a sample of 4412 households, 3977 households, and 4683 households respectively. There are some households with no purchase in one of the two segments, and hence for them the total reduction in SSB only comes from the other segment where they have made purchases.

is small: only around 3g per adult equivalent per week or 156g per adult equivalent per year. The percentage reduction in sugar is around 6 percent for all demographic groups (panel (a) and (b)).

The small variation in the effect of SSB tax on sugar intake away-from-home is supported by our previous descriptive findings in section 2.3. Households across income groups have similar share of SSB expenditures on-the-go. However, high-income households have much higher share in restaurants compared to low-income households. Meanwhile, the elasticity estimates in Table 12 show that high-income households are less price sensitive and hence their reduction in sugar in restaurants given the soda taxes is low. Both results lead to the finding here that the reduction in sugar of the high-income households is similar to that of the low-income households away-from-home.

The number 156g is smaller than the 245g finding in Dubois et al. (2020). They only look at on-the-go while we combine both on-the-go and restaurants into one segment called away-from-home. The difference is mainly due to the diverse consumption behavior between U.K. and U.S. households as well as the difference in data composition and definition. First, in their U.K. data, the amount of on-the-go purchases (including vending machines, convenience stores, kiosks etc.) of soft drinks is three times as large as that of restaurant purchases. In contrast, our U.S. sample exhibits the opposite pattern. In other words, U.S. households purchase much more soft drinks in restaurants than on-the-go and our away-from-home segment is mainly composed of restaurant purchases. Consumers are normally found less elastic in restaurants compared to on-the-go. That implies a smaller elasticity estimates obtained with our dataset. Second, the two datasets differ in the composition of on-the-go purchases. For example, their on-the-go sample contains a significant amount of purchases made at "larger grocery stores when consumed immediately", while ours only contains a few of such observations. Purchases at grocery stores, although consumed immediately, tend to display a similar price sensitivity pattern as at-home purchases, and the sensitivity is larger than that of the away-from-home segment. As a result, their estimated aggregate-level price elasticities are again be larger than ours.

Given the small impact of the tax on sugary soft drinks on the away-from-home segment, all variations in the reduction in total sugar (at-home and away-from-home) across demographic groups is driven by the variation of that in the at-home segment.

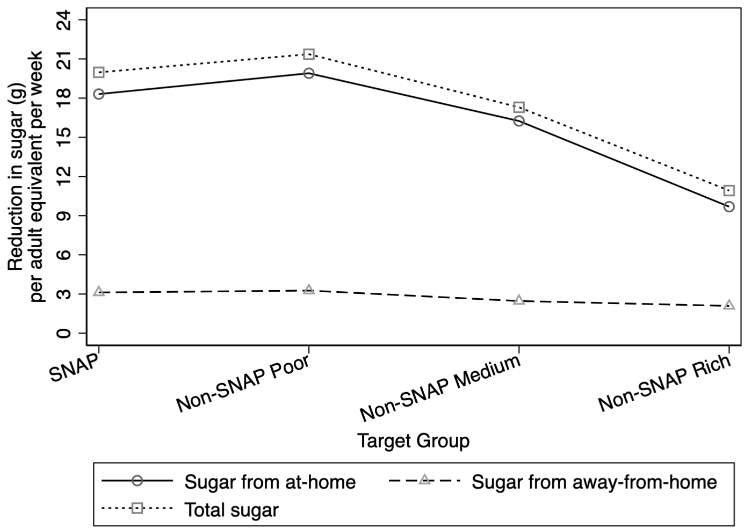
Low-income households are both more likely to be impacted by the policy and, conditional on this, have higher reductions in total SSB consumption than high-income households (panel (a)). The total reduction in sugar of SNAP and non-SNAP poor is around 21g per adult equivalent per week, which doubles the reduction in sugar of the high-income households. This finding is consistent with Dubois *et al.* (2020), who find that households with relatively low total equivalized expenditure (their proxy for income) have relatively large reductions in sugar from soft drinks given SSB taxes. Panel (b) shows that while the average percent reduction in total sugar is slightly lower for the SNAP households (17 percent versus 23 percent across non-SNAP poor and medium-income households), this group obtains a relatively large amount of sugar from products targeted by the tax. This means their level reductions is large.

Panel (c) and (d) show that households without children exhibit higher reduction in both the average level, as well as the average percent in total sugar from SSBs. This finding is also slightly different from Dubois *et al.* (2020), who find that the young consumers are more likely to be impacted by the policy and exhibit bigger level responses than older groups. However, notice that they look at individuals on-the-go while we look at households across all segments. Our finding implies that the effectiveness of soda taxes for the young might not be as large as being found in Dubois *et al.* (2020).

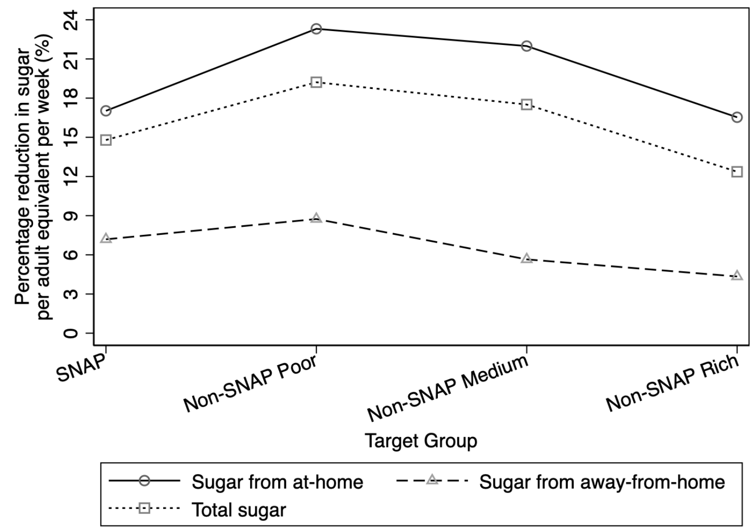
Panel (e) shows that the level reduction in total sugar is positively associated with household total sugar intake from SSBs. In particular, the difference in the reduction in total sugar between high and low total dietary sugar households is as large as 20g per adult equivalent per week. This finding is in sharp contrast to the conclusion made in Dubois *et al.* (2020) who argue that soda taxes are less successful targeting those with high total annual dietary sugar because their response to soda taxes is smaller on average in level terms. We instead find that their total, rather than on-the-go only, sugar reductions from SSBs is larger on average in level.

On the other hand, Dubois *et al.* (2020) and us both find that their response to soda taxes is smaller in percentage terms (panel (f)): for instance, the reduction for households with high decile of added sugar intake from SSB is over 14 percentage points below that for the low decile. They find that the reduction for the top decile of the dietary sugar distribution is over 4 percentage points below that for the bottom decile).

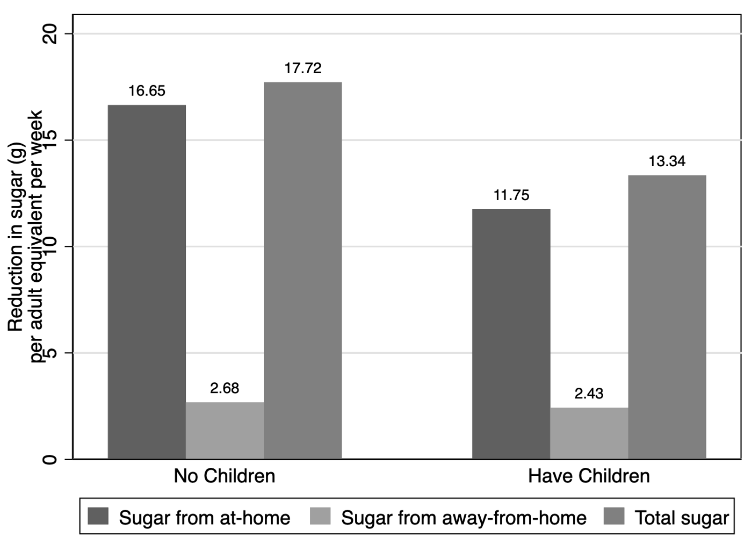
The difference in response across demographic groups can be supported by the pattern of preference variation. Even though the low-income households, those with children, and with high levels of sugar from SSB all have relatively strong SSB preferences, unlike the other groups those with children are less sensitive to prices.



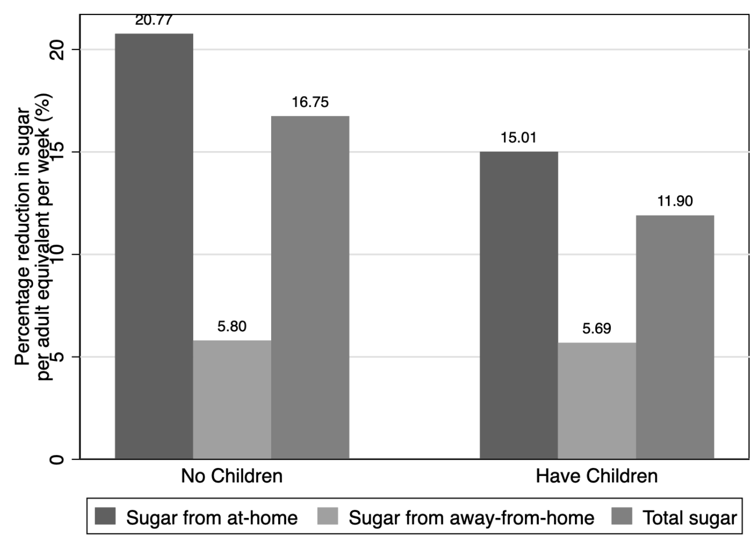
(a) By SNAP status and income level



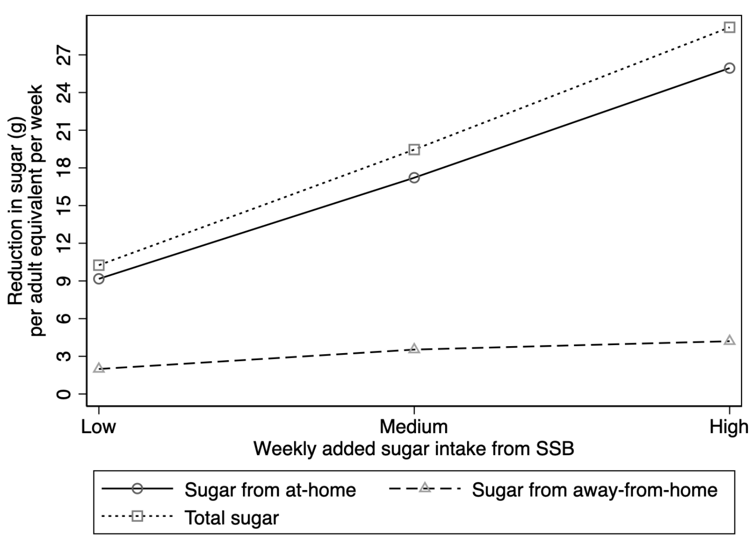
(b) By SNAP status and income level



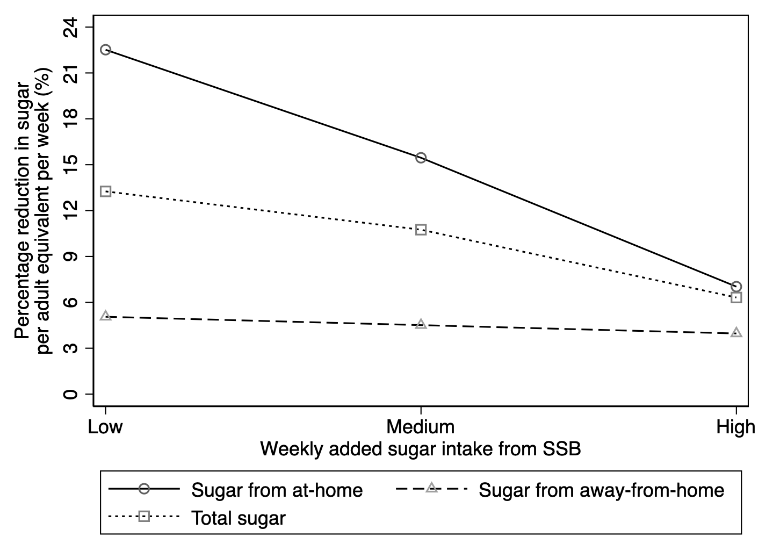
(c) By households with or without children



(d) By households with or without children



(e) By weekly added sugar intake



(f) By weekly added sugar intake

Figure 6: Reductions in Sugar From Drinks

Notes: The at-home segment is calculated based on 4,412 households in the sample, the away-from-home segment is based on 3,977 households, and the total is based on 4,683 households. Figures show how average reductions in SSB consumption varies across SNAP status and income groups, households with or without children and level of sugary diet. Figures (a), (c) and (e) show declines in level; figures (b), (d) and (f) show declines in percentage. In all figures, the pass-through rates is 100 percent in the at-home and away-from-home segments.

4.2 Consumer Welfare and Redistribution

The final question that we ask in this paper is: given the impact of SSB tax on household SSB demand, how will the tax affect consumer welfare? In particular, the tax will create an economic burden on consumers since it raises the price consumers pay. Moreover, with a higher price consumers can obtain less SSBs under the same total expenditure compared to under no tax regime. It is likely that for consumers who purchase SSBs, they will incur a welfare loss through this channel.

To answer this question, we use the standard Small & Rosen (1981) formula to calculate the compensating variation: the additional amount of money an individual would need to reach their initial utility following a change in prices. The compensating variation for a consumer i on choice occasion τ is given by

$$CV_{i\tau} = \frac{1}{\alpha_i}\left[\ln\left(\sum_{l=0}^{1}\left(\sum_{k\in C_l} e^{V^{post}_{ik\tau}(\mu_i)/\lambda}\right)^{\lambda}\right) - \ln\left(\sum_{l=0}^{1}\left(\sum_{k\in C_l} e^{V^{pre}_{ik\tau}(\mu_i)/\lambda}\right)^{\lambda}\right)\right], \quad (8)$$

where $V_{ik\tau}$ is defined in Section 3.1. We then integrated $CV_{i\tau}$ over the distribution of random coefficients and choice occasions to obtain the weekly (per adult equivalent) average compensating variation.

In Figure 7 panels (a)-(c) we describe how average compensating variation varies across SNAP status and income groups, households with or without children, and weekly added sugar intake. Compensating variation is determined by how exposed is the consumer to the tax (that is, whether the consumer buy a lot of the taxed goods) and how willing the consumer is to substitute towards other goods (the behavioral effect in Dubois *et al.* (2020)). In the at-home segment, low-income households and those with high weekly added sugar from SSB obtain more sugar from soft drinks and therefore are more exposed to the tax. Even without accounting for any behavioral effect, they would have higher compensating variation. After accounting for the behavioral effect, Figure 7 shows that the compensating variation remains high for low-income households and those with high weekly added sugar from SSB.

For the away-from-home segment, the picture is slightly reversed if we look at panel (a). The compensating variation is higher for higher-income households even though the difference is not as large as that in the at-home segment. Notice that even though Figure 6 panel (a) shows that the reductions in sugar from SSBs is similar across income groups in the away-from-home segment, the price of SSB is higher in this segment. In other words, higher-income households purchase more SSBs and more expensive SSBs away-from-home than poor households. That is why their compensating variation in the away-from-home segment is slightly higher than that of the poor households. Overall, the total compensating variation (accounting for both segments) is higher for low-income especially SNAP households. Panel (c) shows that high sugar diet households would have higher compensating variation in both the at-home and way-from-home segments. The difference in compensating variation between high and low sugar diet households is much larger in the at-home than the way-from-home segment. This finding is also consistent with previous result in Figure 6 panel (e) that the reduction in sugar is the highest for household with higher weekly added sugar intake from SSB, especially in the at-home segment.

The other thing to notice is that although the level of sugar reduction in Figure 6 is relatively low for the away-from-home segment, the associated compensating variation is somewhat comparable to that of the at-home segment. This is also due to the higher price of the away-from-home sugary drinks. In other words, for a similar level of economic burden, the tax in the away-from-home segment reduces sugar intake by a much smaller amount. For example, Figure 7 panel (c) shows that, in order to retrieve the utility before the tax, a medium sugar diet consumer would require a compensating variation of \$0.26 and \$0.23 for a 6 ounces and 1 ounce reduction in sugar from SSB for the at-home and away-from-home segment respectively. Furthermore, given that the SSB tax is based on volume, the benefits from tax revenue would be similar in the two segments. These imply that the welfare cost will be larger in the away-from-home segment than the at-home segment for a similar level of sugar reduction from SSB.

Even though the findings here suggest that the compensating variation is the highest among low-income households and those with high total dietary sugar, it does not imply that the total negative effect of the tax is the largest for these groups or the tax harms these groups the most. This is because consumers might purchase too much sugary soft drinks without considering the associated future costs (internality). Compensating variation only reflects part of the total consumer welfare effect of a sugary tax.

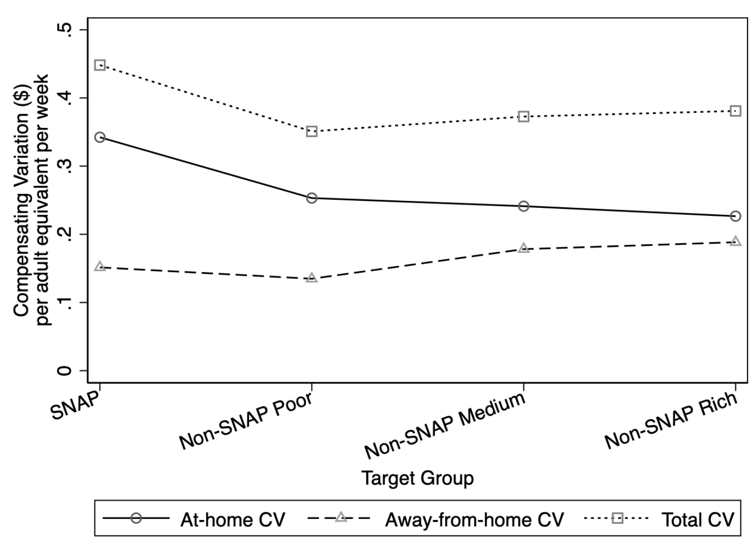
Policymakers are particularly concerned with low-income households. To provide an intuitive sense of the quantitative measure here, we find that in response to the tax consumers who have low income (regardless of SNAP participation), on average, reduce at-home sugar consumption from SSB by 21g per adult equivalent per week (annual: $21 \times 52 = 1092g$) and have average compensating variation of \$0.33 (annual: $\$0.33 \times 52 = \$17.16$). They reduce away-from-home sugar consumption from SSB by 3g per adult equivalent per week (annual: $3 \times 52 = 156g$) and have average compensating variation of \$0.15 (annual: $\$0.15 \times 52 = \$7.8$). The total reduction in sugar consumption from SSB is 24g per adult equivalent per week (annual: $24 \times 52 = 1248g$) and have average total compensating variation of \$0.43 (annual: $\$0.43 \times 52 = \$22.36$).

Following Dubois *et al.* (2020), we use a typical sugary soft drink (a can of Coca-Cola), as our standard unit of comparison; a can of Coca-Cola in the United States is 12 oz (355 ml) and contains 35g of sugar. If we assume that consumers receive no benefits from the tax revenue raised, then this implies that the internality from a can of Coca-Cola would need to be at least \$0.55 (that is $17.16*(35/1092)$) at-home and \$0.22 (that is $7.8*(35/1248)$) away-from-home for this group on average to benefit from the tax. The value \$0.55 at-home is over 5 times larger than the average internality from sugar sweetened soft drinks estimated in Allcott *et al.* (2019a).[33]

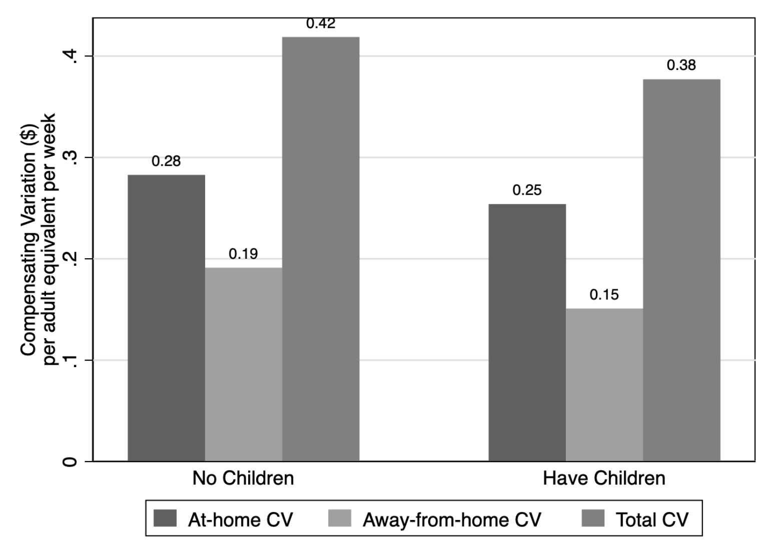
However, we have not accounted for any benefits from the tax revenue raised or any savings from the averted externalities (for example, the health care costs). We could further compare

[33] The value found in Dubois *et al.* (2020) for the on-the-go segment only is over 7 times larger than the average internality from sugar sweetened soft drinks estimated in Allcott *et al.* (2019a).

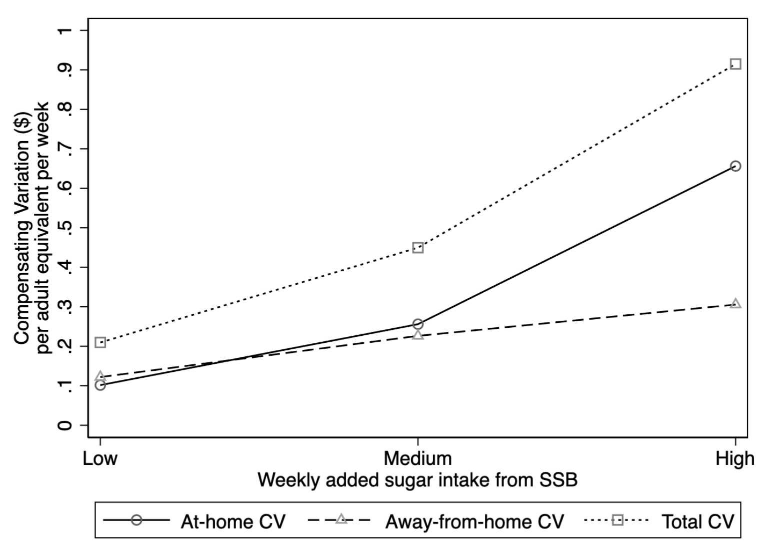
the CV to the calculated tax revenue per consumer for each group. For example, on average, the current tax raises $15.79 per adult equivalent over a year at-home.[34] The annual CV of SNAP consumers, $17.68, is only slightly higher than the average tax revenue $15.79. If the tax revenue can be distributed lump-sum back to the consumers, there only need to be a small positive amount of internality from the reduced soda consumption so that the SNAP participants can benefit from the tax. A similar calculation shows that the average tax revenue per consumer is $7.32 away-from-home, which is very closed to the CV of SNAP consumers ($7.8), implying a higher probability of the tax to be beneficial to them.



(a) By SNAP status and income level



(b) By households with or without children



(c) By weekly added sugar intake

Figure 7: Revealed Consumer Welfare Effect

Notes: The at-home segment is calculated based on 4,412 households in the sample, the away-from-home segment is based on 3,977 households, and the total is based on 4,683 households. Figures show how average compensating variation varies across SNAP status and income groups, households with or without children and level of sugary diet. In all figures, the pass-through rate is 100 percent in the at-home and away-from-home segments.

Another concern about excise taxes is that they are regressive: lower-income households con-

[34] Average tax revenue is calculated based on the post-tax volume consumption of SSB predicted by the model estimates and then multiplied by a tax rate of $0.01 per ounce. The calculation method is similar to Dubois *et al.* (2020), who find the tax revenue to be £3.14 per UK consumer for the on-the-go segment only.

sume more of the taxed goods and hence bare more of the tax burden compared to higher-income households. Figure 2 confirms that, in the case of sugary soft drinks, low-income households are more likely to be soft drink purchasers and to get more sugar from these products; those in the bottom half of the distribution and who purchase soft drinks on average obtain 25 percent more sugar from these products. Notice that previous literature that only look at at-home or on-the-go segment along might overestimate this regressivity concern because low-income households have higher SSB shares at-home compared to high-income households. However, Figure 2 shows that high-income households obtain more SSB shares in restaurants and this finding mitigates the regressivity concern.

In sum, we find that the total CV is not largely different across income and SNAP participation group. Only SNAP households have 10 percent higher CV than other groups. In other words, we do not find the soda taxes to be regressive when we account for both at-home and away-from-home segments. Why is it the case that, low-income households obtain more total sugar (at-home plus away-from-home) from SSB, but their CV is not much larger than higher-income households? The reason is that higher-income households obtain more soft drinks and more expensive drinks away-from-home. Even though low-income households obtain more sugar from SSB especially at-home, but the tax is based on volume rather than the amount of sugar contained in each drink. In other words, low-income and high-income households might buy similar amount (ounces) of soft drinks across segments but the sugar amount in each drink is much higher for low-income households.

This finding is different from Dubois *et al.* (2020) who suggest that compensating variation for a tax on sugary soft drinks is around 19 percent higher, on average, for soft drink purchasers in the bottom half of total equivalized grocery expenditure distribution than for those in the top half). Again, their policy implication is only based on the on-the-go segment, which accounts for only 10 percent of the total SSB expenditure on average of a U.S. household.

5 Conclusion

Beyond the focus on alcohol, tobacco, and gambling, sin taxes recently have been focused on the promotion of healthy eating. That is, the government has extended taxes to food and drinks. There is one main question related to assessing an effective an SSB tax. Who are most affected by a soda tax, and who bears the most of a soda taxes burden. This question is critical in assessing the effectiveness of an SSB tax in deterring excess levels of sugar intake and the welfare change of consumers due to a soda tax.

In this paper, we study the above questions by exploiting a novel dataset that cover household SSB demand from all channels (at-home, on-the-go, and in restaurants) for a representative sample of U.S. households. We utilize the rich demographic information on SNAP participation and eligibility and household income and composition, and nonparametrically estimate a flexible random coefficient nested logit model to document the heterogeneity in preferences and elasticity across household groups.

We find that preferences and elasticity vary with demographics in terms of SNAP participation, income, the existence of children, and the household total sugar from SSB. Such variation pattern is also different for the at-home and away-from-home segment. We find that soda taxes are less effective away-from-home and there is little variation in responses across households. In contrast, we find substantial variation in demand responses at-home across households. soda taxes are relatively effective at targeting the total sugar intake of the poor, those with high sugar consumption, and households without children for the at-home segment. Lastly, our results suggest that ignoring any segment will lead to biased policy implications on the targeting and effectiveness of soda tax.

We also find that, contrary to previous literature, the SSB tax is not highly regressive. The difference in compensating variation is much smaller than the difference in total sugar reduction across household income groups, especially when accounting for the away-from-home segment.

Firms will respond by adjusting their product types, pricing, advertising, the invention of new products, etc. Our results therefore only speak to the short to medium term effect of an SSB tax. Future research on incorporating the firm side responses into the picture will be very interesting and worth exploring.

References

Aguiar, Mark, & Hurst, Erik. 2007. Life-cycle prices and production. *American Economic Review*, **97**(5), 1533–1559. [5]

Allcott, Hunt, Lockwood, Benjamin B, & Taubinsky, Dmitry. 2019a. Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics*, **134**(3), 1557–1626. [2], [4], [5], [36]

Allcott, Hunt, Lockwood, Benjamin B, & Taubinsky, Dmitry. 2019b. Should we tax sugar-sweetened beverages? An overview of theory and evidence. *Journal of Economic Perspectives*, **33**(3), 202–27. [2], [12]

Andreyeva, Tatiana, Long, Michael W, & Brownell, Kelly D. 2010. The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food. *American journal of public health*, **100**(2), 216–222. [30]

Berardi, Nicoletta, Sevestre, Patrick, Tepaut, Marine, & Vigneron, Alexandre. 2016. The impact of a 'soda tax'on prices: evidence from French micro data. *Applied Economics*, **48**(41), 3976–3994. [31]

Bergman, U Michael, & Hansen, Niels Lynggård. 2019. Are Excise Taxes on Beverages Fully Passed through to Prices?: The Danish Evidence. *FinanzArchiv*, **75**(4), 323–356. [31]

Berkson, Joseph. 1950. Are there two regressions? *Journal of the american statistical association*, **45**(250), 164–180. [10]

Birchall, Cameron, & Verboven, Frank. 2022. Estimating Substitution Patterns and Demand Curvature in Discrete-Choice Models of Product Differentiation. [23]

Blundell, Wesley, Gowrisankaran, Gautam, & Langer, Ashley. 2020. Escalation of scrutiny: The gains from dynamic enforcement of environmental regulations. *American Economic Review*, **110**(8), 2558–2585. [3]

Bollinger, Bryan, & Sexton, Steven. 2018. Local excise taxes, sticky prices, and spillovers: evidence from Berkeley's soda tax. *Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax (January 12, 2018)*. [2], [4]

Bonnet, Céline, & Dubois, Pierre. 2010. Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance. *The RAND Journal of Economics*, **41**(1), 139–164. [23]

Bonnet, Céline, & Réquillart, Vincent. 2013. Tax incidence with strategic firms in the soft drink market. *Journal of Public Economics*, **106**, 77–88. [30]

Bonnet, Céline, Requillart, Vincent, *et al.* 2012. Sugar policy reform, tax policy and price transmission in the soft drink industry. *Toulouse School of Economics*. [5]

Browning, Martin, & Meghir, Costas. 1991. The effects of male and female labor supply on commodity demands. *Econometrica: Journal of the Econometric Society*, 925–951. [24]

Cawley, John, & Frisvold, David E. 2017. The pass-through of taxes on sugar-sweetened beverages to retail prices: the case of Berkeley, California. *Journal of Policy Analysis and Management*, **36**(2), 303–326. [31]

Cawley, John, Thow, Anne Marie, Wen, Katherine, & Frisvold, David. 2019. The economics of taxes on sugar-sweetened beverages: a review of the effects on prices, sales, cross-border shopping, and consumption. *Annual review of nutrition*, **39**, 317–338. [4]

Cawley, John, Frisvold, David, Jones, David, & Lensing, Chelsea. 2021. The pass-through of a tax on sugar-sweetened beverages in Boulder, Colorado. *American Journal of Agricultural Economics*, **103**(3), 987–1005. [31]

Chernozhukov, Victor, Hausman, Jerry A, & Newey, Whitney K. 2019. *Demand analysis with many prices*. Tech. rept. National Bureau of Economic Research. [5]

Currie, Janet. 2009. Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *JoUrnal of economIc lIteratUre*, **47**(1), 87–122. [2]

Currie, Janet, DellaVigna, Stefano, Moretti, Enrico, & Pathania, Vikram. 2010. The effect of fast food restaurants on obesity and weight gain. *American Economic Journal: Economic Policy*, **2**(3), 32–63. [2]

Cutler, David M, Glaeser, Edward L, & Shapiro, Jesse M. 2003. Why have Americans become more obese? *Journal of Economic perspectives*, **17**(3), 93–118. [2]

Diamond, Peter A, & Mirrlees, James A. 1971a. Optimal taxation and public production I: Production efficiency. *The American economic review*, **61**(1), 8–27. [5]

Diamond, Peter A, & Mirrlees, James A. 1971b. Optimal taxation and public production II: Tax rules. *The American Economic Review*, **61**(3), 261–278. [5]

Dubé, Jean-Pierre. 2005. Product differentiation and mergers in the carbonated soft drink industry. *Journal of Economics & Management Strategy*, **14**(4), 879–904. [9], [30]

Dubois, Pierre, Griffith, Rachel, & Nevo, Aviv. 2014. Do prices and attributes explain international differences in food purchases? *American Economic Review*, **104**(3), 832–867. [5]

Dubois, Pierre, Griffith, Rachel, & O'Connell, Martin. 2020. How well targeted are soda taxes? *American Economic Review*, **110**(11), 3661–3704. [2], [3], [4], [5], [8], [9], [12], [21], [24], [27], [29], [32], [33], [35], [36], [37], [38] Eizenberg, Alon, & Salvo, Alberto. 2015. The rise of fringe competitors in the wake of an emerging middle class: An empirical analysis. *American Economic Journal: Applied Economics*, **7**(3), 85–122. [23]

Falbe, Jennifer, Rojas, Nadia, Grummon, Anna H, & Madsen, Kristine A. 2015. Higher retail prices of sugar-sweetened beverages 3 months after implementation of an excise tax in Berkeley, California. *American journal of public health*, **105**(11), 2194–2201. [31]

Fox, Jeremy T, Kim, Kyoo Il, Ryan, Stephen P, & Bajari, Patrick. 2011. A simple estimator for the distribution of random coefficients. *Quantitative Economics*, **2**(3), 381–418. [3], [21], [23], [24]

Fox, Jeremy T, il Kim, Kyoo, & Yang, Chenyu. 2016. A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *Journal of Econometrics*, **195**(2), 236–254. [3], [21], [23], [24]

Gortmaker, S, Long, M, & Wang, YC. 2009. *The negative impact of sugar-sweetened beverages on children's health: A research synthesis, Robert Wood Johnson Foundation. Retrieved August 10, 2010.* [2]

Griffith, Rachel, O'Connell, Martin, Smith, Kate, & Stroud, Rebekah. 2020. What's on the menu? Policies to reduce young people's sugar consumption. *Fiscal studies*, **41**(1), 165–197. [2]

Griffith, Rachel, Jin, Wenchao, & Lechene, Valérie. 2022. The decline of home-cooked food. *Fiscal Studies*, **43**(2), 105–120. [5]

Grigolon, Laura, & Verboven, Frank. 2014. Nested logit or random coefficients logit? A comparison of alternative discrete choice models of product differentiation. *Review of Economics and Statistics*, **96**(5), 916–935. [21], [27]

Grogger, Jeffrey. 2017. *Soda taxes and the prices of sodas and other drinks: evidence from Mexico.* [31]

Gundersen, Craig. 2021. A proposal to reconstruct the Supplemental Nutrition Assistance Program (SNAP) into a universal basic income program for food. *Food Policy*, **101**, 102096. [13]

Harding, Matthew, & Lovenheim, Michael. 2017. The effect of prices on nutrition: comparing the impact of product-and nutrient-specific taxes. *Journal of Health Economics*, **53**, 53–71. [5]

Kit, Brian K, Fakhouri, Tala HI, Park, Sohyun, Nielsen, Samara Joy, & Ogden, Cynthia L. 2013. Trends in sugar-sweetened beverage consumption among youth and adults in the United States: 1999–2010. *The American journal of clinical nutrition*, **98**(1), 180–188. [3]

Lin, Xirong. 2023. Food Demand and Cash Transfers: A Collective Household Approach with Homescan Data. *Journal of Economic Behavior and Organization*, **212**, 233–259. [5]

Lopez, Rigoberto A, & Fantuzzi, Kristen L. 2012. Demand for carbonated soft drinks: implications for obesity policy. *Applied Economics*, **44**(22), 2859–2865. [29]

Marinello, Samantha, Pipito, Andrea A, Leider, Julien, Pugach, Oksana, & Powell, Lisa M. 2021. Longer-term impacts of sugar-sweetened beverage taxes on fast-food beverage prices: evidence from Oakland, California, 2-year post-tax. *Public Health Nutrition*, **24**(11), 3571–3575. [31]

Marshall, Guillermo. 2015. Hassle costs and price discrimination: An empirical welfare analysis. *American Economic Journal: Applied Economics*, **7**(3), 123–146. [9], [22], [27]

Martin, Crescent B, Wambogo, Edwina A, Ahluwalia, Namanjeet, & Ogden, Cynthia L. 2020. Nonalcoholic beverage consumption among adults: United States, 2015–2018. [17]

McFadden, Daniel. 1978. MODELING THE CHOICE OF RESIDENTIAL LOCATION. *Transportation Research Record*. [23]

Miravete, Eugenio J, Seim, Katja, & Thurk, Jeff. 2020. One markup to rule them all: Taxation by liquor pricing regulation. *American Economic Journal: Microeconomics*, **12**(1), 1–41. [5]

Miravete, Eugenio J, Seim, Katja, & Thurk, Jeff. 2022. *Robust pass-through estimation in discrete choice models*. Tech. rept. Working paper, Department of Economics, University of Pennsylvania. [23]

Moran, Alyssa J, Subramanian, SV, Rimm, Eric B, & Bleich, Sara N. 2019. Characteristics Associated with Household Purchases of Sugar-Sweetened Beverages in US Restaurants. *Obesity*, **27**(2), 339–348. [2], [3], [5]

Nevo, Aviv, Turner, John L, & Williams, Jonathan W. 2016. Usage-based pricing and demand for residential broadband. *Econometrica*, **84**(2), 411–443. [3]

O'Connell, Martin, & Smith, Kate. 2020. Corrective tax design and market power. [5], [8], [24]

Okrent, Abigail, & Alston, Julian. 2012. The demand for disaggregated food-away-from-home and food-at-home products in the United States. *USDA-ERS Economic Research Report*. [5]

Ramsey, Frank P. 1927. A Contribution to the Theory of Taxation. *The economic journal*, **37**(145), 47–61. [5]

Rojas, Christian, & Wang, Emily Yucai. 2017. Do taxes for soda and sugary drinks work? Scanner data evidence from Berkeley and Washington. *Scanner Data Evidence from Berkeley and Washington (September 23, 2017)*. [4]

Saksena, Michelle J, Okrent, Abigail M, Anekwe, Tobenna D, Cho, Clare, Dicken, Christopher, Effland, Anne, Elitzak, Howard, Guthrie, Joanne, Hamrick, Karen S, Hyman, Jeffrey, *et al.* 2018. *America's eating habits: food away from home*. Tech. rept. United States Department of Agriculture, Economic Research Service. [5]

Schennach, Susanne M. 2013. Regressions with Berkson errors in covariates—a nonparametric approach. *The Annals of Statistics*, 1642–1668. [10]

Seiler, Stephan, Tuchman, Anna, & Yao, Song. 2021. The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects. *Journal of Marketing Research*, **58**(1), 22–49. [4]

Small, Kenneth A, & Rosen, Harvey S. 1981. Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, 105–130. [35]

Wang, Emily Yucai. 2015. The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity. *The RAND Journal of Economics*, **46**(2), 409–441. [5]

Appendices

1 FoodAPS Data Collection Process

A screening interview determined whether the household at a sampled residence was eligible to participate in FoodAPS based on household income and SNAP participation. If eligible, the FoodAPS screener identified the main food shopper or meal planner in the household and invited him or her to participate in the week-long data collection.

The PR was asked to complete two in-person interviews and to call the study's telephone center for three brief telephone interviews regarding food acquisition events over the course of one week. The PR food book included both Blue pages to report details for "food at home" and Red pages to report "food away from home" acquisitions. The PR was responsible for recording food acquisitions by members under 11 years old.

Households were asked to scan barcodes on foods, save their receipts from stores and restaurants, and write information in their food books. For food-at-home acquisitions, the scanned barcodes were intended to be the primary source of item-level descriptions, while the receipts were intended to provide the price or expenditure information for each item. The Food Book (Blue) pages would provide the rest of the information and saved receipts would be used to verify this information and/or fill in missing information from the Blue page. For food-away-from-home acquisitions, the phone calls were intended to be the main source of item descriptions, details about the event, and price/expenditure information. The Red pages were reviewed to identify and capture any information that had not been reported during a phone call.

2 Details of Data Construction

Product

We group the items by three characteristics dimensions: 10 different beverage types (soft drinks, fruit drinks, sport and energy drinks, sweetened coffee and tea, diet drinks, fruit and vegetable juice, unsweetened coffee and tea, flavored milk, flavored water and water), three different packaging formats (regular, large, and multi-pack), and two different segments (at-home and away-from-home). A product is defined as either a type-packaging-segment or a type-segment combination, with each product referring to a group of items sharing the same characteristics.

Price

Price varies across products, time period and retailer type. The price information of near 40% of beverage products is missing in the away-from-home dataset. Most of them are sold in a meal bundle. We use the price of other items of the same product within the same place category to impute missing prices. Specifically we apply a linear interpolation of transaction price on package size for missing values of price, and perform this calculation separately for each product-place category combination.

Package Size

For a given UPC, the multi-pack information is not provided in the FoodAPS data. That is, we do not observe how many of those goods appear in a given pack (e.g. a 6 pack of soda). The *pkgsize* variable is the multiplication of the size of individual packaging and the number of individual packaging. Thus it measures the size of an item defined by a given UPC. There is another variable *quantity* in the item level data, which indicates how many of that item is purchased.

First, we restrict the sample to households who purchased a single item. Sometimes the *pkgsize* and *quantity* information is inconsistent with the UPC information. For example, an item is a 20 oz bottle-8 counts according to the UPC, so the correct item-level size should be $20 * 8$ oz. However, in the FoodAPS data, some of them are incorrectly recorded as $pkgsize = 160$ and $quantity = 8$, yielding a size of $160 * 8$ oz. It is hard to identify items with wrong size information like this. Therefore we eliminate all transactions with $quantity > 1$. The fraction of households buying multiple units of a given item is small (around 5%). By doing so, we abstract away multi-units purchasing behavior. Note that we did not eliminate multi-pack items. For example, we allow for the households to buy a 6 pack of coke, but we don't allow for 6 bottles of single-bottle coke.

Second, we look up the UPC code to recover the multi-pack information using the *pkgsize* variable. Almost always, a given *pkgsize* corresponds to a unique combination of number of goods in a pack and the size of the individual packaging good. Thus we are able to identify whether the item is a multi-pack product and the size of each single bottle using the *pkgsize* variable.

Third, near 15% of SSB transactions have missing package size information. Only less than 10% of the missing package size information might be imputed using UPC code. There are two things we need to impute: (a) package size, i.e., size of the item in ounces; (b) package type, i.e., regular single-bottle, large single-bottle, or multi-pack. We impute the missing package size by linear interpolation of package size on transaction price. This calculation is performed separately for each drink category within a given month. Products with imputed weight less than or equal to 32 ounces are classified as regular-size single-bottle. Products with imputed weight less than or equal to 32 ounces are either large single-bottle or multi-pack. In order to impute the missing information of package type, we take the subsample of observations that are either large single-bottle or multi-pack with non-missing information on the package type and fit a random forest classification model of package type (either large-bottled or multi-pack) on package size in ounce and transaction price. Based on the estimated model, we predict the package type for the set of observations with missing values.

3 Separability Test

We consider two sets of separability tests. First, we include the average price of away-from-home soda in the at-home demand equations and vice versa. Second, we include a dummy for whether there were any away-from-home SSB purchases, and vice versa. More specifically, we estimate

$$q_i^{\text{AH}} = x_i'\beta + \gamma p_i^{\text{AFH}} + \varepsilon_i,$$

and

$$q_i^{\text{AH}} = x_i'\beta + \alpha d_i^{\text{AFH}} + \varepsilon_i,$$

and similar for the away-from-home demand equations. q_i^{AH} is the volume consumption of SSB at-home for household i, p_i^{AFH} is the average prices of away-from-home soda, and d_i^{AFH} is a dummy variable that equals one if there were any away-from-home purchases of SSB. Endogeneity might arise because households that demand more soda at-home might have unobserved characteristics that also cause them to purchase soda away from home, or visit places that offer low (or high) price soda. Due to the short time period of the data, it is infeasible to include a full set of household fixed effects to control for the unobservables. Instead, We include a rich set of household characteristics, x_i, to deal with potential endogeneity. x_i include the constant term, household size, whether the household has children, income, age of primary respondent, average BMI, diet status, and knowledge of nutrition.

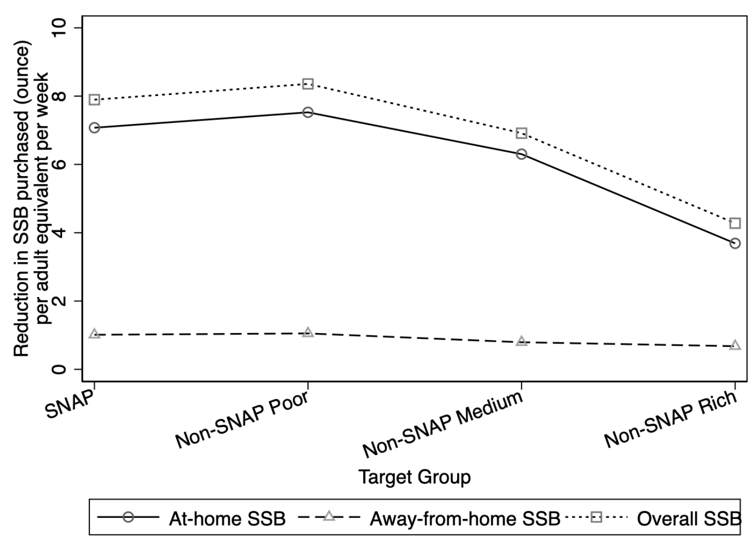
Table A1: Separability Tests

	AH Demand	AFH Demand	AH Demand	AFH Demand
At-home price	-98.90***	0.892	-87.79***	
	(8.410)	(2.295)	(6.581)	
Away-from-home price	2.135	-12.52***		-13.05***
	(5.911)	(1.613)		(1.188)
At-home soda purchased				2.884
				(3.823)
Away-from-home soda purchased			19.99	
			(16.44)	
Child	59.60*	11.58	51.43**	8.334
	(23.46)	(6.401)	(19.55)	(4.788)
Household size	26.37***	8.649***	31.51***	9.769***
	(7.053)	(1.925)	(5.877)	(1.507)
Age	-0.535	0.0298	-0.371	-0.0623
	(0.633)	(0.173)	(0.499)	(0.123)
BMI	5.118***	0.535	4.540***	0.616*
	(1.296)	(0.354)	(1.069)	(0.272)
Constant	240.0**	51.26*	161.3**	51.43***
	(74.91)	(20.45)	(57.40)	(14.85)
Income group	Yes	Yes	Yes	Yes
Diet status	Yes	Yes	Yes	Yes
Nutrition fact	Yes	Yes	Yes	Yes

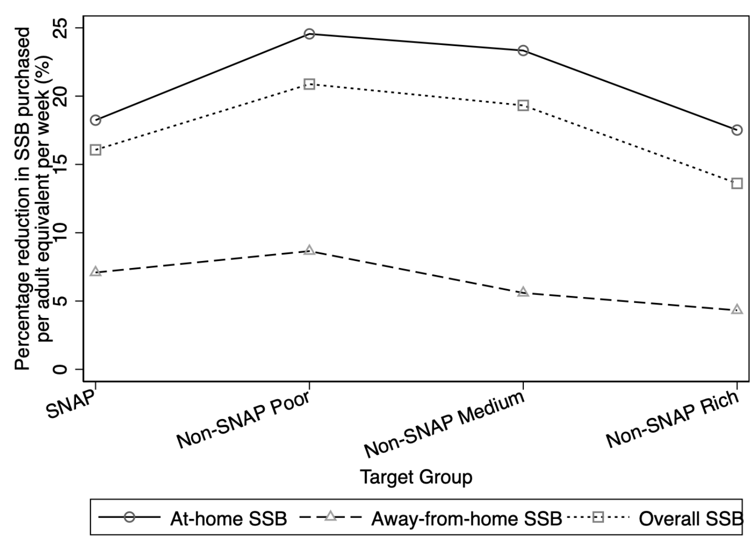
Notes: Standard errors in parentheses. Estimates of categorical variables are omitted from the table. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In Table A1, we report our results. For example, column 1 shows that the estimated value of γ is 2.135, with a large standard error 5.911, suggesting that the price of away-from-home soda has insignificant effect on the demand at home. Similarly, column 3 shows that whether there were any away-from-home soda purchases has no significant impact on the amount of soda consumed at home ($\hat{\alpha} = 19.99$ and standard error is 16.44). The results supporting separability here are consistent with other results in the literature.

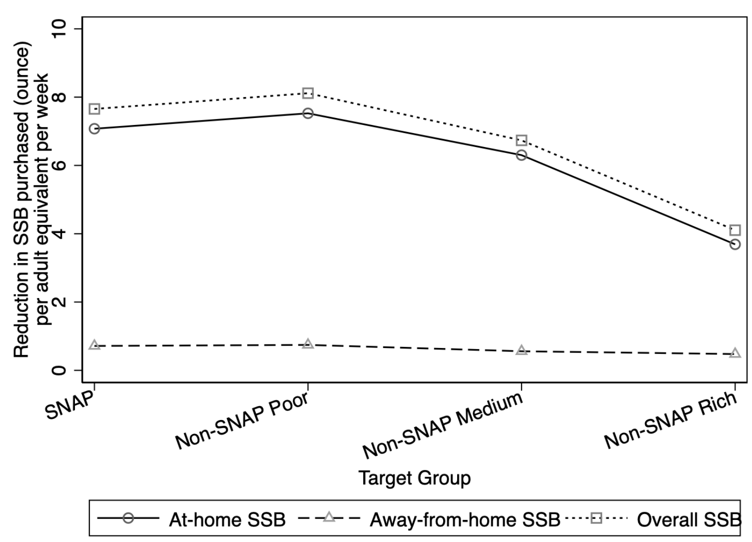
4 Additional Counterfactual Results



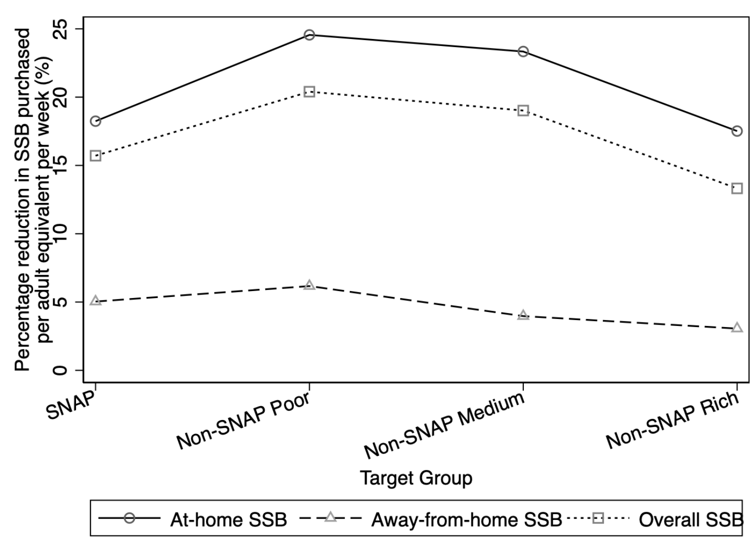
(a) Level reductions, at-home 100% pass-through, away-from-home 100% pass-through



(b) Percentage reductions, at-home 100% pass-through, away-from-home 100% pass-through



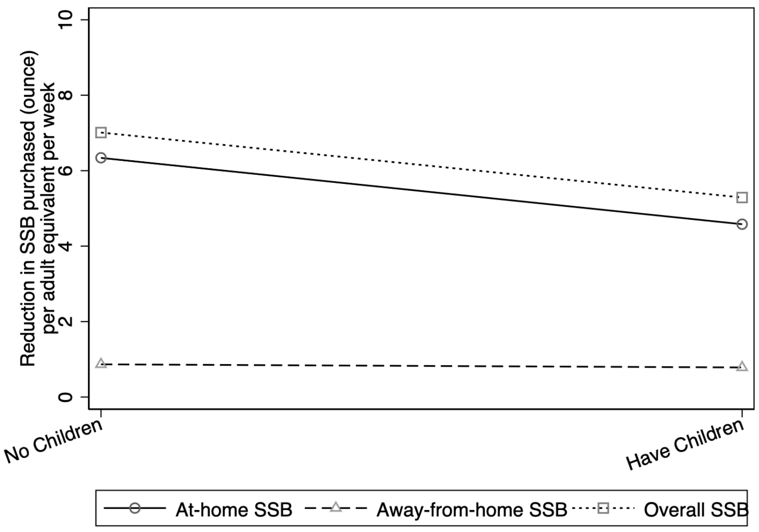
(c) Level reductions, at-home 100% pass-through, away-from-home 70% pass-through



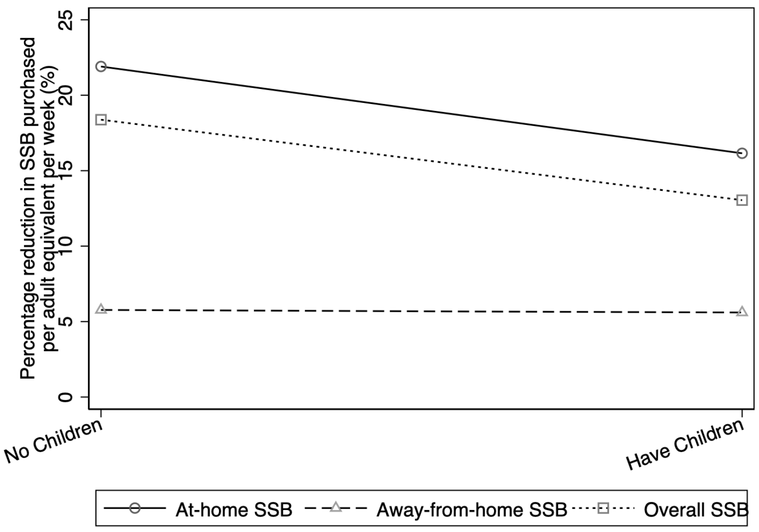
(d) Percentage reductions, at-home 100% pass-through, away-from-home 70% pass-through

Figure A1: Reductions in SSB Purchases by SNAP Status and Income Level

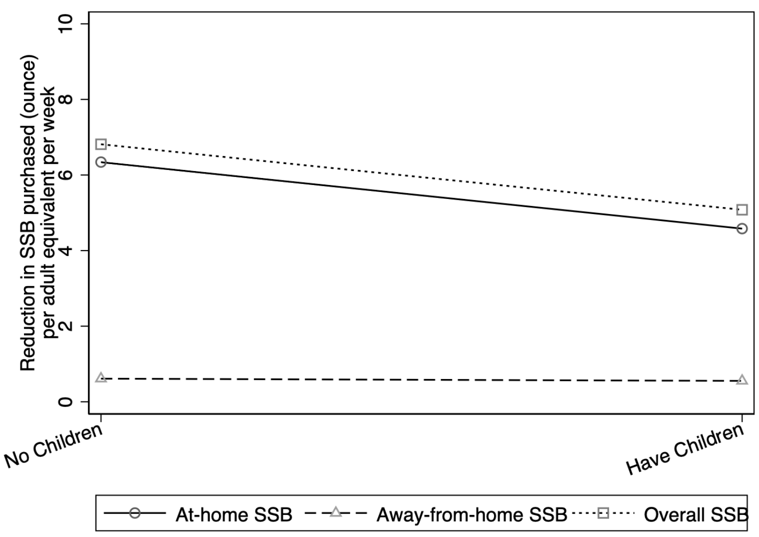
Notes: Figures show how average reductions in SSB consumption varies across SNAP status and income groups. In all figures, the pass-through rates is 100 percent in the at-home segment. Figures (a) and (b) show the results with 100 percent pass-through rates in away-from-home, and (c) and (d) are 70 percent pass-through rates in away-from-home.



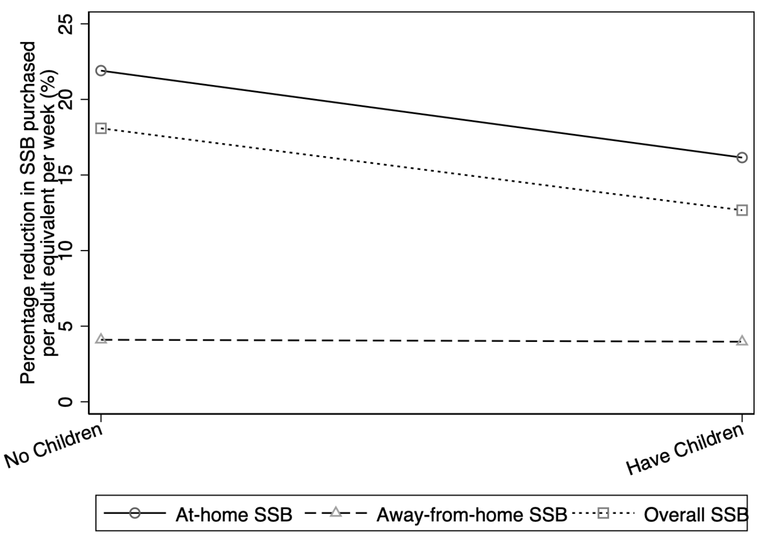
(a) Level reductions, at-home 100% pass-through, away-from-home 100% pass-through



(b) Percentage reductions, at-home 100% pass-through, away-from-home 100% pass-through



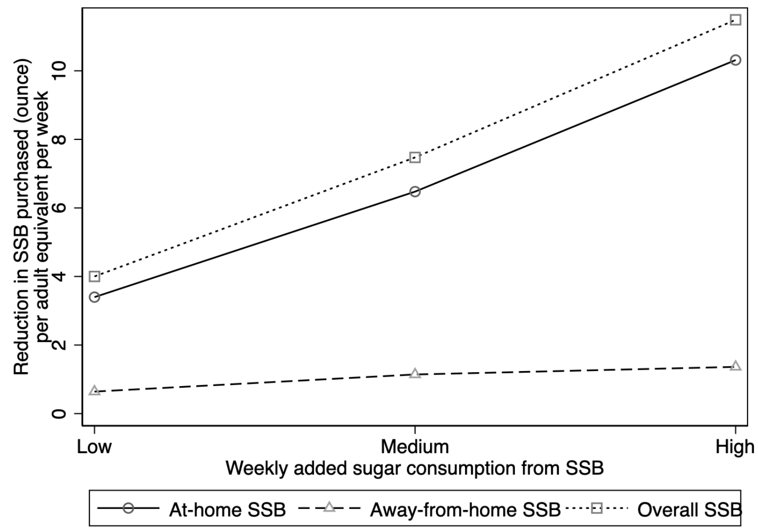
(c) Level reductions, at-home 100% pass-through, away-from-home 70% pass-through



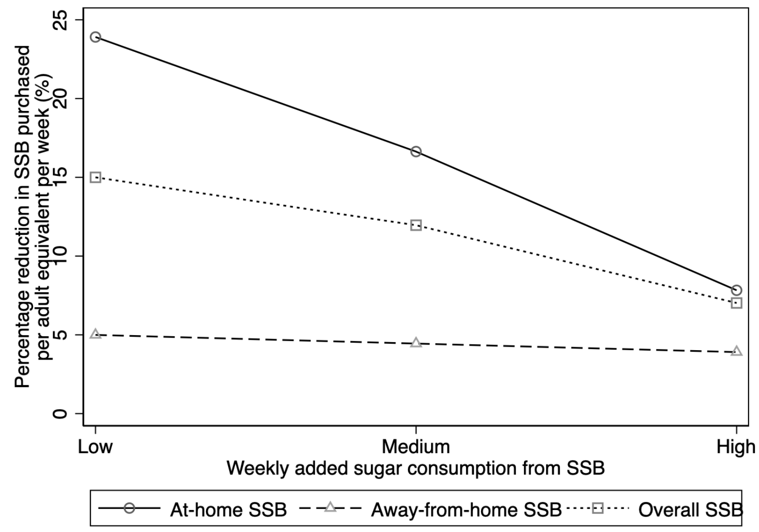
(d) Percentage reductions, at-home 100% pass-through, away-from-home 70% pass-through

Figure A2: Reductions in SSB Purchases by Households with or without Children

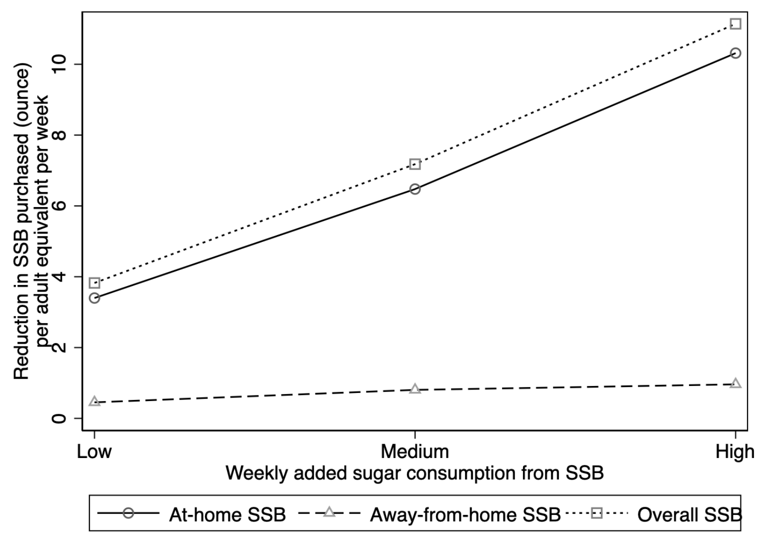
Notes: Figures show how average reductions in SSB consumption varies across households with or without children. In all figures, the pass-through rates is 100 percent in the at-home segment. Figures (a) and (b) show the results with 100 percent pass-through rates in away-from-home, and (c) and (d) are 70 percent pass-through rates in away-from-home.



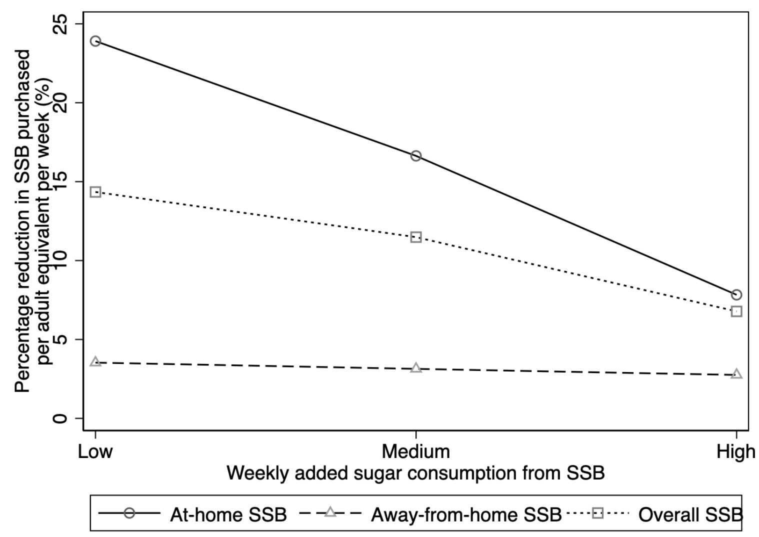
(a) Level reductions, at-home 100% pass-through, away-from-home 100% pass-through



(b) Percentage reductions, at-home 100% pass-through, away-from-home 100% pass-through



(c) Level reductions, at-home 100% pass-through, away-from-home 70% pass-through



(d) Percentage reductions, at-home 100% pass-through, away-from-home 70% pass-through

Figure A3: Reductions in SSB Purchases by Weekly Added Sugar Intake

Notes: Figures show how average reductions in SSB consumption varies across level of sugary diet. In all figures, the pass-through rates is 100 percent in the at-home segment. Figures (a) and (b) show the results with 100 percent pass-through rates in away-from-home, and (c) and (d) are 70 percent pass-through rates in away-from-home.

5 Additional Results of Price Elasticity

In this section we show product level elasticities. Tables A2 and A3 report the demand change for alternative drink options resulting from a 1% price increase of each product. Tables A4 and A5 provide the full matrix of own- and cross-price elasticities for all products. For example, Tables A2 shows that a 1% increases in price of the regular sized soft drink category would result in a reduction of demand of 0.25% while demand for non-SSB drinks would rise by 0.012%.

Table A2: At-Home Product Level Price Elasticities

	Own	FAH SSB	FAH non-SSB
Soft Drinks	-0.247	0.014	0.012
Large-Bottle Soft Drinks	-0.235	0.018	0.016
Pack Soft Drinks	-0.467	0.031	0.025
Fruit Drinks	-0.245	0.008	0.007
Large-Bottle Fruit Drinks	-0.341	0.016	0.013
Pack Fruit Drinks	-0.488	0.009	0.008
Sport and Energy Drinks	-0.239	0.005	0.005
Pack Sport and Energy Drinks	-0.875	0.011	0.008
Sweetened Coffee and Tea	-0.252	0.002	0.002
Large-Bottle Sweetened Coffee and Tea	-0.427	0.005	0.005
Pack Sweetened Coffee and Tea	-0.619	0.005	0.004
Diet Drinks	-0.368	0.003	0.005
Large-Bottle Diet Drinks	-0.390	0.004	0.007
Pack Diet Drinks	-0.854	0.007	0.010
Fruit and Vegetable Juice	-0.702	0.027	0.040
Unsweetened Coffee and Tea	-0.447	0.001	0.003
Flavored Milk	-0.545	0.004	0.006
Flavored and Enhanced Water	-0.377	0.005	0.008
Water	-0.559	0.022	0.032

Notes: For each of the products we compute the change in demand for that product, for other SSB alternatives and for non-SSB alternatives resulting from a 1% price increase. Numbers are averaged across time and place types.

Table A3: Away-From-Home Product Level Price Elasticities

	Own	FAFH SSB	FAFH non-SSB
FAFH Soft Drinks	-0.277	0.065	0.049
FAFH Large-Bottle Soft Drinks	-0.374	0.016	0.011
FAFH Fruit Drinks	-0.442	0.016	0.011
FAFH Large-Bottle Fruit Drinks	-0.557	0.004	0.003
FAFH Sport and Energy Drinks	-0.423	0.011	0.008
FAFH Large-Bottle Sport and Energy Drinks	-0.738	0.003	0.002
FAFH Sweetened Coffee and Tea	-0.410	0.048	0.034
FAFH Large-Bottle Sweetened Coffee and Tea	-0.382	0.005	0.004
FAFH Diet Drinks	-0.364	0.020	0.027
FAFH Large-Bottle Diet Drinks	-0.396	0.004	0.005
FAFH Fruit and Vegetable Juice	-0.443	0.006	0.008
FAFH Unsweetened Coffee and Tea	-0.348	0.033	0.044
FAFH Flavored Milk	-0.431	0.006	0.008
FAFH Flavored and Enhanced Water	-0.426	0.002	0.003
FAFH Water	-0.343	0.012	0.017

Notes: For each of the products we compute the change in demand for that product, for other SSB alternatives and for non-SSB alternatives resulting from a 1% price increase. Numbers are averaged across time and place types.

Table A4: At-Home Product Level Own- and Cross-Price Elasticities

	Soft	L Soft	P Soft	Fruit	L Fruit	P Fruit	Sport	P Sport	Sw Coffee	L Sw Coffee	P Sw Coffee	Diet	L Diet	P Diet	Juice	Unsw Coffee	Milk	Flavored Water	Water
Soft	-0.247	0.014	0.014	0.014	0.014	0.015	0.014	0.013	0.014	0.014	0.011	0.012	0.012	0.013	0.012	0.012	0.012	0.012	0.012
L Soft	0.018	-0.235	0.018	0.018	0.018	0.018	0.018	0.017	0.018	0.018	0.014	0.016	0.016	0.016	0.016	0.016	0.016	0.015	0.016
P Soft	0.031	0.031	-0.467	0.031	0.031	0.033	0.031	0.029	0.031	0.031	0.028	0.024	0.024	0.028	0.025	0.024	0.024	0.024	0.024
Fruit	0.008	0.008	0.008	-0.245	0.008	0.008	0.009	0.008	0.008	0.008	0.009	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
L Fruit	0.016	0.016	0.016	0.016	-0.341	0.015	0.016	0.015	0.016	0.016	0.017	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
P Fruit	0.010	0.009	0.009	0.009	0.009	-0.488	0.010	0.009	0.009	0.010	0.010	0.007	0.007	0.010	0.008	0.007	0.007	0.007	0.008
Sport	0.006	0.006	0.005	0.006	0.006	0.006	-0.239	0.005	0.006	0.006	0.004	0.005	0.005	0.006	0.005	0.005	0.005	0.005	0.005
P Sport	0.011	0.011	0.011	0.011	0.011	0.012	0.011	-0.875	0.011	0.011	0.012	0.007	0.007	0.009	0.008	0.008	0.008	0.007	0.008
Sw Coffee	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.002	-0.252	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
L Sw Coffee	0.005	0.006	0.005	0.006	0.006	0.006	0.006	0.005	0.006	-0.427	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.004	0.005
P Sw Coffee	0.005	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.005	-0.619	0.003	0.003	0.003	0.004	0.004	0.004	0.003	0.004
Diet	0.004	0.003	0.003	0.004	0.003	0.004	0.004	0.003	0.004	0.003	0.002	-0.368	0.006	0.006	0.005	0.006	0.006	0.006	0.006
L Diet	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.004	0.008	-0.390	0.008	0.007	0.008	0.008	0.008	0.008
P Diet	0.008	0.008	0.007	0.008	0.007	0.009	0.008	0.006	0.008	0.007	0.006	0.011	0.011	-0.854	0.010	0.011	0.010	0.011	0.010
Juice	0.028	0.028	0.026	0.028	0.027	0.027	0.029	0.022	0.028	0.027	0.028	0.042	0.042	0.039	-0.702	0.041	0.040	0.042	0.040
Unsw Coffee	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.001	0.002	0.002	0.001	0.003	0.003	0.002	0.002	-0.447	0.003	0.003	0.003
Milk	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.004	0.004	0.004	0.007	0.007	0.007	0.006	0.007	-0.545	0.007	0.007
Flavored Water	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.008	0.008	0.009	0.007	0.008	0.008	-0.377	0.008
Water	0.024	0.024	0.022	0.024	0.023	0.022	0.024	0.018	0.024	0.022	0.021	0.035	0.035	0.031	0.032	0.034	0.033	0.035	-0.559

Notes: For each of the at-home products we compute the own- and cross-price elasticities. Numbers are averaged across time and place types.

Table A5: Away-From-Home Product Level Own- and Cross-Price Elasticities

	Soft	L Soft	Fruit	L Fruit	Sport	L Sport	Sweetened Coffee	L Sweetened Coffee	Diet	L Diet	Juice	Unsweetened Coffee	Milk	Flavored Water	Water
Soft	-0.277	0.067	0.066	0.066	0.066	0.065	0.066	0.075	0.050	0.055	0.049	0.048	0.051	0.050	0.050
L Soft	0.016	-0.374	0.015	0.018	0.016	0.013	0.016	0.020	0.012	0.014	0.013	0.011	0.013	0.012	0.013
Fruit	0.016	0.016	-0.442	0.015	0.016	0.012	0.016	0.015	0.011	0.010	0.011	0.011	0.011	0.011	0.011
L Fruit	0.004	0.005	0.004	-0.557	0.004	0.002	0.004	0.004	0.003	0.003	0.003	0.003	0.004	0.003	0.003
Sport	0.011	0.012	0.011	0.012	-0.423	0.020	0.011	0.013	0.008	0.010	0.009	0.008	0.009	0.008	0.009
L Sport	0.003	0.003	0.003	0.004	0.003	-0.738	0.003	0.004	0.002	0.003	0.002	0.002	0.003	0.002	0.002
Sweetened Coffee	0.048	0.051	0.046	0.055	0.047	0.018	-0.410	0.024	0.038	0.018	0.036	0.034	0.040	0.037	0.036
L Sweetened Coffee	0.005	0.006	0.005	0.006	0.005	0.005	0.005	-0.382	0.004	0.005	0.004	0.004	0.004	0.004	0.004
Diet	0.020	0.014	0.020	0.016	0.020	0.019	0.020	0.016	-0.364	0.021	0.019	0.027	0.022	0.030	0.019
L Diet	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.005	-0.396	0.005	0.005	0.005	0.005	0.005
Juice	0.006	0.007	0.006	0.006	0.006	0.006	0.006	0.007	0.007	0.010	-0.443	0.009	0.008	0.007	0.010
Unsweetened Coffee	0.034	0.036	0.032	0.037	0.033	0.025	0.033	0.024	0.046	0.032	0.045	-0.348	0.048	0.044	0.047
Milk	0.006	0.007	0.006	0.008	0.006	0.004	0.006	0.005	0.009	0.007	0.009	0.008	-0.431	0.009	0.009
Flavored Water	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.003	0.002	0.002	0.003	0.002	-0.426	0.002
Water	0.013	0.014	0.012	0.010	0.012	0.005	0.012	0.014	0.011	0.019	0.019	0.017	0.013	0.011	-0.343

Notes: For each of the away-from-home products we compute the own- and cross-price elasticities. Numbers are averaged across time and place types.