Soft Drink Taxation and Habit Formation

Renke Schmacker *

May 17, 2020

Abstract

Despite its popularity among policymakers, the effectiveness of taxation in reducing soft drink consumption is still under dispute. To date, most studies using discrete choice demand models and scanner data do not consider the role of habit formation. This paper uses a structural demand model to analyze the impact of soft drink taxes in the presence of habit formation and stockpiling. The model is estimated using nested logit and incorporates unobserved heterogeneity in tastes. The estimated model is used to simulate short-run and long-run price elasticities, as well as the simulated impact of different soft drink taxes. The results show that long-run price elasticities are approximately 20 percent larger than short-run elasticities due to habit formation. Moreover, excise taxes on sugary soft drinks are more effective in reducing sugar consumption than *ad valorem* taxes and excise taxes that do not distinguish between sugary and diet beverages.

JEL-codes: H20, D12, I18
Keywords: soft drink tax, sin tax, state dependence, addiction

*DIW Berlin, `rschmacker[at]diw[dot]de`

I thank Peter Haan, Daniel Kemptner, Kevin Tran and Hannes Ullrich for helpful comments. I would like to thank IRI for making the data available. All estimates and analyses in this paper, based on data provided by IRI, are by the authors and not by IRI. Financial support by the Leibniz Competition through the project GlobalFood (SAW-2015-DIW-4) is gratefully acknowledged.

1 Introduction

In the fight against obesity, the World Health Organization (WHO, 2016) recommends the introduction of a tax on sugary soft drinks. Although these taxes have been studied extensively (see the surveys in Cawley *et al.*, 2019b; Allcott *et al.*, 2019b), it is still under dispute whether they are effective in curbing consumption. Hence, there is a sustained need for demand models that allow to analyze the effects of soft drink taxes before they are introduced.

In order to simulate the total effect of soft drink taxes on purchases, it is important to take into account consumption dynamics. On the one hand, since consumers can stockpile soft drinks when prices are low, static demand models tend to overestimate price elasticities (Wang, 2015). On the other hand, if consumption of soft drinks is habit-forming, static demand models underestimate long-run price elasticities. The reason is that the tax does not just discourage instantaneous consumption, but due to habit formation, it also reduces the utility of consumption in the next periods. In fact, biological research shows evidence for addiction to sugar and sweetness (Ahmed *et al.*, 2013; Mennella *et al.*, 2016) and market-level data shows a high degree of habit persistence in demand for soft drinks (Zhen *et al.*, 2011). However, until now the literature on soft drink taxation has largely neglected these two sources of state dependence and assumes consumption to be time-separable.

In this paper, I estimate a structural model of soft drink demand that accounts for both stockpiling and habit formation. I use weekly scanner data of household purchases in the US from 2003 to 2004 in order to analyze the effects of soft drink taxes in the presence of state dependence. The analysis proceeds in two steps. In the first step, I provide descriptive evidence for positive and negative state dependence. I follow the approach in Tuchman (2019) and find a very similar dynamic purchasing pattern compared to her analysis of the demand for cigarettes. After controlling for stockpiling, the proxy variables for habit formation have the expected sign and are highly significant. This highlights the importance of incorporating habit formation into the demand model. In the second step, I estimate a discrete choice model of soft drink demand. The model incorporates habit formation, as represented by the lagged purchase of the previous period, and stockpiling, as proxied by purchases made on sale. I estimate the model using nested logit to allow for correlated shocks to utility. Moreover, the panel data allows to control for unobserved heterogeneity in tastes, which is important in order to distinguish true from spurious state dependence (Heckman, 1981). The latter describes the situation when state dependence mainly picks up persistent taste differences.

The estimation results of the structural model provide evidence for habit formation in soft drink demand. As expected, the coefficients for state dependence decrease after controlling for persistent taste differences, but they remain sizeable. The effect of state dependence can be illustrated by the difference between short-run and long-run price elasticities. While the former only considers the price effect on instantaneous demand, the latter takes into account that a changed consumption pattern also affects future demand. I estimate a short-run price elasticity of 0.87 for sugary soft drinks and a long-run price elasticity of 1.07. Hence, the long-run price elasticity is approximately 20 percent larger than the short-run elasticity.

I use the estimated model to simulate the effect of different soft drink taxes on demand. The simulated taxes resemble actually implemented tax designs: (1) an excise tax on sugary beverages (as, e.g., in Mexico), (2) an *ad valorem* tax on sugary soft drinks (as in Chile), and (3) an excise tax on all soft drinks (as in France). The taxes are calibrated such that they generate the same relative price increase for the average product (22 percent). The simulation results show that the excise tax on sugary beverages is most effective and reduces the probability to buy a sugary soft drink by 24.4 percent. The *ad valorem* tax leads to a smaller reduction of 22.0 percent and consumers substitute to larger packaging sizes that experience a smaller price increase. The excise tax on all soft drinks leads to the smallest reduction by only 16.9 percent as it does not incentivize consumers to switch to diet soft drinks. The long-run responses to the simulated taxes are between 16 and 23 percent larger than the short-run responses. Finally, the taxes lead to a relatively uniform reduction in purchases across income groups and across household sizes.

This paper contributes to the literature that uses demand models to simulate the effect of soft drink taxes. This literature uses naturally occurring price variation and structural models to estimate demand parameters. While the literature used to rely mostly on market level data (see the surveys in Andreyeva *et al.*, 2010; Powell *et al.*, 2013), a rapidly growing number of studies uses disaggregated purchase data to estimate demand on the consumer level (Allcott *et al.*, 2019a; Dubois *et al.*, 2019; O'Connell and Smith, 2020; Wang, 2015; Bonnet and Réquillart, 2013). The use of disaggregated data has the advantage that it allows to study differentiated products and heterogeneity in consumer demand. However, the majority of these papers assumes consumption to be time-separable. Notable exceptions are Wang (2015) and Serse (2019).[1] Wang (2015) estimates a dynamic discrete choice model that explicitly models stockpiling, i.e., households stock up their inventory during price promotions and are less likely to buy when prices go up again. She finds that the price elasticity is substantially smaller compared to the results from static demand models. However, she does not consider positive state dependence arising from habit formation or addiction, which could lead to underestimation of the long-run price elasticity. Although she controls for persistent brand and sugary/diet preferences, I find that there is evidence for state dependence over and above unobserved heterogeneity. Serse (2019) estimates a mixed logit model for cola demand that incorporates habit formation. He circumvents the issue of stockpiling by modeling product choice conditional on making a purchase. He finds evidence for positive state dependence, which leads the long-run demand response to simulated taxes to be larger

[1] O'Connell and Smith (2020) provide reduced-form evidence suggesting that habit formation and stockpiling are less relevant in their dataset from the United Kingdom compared to previous studies using US data.

than the instantaneous response. However, he focuses only on the substitution between regular and diet cola beverages. In this paper, I also model the decision to make a purchase in order to provide a more comprehensive picture of the overall purchase response to taxes.

Moreover, I contribute to a literature that studies habit formation in soft drinks using other methodological approaches. Zhen *et al.* (2011) use market level data and a dynamic almost ideal demand system (AIDS). They show that long-run tax revenue is 15 to 20 percent lower than short-run revenue when habit formation is considered. Hence, they find a similar impact of habit formation compared to this study despite using different methods. Colchero *et al.* (2017) find a stronger long-term than short-term effect of the Mexican soft drink tax and conjecture that habit formation is the reason for it. Liem and de Graaf (2004) conduct randomized experiments with children and find that repeated exposure to sweetened beverages increases preference for sweetness already after a couple of days.

Finally, this paper relates to the literature that empirically estimates habit formation and state dependence. Here, habit formation is construed as intertemporal complementarities in consumption and is, following the convention in the economics literature, used synonymously with the concept of addiction (Becker and Murphy, 1988; Pollak, 1970). While the marketing literature typically understands habit formation as complementarities in brand choice probability (e.g. Keane, 1997; Dubé *et al.*, 2010; Guadagni and Little, 1983), I consider category-level habit formation. This approach is also pursued in previous studies that use both static (Tuchman, 2019) and dynamic structural models (Gordon and Sun, 2015) of addiction to analyze demand for cigarettes. In this paper, I use a static demand model, in which the previous purchase enters utility as a lagged variable. Hence, I employ a myopic model of habit formation.

The rest of the article proceeds as follows. Section 2 describes the data and Section 3 provides reduced-form evidence indicating that stockpiling and habit formation are relevant. Section 4 introduces the model and Section 5 presents the results from estimating the model, including tax simulations using the estimated parameters.

2 Data

I use weekly scanner data collected by Information Resources, Inc. (IRI, Bronnenberg *et al.*, 2008), which covers the years 2003 and 2004. The data consists of a household panel and a store panel. The household panel is comprised of households from Eau Claire, Wisconsin, and Pittsfield, Massachussetts. It tracks each household's purchases with respect to universal product code (UPC), quantity, store, and price paid. Moreover, some demographic characteristics of panelists such as income and household size are captured. The household panel is matched to the store panel based on the store, in which a household shopped in the respective week. The store panel contains store-level data on the prices charged and potential

Table 1: Market shares of soft drink products

Brand	Brand share (%)	Type	Can < 12 Cans	Can ≥ 12 Cans	Bottle < 2 Liter	Bottle ≥ 2 Liter
Coca Cola	27.96	Sugary	0.87	6.68	0.49	4.47
		Diet	0.62	9.41	0.79	4.62
Pepsi	25.09	Sugary	0.41	6.44	0.61	5.11
		Diet	0.24	6.57	0.80	4.91
Dr Pepper	2.38	Sugary	0.00	0.66	0.12	0.49
		Diet	0.00	0.84	0.06	0.21
Mountain Dew	6.68	Sugary	0.04	2.90	0.46	1.16
		Diet	0.02	1.44	0.19	0.47
Sierra Mist	4.13	Sugary	0.05	0.97	0.17	0.80
		Diet	0.02	1.41	0.13	0.59
Sprite	5.05	Sugary	0.38	1.45	0.39	1.80
		Diet	0.07	0.51	0.01	0.45
Private Label	5.61	Sugary	0.18	1.74	0.16	2.18
		Diet	0.01	0.51	0.00	0.83
Other	23.11	Sugary	1.63	5.13	1.43	5.91
		Diet	0.68	3.79	0.90	3.63
Total	100.00		5.22	50.46	6.71	37.61

Notes: Table shows market shares in percent for products purchased. Products are differentiated by brand, packaging, and sugar content.

promotional activities. This latter data allows to construct the choice set that the household faced during each shopping trip.

For the analysis, I only consider households who buy at least one soft drink product per year in the period from 2003 to 2004.[2] This leaves me with 3,062 households that make 64,222 soft drink purchases in 2013 and 59,761 purchases in 2014.[3] This dataset is merged with information on weekly shopping trips, during which no soft drink was purchased, adding a total of 159,484 store trips with no purchase. Table A.1 shows some demographic statistics about the households in the sample.

Table 1 shows the products considered and their respective market shares. A product is defined as the combination of brand, packaging type, and whether it is sugary or diet. While products in the dataset are observed on the UPC level, I follow the convention in the literature to aggregate UPCs to the brand/packaging level to construct tractable choice sets (e.g. Dubois *et al.*, 2019; Gordon *et al.*, 2013).[4] I consider the regular and diet products of eight major brands according to market share (Coca Cola, Pepsi, Dr. Pepper, Mountain Dew, Sierra

[2]Due to the computational burden the structural model is estimated using the data from 2003. The descriptive analysis uses the data from both years.

[3]If households buy more than one soft drink product per week, I keep one purchase decision per week at random.

[4]The store data provides price information for all products that were purchased during a given week on the UPC level. I calculate the price index for each brand/packaging combination by aggregating the share-weighted UPC prices per ounce in a given store and week. The share-weight of a UPC is calculated by dividing in each week the volume sold of that UPC by the total volume sold of the brand/packaging it belongs to.

Mist, Sprite) as well as those categorized as "Private Label". Together, these products have a cumulated market share of 76.9 percent and the remaining products constitute a composite brand named "Other". I further differentiate the products into cans or bottles and whether the packaging size is small (less than 12 cans and bottles of less than 2 liters) or large (at least 12 cans or bottles of 2 liters).

Using the store data, I can observe the marketing variables of all products in consumers' choice sets. First, the dataset provides information on prices of sold products in a given week. Table A.2 shows the average prices of each product. To make prices comparable, I denote prices in cents per ounce. The table shows that there is heterogeneity in prices across brands and across packaging types. Moreover, the dataset contains a variable that indicates whether products are on a temporary price reduction. The indicator equals one if the price reduction is at least 5 percent below its usual price. I use this variable as a proxy for stockpiling as discussed below.

3 Descriptive Evidence for Positive and Negative State Dependence

This section provides descriptive evidence for state dependence in soft drink demand. Two sources of state dependence in demand have been frequently identified: On the one hand, when addiction or habit formation matter, consumption in the current period increases the marginal utility of consumption in the following period (Becker and Murphy, 1988). On the other hand, when stockpiling matters, stockpiling purchases in the current period decrease the need to purchase in the next period (Hendel and Nevo, 2006). Thus, habit formation results in positive and stockpiling in negative state dependence of demand. In order to disentangle positive and negative state dependence, I follow the approach that Tuchman (2019) employs in her analysis of cigarette purchases.

I regress soft drink purchases in period t on lagged purchase variables which stand in as a proxy for habit formation and stockpiling:

$$X_{it} = \mu + \beta_1 \tilde{x}_{it-1} + \beta_2 x_{it-1} + \beta_3 \sum\nolimits_{k=1}^{4} x_{it-k} + \beta_4 \mathrm{pr}_{it-1} \tilde{x}_{it-1} + \alpha_i + \alpha_t + \epsilon_{it}, \tag{1}$$

where the dependent variable X_{it} is either the incidence of a soft drink purchase, $\tilde{x}_{it}$, or the purchase quantity, x_{it}. α_i and α_t capture household and week fixed effects, $\tilde{x}_{it-1}$ is a binary variable indicating whether the consumer bought soft drinks at all in the previous period, x_{it-1} is the amount of soft drinks bought in the previous period and $\sum_{k=1}^{4} x_{it-k}$ is a stock variable that represents the total amount of soft drinks purchased in the previous four

Table 2: Reduced form evidence for habit formation and stockpiling

	(1) Incidence	(2) Quantity	(3) Incidence	(4) Quantity
Soda purchase incidence previous week	0.019***	0.023***	0.022***	0.038***
	(0.003)	(0.008)	(0.003)	(0.009)
Soda purchase quantity previous week	-0.019***	-0.052***	-0.019***	-0.050***
	(0.001)	(0.004)	(0.001)	(0.004)
Soda purchase quantity previous 4 weeks	0.002***	0.014***	0.002***	0.014***
	(0.000)	(0.002)	(0.000)	(0.002)
Previous soda purchase was on sale			-0.009**	-0.033***
			(0.003)	(0.009)
Constant	0.438***	0.813***	0.438***	0.813***
	(0.002)	(0.007)	(0.002)	(0.007)
n	282,902	282,902	282,902	282,902
HH FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes

Notes: Table shows estimation results of regressing purchase incidence and quantity, respectively, on proxies for habit formation (incidence previous week, quantity previous 4 weeks) and stockpiling (quantity previous week, last purchase on sale). The estimations are performed on data from the years 2003 and 2004. Clustered standard errors on the household level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

weeks. pr_{it-1} indicates whether the majority of purchased soft drinks in the previous period were on sale.

If habit formation matters, $\tilde{x}_{it-1}$ captures positive dependence between purchases of the previous period and the current period. However, if consumers stockpile, buying *more* in the previous period (x_{it-1}) decreases the probability that they buy in the current period. Conditional on the last period, soft drink purchases of the four preceding weeks (X_{it}) are arguably not so much affected by stockpiling and, thus, mainly capture the influence of habit formation. Finally, if a product was purchased on sale (pr_{it-1}), the likelihood is larger that it was stockpiled.

Table 2 shows the regression results. In the first column, I use purchase incidence as dependent variable. The probability to purchase soft drinks in the current week is larger if a consumer purchased in the previous week, but having purchased *more* soft drinks in the week before is associated with a lower propensity to purchase. Thus, demand shows positive persistence in general, but proxies for stockpiling suggest a reduced need to purchase. Moreover, the stock of purchases in the previous four weeks is associated with a higher propensity to purchase. These estimates support the notion that both habit formation and stockpiling are relevant. The results are supported when using purchase quantity as dependent variable in the second column. Again, consistent with addiction and stockpiling, purchase incidence in the preceding week and purchase quantity in the preceding four weeks are associated with a larger quantity of purchased. In contrast, the larger the purchase quantity in the previous week, the smaller the quantity in the current week. All coefficients have the same sign as the ones Tuchman (2019) obtains for cigarette purchases.

In the third and fourth columns, I additionally control whether the household purchased a soft drink that was on sale during the previous purchase occasion. The third column shows that the proxies for positive and negative state dependence are similar as before and that a consumer is indeed less likely to make a soft drink purchase if she bought a soft drink during a sale the period before. The fourth column gives a similar picture when considering the total quantity purchased in a given week. While the coefficient for positive state dependence becomes more pronounced, having purchased during a price reduction the period before is associated with a lower quantity in the current period.

These results lead me to conclude that a model of soft drink demand should incorporate both positive and negative state dependence. Hence, in the next section, a model is introduced that incorporates both habit formation and stockpiling.

4 Model

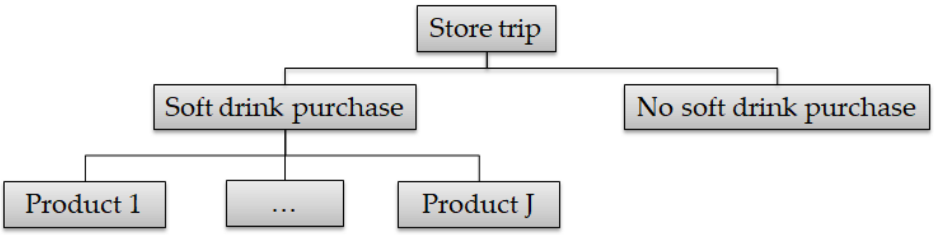
Conditional on a store trip in week t, a consumer i, who belongs to income group $h(i)$, maximizes her utility from buying product j, which belongs to the sugary/diet segment $s(j)$:

$$\begin{aligned} U_{ijt} =& \theta_i x_j + \alpha_{h(i)} p_{jt} + Z_i \beta + \xi_t + \kappa_{s(j)} \text{purch}_{is(j),t-1} + \zeta \text{pr}_{ijt} \\ &+ \delta_{s(j)} \text{purch}_{is(j),t-1} \text{pr}_{is(j),t-1} + \gamma_{s(j)} \text{purch}_{is(j),0} + \epsilon_{ijt}. \end{aligned} \tag{2}$$

θ_i is a vector of consumer-specific, time-constant preferences for product characteristics x_j. The latter consist of a general preference for soft drinks, as well as brand, diet and packaging (large vs. small and can vs. bottle) fixed effects. For identification, the brand fixed effect of "Private Label" is normalized to zero. p_{jt} is the price of product j in week t and price sensitivity $\alpha_{h(i)}$ is allowed to vary by income group $h(i)$. Z_i contains income and household size as consumer-specific demographics. Thus, the general propensity to buy a soft drink depends on income and household size. ξ_t are quarterly fixed effects. $\kappa_{s(j)}$ captures the influence of state dependence which is allowed to differ between sugary and diet soft drinks. $\text{purch}_{is(j),t-1}$ indicates if a consumer purchased a sugary or diet soft drink in the previous period. pr_{ijt} indicates if a product is on sale and ζ measures the impact of the sale on purchase probability over and above the pure price effect. Moreover, the lagged purchase indicator is interacted with $\text{pr}_{is(j),t-1}$, which indicates whether the previous purchase of a sugary or diet product took place during a sale.[5] Hence, $\delta_{s(j)}$ is supposed to capture stockpiling. Finally, $\text{purch}_{is(j),0}$ indicates for sugary and diet products if a consumer purchased a product of the

[5] The dummy variable "price reduction" is given in the IRI data and indicates a temporary price reduction of at least 5 percent. If a consumer bought more than one product, the dummy equals one if at least half

Figure 1: Nesting structure of empirical model



respective segment in the first observed period. By controlling for the first observed decision and exogenous covariates, Z_i, I aim to deal with the initial conditions problem (Wooldridge, 2005). I assume that the idiosynratic, unobserved shocks to utility, ϵ_{ijt}, are type 1 extreme value distributed. The outside option to not make a soft drink purchase is normalized to $U_{ijt} = \epsilon_{ijt}$.

I estimate the model using nested logit, as shown in Figure 1. Conditional on a store trip, a consumer first decides whether to purchase a soft drink or not, and, in case of a purchase, which product to buy. Thereby, I allow unobserved shocks to be correlated in case of a soft drink purchase but not with the outside option. This relaxes the independence of irrelevant alternatives (IIA) assumption.[6] In such a model, a price increase of a soft drink leads to a proportionate shift in the probabilities to buy soft drinks but not necessarily in the probability to choose the outside option. To allow more realistic substitution patterns between products, I allow for random coefficients in the product specific preferences θ_i. This allows products that are closer in characteristics space to be closer substitutes to each other. The unobserved heterogeneity is estimated non-parametrically using discrete mass points as described below.

Likelihood. Following Train (2003), I decompose U_{ijt} into utility components that are constant over products and the outside option, V_{it}, and components that depend on variables that vary between products, W_{ijt}. The individual likelihood contribution of consumer i over all weeks and choice alternatives can be written as

(3)
$$L_i = \prod_t \prod_j \left(\frac{exp(V_{it} + \lambda I_{it})}{1 + exp(V_{it} + \lambda I_{it})} \right)^{T_{it}} \left(\frac{1}{1 + exp(V_{it} + \lambda I_{it})} \right)^{1-T_{it}} \left(\frac{exp(W_{ijt}/\lambda)}{\sum_k exp(W_{ikt}\lambda)} \right)^{Y_{ijt}},$$

where T_{it} indicates whether individual i makes a soft drink purchase in a week with a store visit t and Y_{ijt} indicates whether individual i buys product j in week t. $I_{it} =$

the quantity purchased was on sale. Note that $pr_{is(j),t-1}$ refers to the previous calendar week as stockpiling should not be relevant for longer periods.

[6] The IIA assumption implies that, e.g., a price change of one particular soft drink leads to a proportional change in the purchase probability of all other soft drinks and the outside option.

$\ln(\sum_k exp(W_{ikt}/\lambda))$ is the inclusive value of making a soft drink purchase and links the purchase incidence decision and the product choice. It represents the expected utility that an individual receives from choosing a product. Its coefficient λ approximates the dissimilarity of alternatives within a nest (Train, 2003). If $\lambda = 1$, the unobserved utilities of alternatives are completely independent and the choice probabilities become logit. If λ lies within the unit interval, the model is consistent with random utility maximization (McFadden, 1978).

Consumer Heterogeneity. In order to identify true from spurious state dependence, it is crucial to allow for a rich and flexible structure of heterogeneity (Heckman, 1981; Keane, 1997). Here, I allow for random coefficients in the persistent preferences for each brand, as well as tastes for diet soft drinks (in contrast to sugary soft drinks), large packaging sizes, and cans (in contrast to bottles). The random coefficients are estimated non-parametrically with M discrete mass points (Heckman and Singer, 1984). The random coefficient vector is constant for each mass point. Hence, the mass points can be interpreted as different consumer types. The share of consumer type m is denoted by π^m:

$$\pi^m = \frac{exp(q^m)}{1 + \sum_{m'=2}^{M} exp(q^{m'})}, \quad \sum_{m=1}^{M} \pi^m = 1 \tag{4}$$

where the parameters q^m have to be estimated. For identification, q^1 is set to zero. The number of mass points is determined by successively increasing the number of types and evaluating the Bayesian Information Criterion (BIC) until there is no further increase.

The log-likelihood over all consumers thus becomes

$$LogL = \sum_{i=1}^{I} \log(\sum_{m=1}^{M} \pi_i^m L_i). \tag{5}$$

Empirical Bayes. To be able to perform simulations, I have to allocate values to the the random intercepts and to the random coefficients. Therefore, I assign each consumer to one of the discrete types m by applying Bayes rule (Bucklin and Gupta, 1992; Skrondal and Rabe-Hesketh, 2004). The average type probability in the sample π^m can be seen as the *prior* probability of belonging to a type. The posterior probability p_i^m is calculated using the likelihood of the purchase history conditional on each type:

$$p_i^m = \frac{\pi^m L_i^m}{\sum_{n=1}^{M} \pi^n L_i^n}. \tag{6}$$

where L_i^m is the likelihood of consumer i's observed purchase history given that she belongs to type m. I assign consumers the vector of random coefficients that belongs to the type m with the highest posterior, p_i^m.

Identification. Since the ultimate goal of the model is to perform tax simulations, a key parameter to be identified is the price sensitivity. I exploit price variation over time and across retailers (following Dubois *et al.*, 2019). First, there is variation in the choice sets that consumers face since they make purchases at different retailers over time. This variation stems from different cost structures, varying product availability and occasional price reductions that I assume to be random. As I do not model the store choice, I have to assume that consumers do not select the store based on the prices of soft drinks. The second source of price variation is non-linear pricing within brands and across package sizes. As can be seen in Table A.2, cans are typically more expensive than plastic bottles and retailers offer discounts for larger sizes. The extent of non-linear pricing differs between retailers and over time.

The identifying assumption is that, conditional on product-specific controls, the remaining price variation is exogenous. This assumption implies that demand shocks are not correlated with price changes. To make this assumption more plausible, I include (quarterly) time fixed effects in the decision to make a purchase. Thereby, I control for product-invariant, aggregate demand shocks (e.g. christmas) that are anticipated by the retailer in their pricing decision.

Moreover, the model aims to separately identify habit formation and stockpiling as sources for positive and negative state dependence. Therefore, I assume that consumers predominantly stockpile during sales periods. This is reasonable since they could otherwise wait until the next period allowing them, first, to not incur the storage costs, and, second, to hope for a price reduction in the next period.[7] Under this assumption, I can interact the lagged purchase indicator and the lagged price reduction indicator to take out the stockpiling component of state dependence. The remaining state dependence is interpreted as the impact of habit formation. This interpretation requires to assume that, conditional on the rich set of control variables, the lagged purchase coefficient is not correlated with any remaining unobserved preference heterogeneity. Under this assumption, habit formation can be identified from variation in choice sets and prices, which leads individuals to deviate from their purchasing pattern with longer lasting effects.

[7] This argument abstracts from transaction costs associated with making a shopping trip. However, in the dataset I observe most households making at least one shopping trip per week. If households make a shopping trip anyway, the additional costs of buying a soft drink are small.

5 Results

5.1 Parameter Estimates

Table 3 presents the estimated coefficients of the models with homogenous and heterogenous preferences. In the model with homogenous preferences, I do not account for unobserved consumer heterogeneity and assume all individual-specific coefficients to be fixed. In the model with heterogenous preferences, I introduce random coefficients for the taste parameters with three mass points, as described in Section 4. The first column of the heterogenous tastes model shows the baseline coefficients and the second and third column the interaction terms for the respective types. The average probability to belong to each type is 42.6, 38.4 and 19.1 percent, respectively. When introducing random coefficients, the Bayesian Information Criterion (BIC) substantially increases, hence, the model with heterogenous preferences is the preferred model.[8]

Panel A of Table 3 shows the variables that relate to product choice. In both models, the price coefficients are negative and the richest households are slightly less price sensitive. The indicator for habit formation (purch_{t-1}) is positive and significant in both specifications. However, the magnitude of state dependence is smaller in the model with heterogenous preferences. This suggests that the coefficient in the homogenous model picks up persistent taste heterogeneity. Such "spurious state dependence" (Heckman, 1981) reiterates that it is important to account for unobserved heterogeneity. Moreover, the coefficient indicating a purchase on sale in the previous period is negative, consistent with the idea that an individual who stockpiled in the previous period is less likely to purchase in this period. Regarding product preferences, there is substantial heterogeneity in brand and diet/sugary preferences when introducing consumer heterogeneity.

Panel B of Table 3 displays the variables that describe the decision to make a soft drink purchase. The coefficient of the inclusive value is close to one in the homogenous tastes model and smaller than one in the model with heterogenous tastes. This suggests that the model with heterogenous tastes is consistent with random utility maximization. The probability to make a purchase increases with the household size. Moreover, the constant, which measures a general preference for soft drinks, exhibits considerable heterogeneity between types.

5.2 Model Fit

In order to assess the in-sample fit of the estimated model, I simulate purchase decisions given the estimation sample. First, I assign values to the random intercepts and random coefficients using the "Empirical Bayes" approach described in Section 4. The resulting distribution of

[8] I choose the model with three mass points since the BIC is larger compared to a model with two mass points, whereas a model with a fourth mass point could not be identified given the data at hand.

Table 3: Parameter estimates

	Homogenous preferences	Heterogenous preferences		
		Baseline	Type 2	Type 3
Panel A: Product choice				
Price (HH inc=1)	-0.376	-0.294		
	(0.010)	(0.009)		
Price (HH inc=2)	-0.387	-0.306		
	(0.008)	(0.008)		
Price (HH inc=3)	-0.354	-0.285		
	(0.008)	(0.008)		
Purch_{t-1} (sugary)	0.962	0.894		
	(0.022)	(0.021)		
Purch_{t-1} (diet)	1.202	0.696		
	(0.025)	(0.025)		
Price reduction (pr)	0.179	0.141		
	(0.011)	0.010		
Purch_{t-1} x pr_{t-1} (sugary)	-0.092	-0.094		
	(0.024)	(0.023)		
Purch_{t-1} x pr_{t-1} (diet)	-0.142	-0.182		
	(0.028)	(0.027)		
Diet Drinks	-0.180	-0.903	1.139	1.201
	(0.014)	(0.034)	(0.040)	(0.045)
Can Package	0.517	0.077	1.539	-0.853
	(0.013)	(0.025)	(0.048)	(0.045)
Large Size	0.412	-0.002	0.901	0.209
	(0.020)	(0.025)	(0.047)	(0.049)
Coca Cola	1.776	0.378	2.296	1.392
	(0.036)	(0.037)	(0.080)	(0.063)
Pepsi	0.026	-0.042	0.740	-0.080
	(0.029)	(0.036)	(0.079)	(0.068)
Dr Pepper	1.681	0.481	1.910	1.362
	(0.034)	(0.038)	(0.079)	(0.063)
Mountain Dew	0.378	-0.514	2.266	-0.223
	(0.028)	(0.049)	(0.090)	(0.091)
Sierra Mist	-0.166	-0.543	1.507	0.160
	(0.031)	(0.045)	(0.083)	(0.076)
Sprite	-0.616	-1.294	2.157	0.075
	(0.038)	(0.072)	(0.104)	(0.144)
Other	1.363	0.912	1.012	0.249
	(0.030)	(0.034)	(0.065)	(0.054)
Panel B: Purchase incidence				
Inclusive value	1.041	0.869		
	(0.016)	(0.018)		
HH size	0.177	0.182		
	(0.005)	(0.006)		
HH inc=2	0.050	0.030		
	(0.041)	(0.037)		
HH inc=3	-0.104	-0.079		
	(0.043)	(0.039)		
Constant	-5.010	-3.618	-3.035	0.196
	(0.080)	(0.066)	(0.112)	(0.065)
Type Share	1.000	0.426	0.384	0.191
BIC	-499,813	-472,696		

Notes: Table shows the estimated coefficients for a model with homogenous preferences (i.e. only fixed coefficients) and a model with heterogenous preferences (i.e. random coefficients with three discrete mass points). All models also control for quarter fixed effects and the observed purchase decision in the first period to account for the initial conditions problem. The model is estimated on the data from 2003. Standard errors are in parentheses.

Table 4: Sample fit of main model

	Observed (%)	Estimated (%)
Purchase incidence	46.88	46.02
Sugary products	55.24	59.98
Can packaging	55.68	54.55
Large sizes	88.07	87.77
Market shares (%)		
Coca Cola	25.55	27.96
Pepsi	5.77	5.05
Dr Pepper	23.35	25.09
Mountain Dew	6.34	6.68
Sierra Mist	4.09	4.13
Sprite	2.14	2.38
Private Label	6.97	5.61
Other	25.80	23.11

Notes: Table shows actually observed frequencies ("Observed") and simulated probabilities from the model with heterogenous preferences ("Estimated"). In the simulations, the first decision of each consumer is taken as given and the subsequent decisions are forward simulated conditional on the observed state space and the choice probabilities of the previous period. Choice probabilities are averaged over consumers.

types closely resembles the average type probabilities in Table 3: The frequency of assigned types is 42.5 percent for type 1, 38.5 percent for type 2, and 19.0 percent for type 3. The average posterior probability of belonging to the assigned type is 95.9 percent.[9] Next, I take the first purchase decision as given and forward simulate the purchase decisions for all following periods. That means, in each period, I simulate the purchase decision conditional on the observed choice set (including product availability, prices, and price reductions) and the simulated choice probabilities in the previous period (that enter as lagged variables).

Table 4 shows the sample fit of the model. It can be seen that both the probability to make a purchase and the probability to buy a respective product are close to the empirically observed frequencies. Similarly, the estimated market shares of brands are close to the empirically observed market shares. This indicates that the model fits the data well.

5.3 Short- and Long-run Elasticities

To illustrate the impact of state dependence, I use the estimates from the model to simulate short- and long-run price elasticities of soft drink demand. In order to compute elasticities, I simulate the purchase decision a consumer makes with and without a 1 percent increase in prices. In the first simulated period, the price change affects the purchase probability only via the instantaneous utility. Hence, the arc elasticity in the first period is interpreted as *short-run elasticity*.[10] In the following periods, I forward simulate the purchase decision, taking

[9] Alternatively, the posterior probabilities could also be used as weights for the assignment of the estimated random coefficients (e.g. Haan, 2010). However, since the posterior is close to one for most consumers, the results would only change very marginally.

[10] A different notion of short-run elasticity is the impact of a 1 percent price change that does not affect the lagged purchase probability in the next period (instead the actually observed purchase decisions enter the state space as lagged variable). This approach yields a very similar short-run elasticity.

Table 5: Price elasticities over time (sugary soft drinks)

	Sugary soft drinks			All soft drinks		
Period	Price Elasticity	Difference to Period 1		Price Elasticity	Difference to Period 1	
1	0.87			0.33		
2	1.02	0.15	(0.03)	0.38	0.05	(0.01)
3	1.05	0.18	(0.04)	0.39	0.06	(0.02)
4	1.07	0.20	(0.04)	0.39	0.06	(0.02)
5	1.08	0.21	(0.04)	0.39	0.06	(0.02)
6	1.07	0.20	(0.04)	0.39	0.06	(0.02)
7	1.07	0.20	(0.04)	0.39	0.06	(0.03)
8	1.06	0.19	(0.04)	0.38	0.05	(0.03)
9	1.07	0.20	(0.04)	0.39	0.06	(0.02)
10	1.07	0.20	(0.04)	0.39	0.06	(0.02)

Notes: Table shows the respective elasticity over time starting with the period of the price increase. The elasticity is computed as the difference between simulated purchase probabilities for baseline prices and for a 1 percent increase in sugary soft drink prices. The simulations are performed for price increases at ten different points in time and elasticities are averaged over these simulations. Choice probabilities are averaged over consumers. Standard errors of the difference compared to the first period are bootstrapped with 200 replications and shown in parentheses.

into account the changed purchase probability in the preceding periods.[11] Since the model includes a lagged dependent variable, it can be described by a first-order Markov process, which is known to converge in the long run. I call the price elasticity that constitutes the limiting value after a number of periods *long-run elasticity.*

Table 5 shows the change in purchase probabilities when the prices of sugary soft drinks are increased by 1 percent. Since state dependence in this model is specific to sugary and diet beverages, there are two ways in which state dependence can lead to differing short- and long-run elasticities. First, individuals can get used to not consuming soft drinks altogether, and, second, individuals can get used to consuming diet instead of sugary soft drinks. Table 5 shows that purchases of sugary soft drinks decrease by 0.87 percent in the short run and after ten periods the drop in purchases increases to 1.07 percent. Hence, the long-run elasticity is more than 20 percent larger than the short-run elasticity. The bootstrapped standard errors show that the difference between short- and long-run elasticity is already significant after the first period.[12] Table 5 also shows that a 1 percent price increase of sugary soft drinks leads to a drop in overall soft drink purchases by 0.33 percent, an effect that grows to 0.39 percent over time.

[11] Since the reaction to a price increase partly depends on the market environment in the period of the price increase (e.g. due to the prices in the respective period and the availability of price reductions), I run ten simulations, in which I introduce the price increase at different points in time. The elasticities starting from the period of the price increase are averaged over simulations.

[12] The bootstrap is performed by repeatedly drawing new parameter vectors from the multivariate normal distribution that is characterized by estimated parameters as distribution mean and by the estimated covariance matrix of parameters.

Table 6: Tax simulations, change in purchase probabilities (in percent)

	Excise tax (sugary, 0.5 ct/oz)		Ad valorem tax (sugary, 22%)		Excise tax (all drinks, 0.5 ct/oz)	
	Short-run	Long-run	Short-run	Long-run	Short-run	Long-run
Panel A: Sugary and diet						
Sugary soft drink	−20.13	−24.39	−17.84	−21.99	−14.56	−16.90
	(0.64)	(0.66)	(0.61)	(0.63)	(0.72)	(0.75)
Diet soft drink	7.68	10.76	6.79	9.65	−12.69	−13.34
	(0.90)	(0.95)	(0.88)	(0.93)	(0.86)	(0.89)
Any soft drink	−7.71	−8.74	−6.84	−7.90	−13.72	−15.31
	(0.49)	(0.50)	(0.47)	(0.48)	(0.57)	(0.60)
Panel B: Packaging and brand						
Sugary, large size	−20.05	−24.46	−16.16	−19.93	−14.41	−16.72
	(0.65)	(0.67)	(0.63)	(0.65)	(0.67)	(0.71)
Sugary, small size	−20.03	−24.14	−29.30	−32.39	−15.67	−18.33
	(1.31)	(1.30)	(1.16)	(1.14)	(1.36)	(1.35)
Sugary, private label	−19.89	−23.96	−8.31	−12.87	−16.13	−18.92
	(1.57)	(1.52)	(1.76)	(1.70)	(1.62)	(1.59)

Notes: Table shows the change in simulated purchase probabilities after imposing the respective tax compared to the baseline simulation. In the simulations, the first decision of each consumer is taken as given and the subsequent decisions are forward simulated conditional on the observed state space and the choice probabilities of the previous period. Choice probabilities are averaged over consumers. Long-run price elasticities are measured ten weeks after the tax introduction. The tax is implemented in ten different weeks and elasticities are averaged over these ten weeks. Bootstrapped standard errors with 200 replications are in parentheses.

5.4 Soft Drink Tax Simulations

Using the estimates from the model, I simulate the impact of different taxes on demand. The considered taxes resemble taxes that are actually implemented in jurisdictions around the world. First, I consider an excise tax imposed on all sugary beverages as implemented in, among others, Mexico and in many cities in the United States. For example, Berkeley implemented a 1 cent per ounce tax in 2015 and most studies find that less than half of the tax was passed through to prices (Cawley *et al.*, 2019b). Hence, I simulate the effect of a 0.5 cent per ounce price increase. Considering the mean price of all products weighted by their market shares, this amounts to an average price increase of 22 percent in my sample. Second, I study the impact of a 22 percent *ad valorem* tax on sugary beverages. The tax rate is chosen to raise the price of the mean product by the same magnitude as the excise tax. *Ad valorem* taxes are implemented, for example, in Thailand and Chile (see e.g. Nakamura *et al.*, 2018). Third, I simulate the impact of a 0.5 cent per ounce excise tax on all soft drinks irrespective of whether they are sugary or diet. Soft drink taxes that do not distinguish between sugary and diet are implemented, for instance, in France and Philadelphia, Pennsylvania (Capacci *et al.*, 2019; Cawley *et al.*, 2019a).

Table 6 shows the tax simulation results. While the excise tax on sugary beverages leads to an immediate drop in purchases of 20.1 percent, the effect increases up to 24.4 percent after ten weeks. Hence, the long-run effect is approximately 20 percent larger than the short-run effect. The impact of the tax can be decomposed into substituting to diet soft drinks and stopping to purchase soft drinks altogether. While there is an immediate increase in diet purchases by 7.7 percent, the effect grows substantially over time up to 10.8 percent. The probability to purchase any soft drink decreases by 7.7 percent and increases slightly up to 8.7 percent in the long run.

In Panel B of Table 6, I analyze if consumers substitute – within the sugary soft drink category – to cheaper soft drinks. As can be seen in Table A.2, larger package sizes and private label products are, for example, cheaper on average. However, we observe that the volumetric excise tax induces a uniform reduction in purchases across packaging types and brands.

The third and fourth column of Table 6 show the effect of the *ad valorem* tax on purchase probability. Although the *ad valorem* tax increases prices of the average product by the same extent as the excise tax, it is less effective in decreasing demand. The purchase probability decreases only by 17.8 percent in the short run and 22.0 percent in the long run. In contrast to the excise tax, there is less substitution to diet beverages and consumers are less likely to stop consuming soft drinks altogether. Panel B illustrates the reason for this pattern. Since the *ad valorem* tax leads to smaller price changes of cheaper products, consumers substitute to products that offer more value for money. While they strongly reduce their purchases of small packages (which are relatively expensive), they reduce their purchases of large sizes less strongly. This substitution decreases the effectiveness of the tax further as larger sizes contain more sugar.[13] Moreover, consumers substitute to the cheaper private label products.

Finally, in the fifth and sixth column, I simulate the impact of an excise tax of 0.5 cents per ounce on all soft drinks, irrespective of whether they are sugary or diet. As expected, the tax leads to a reduction in the purchase probability of both sugary and diet soft drinks. While the probability to buy a sugary soft drink decreases by 14.6 percent (16.9 percent) in the short run (long run), the probability to purchase a diet soft drink decreases by 12.7 percent (13.3 percent) in the short run (long run). Among all taxes, the reduction in the probability to buy a sugary soft drink is the smallest for the excise tax on all soft drinks. This is explained by the observation that the tax on all soft drinks discourages consumers from substituting to diet soft drinks, making this tax the least effective in reducing sugar consumption.

Table 7 addresses the question if taxes have heterogenous effects on different consumer groups. Since the discrete choice framework allows for observed and unobserved heterogeneity in consumer preferences, I can analyze heterogenous effects for different demographic

[13]While the small sizes of cans and bottles in the dataset contain on average 59.1 and 25.2 ounces, respectively, the large sizes contain 195.6 and 78.9 ounces. Thus, since the amount of sugar per ounce does not differ, large sizes contain substantially more sugar.

Table 7: Heterogenous changes in purchase probabilities of sugary soft drinks (in percent)

	Excise tax (sugary, 0.5 ct/oz)		Ad valorem tax (sugary, 22%)		Excise tax (all drinks, 0.5 ct/oz)	
	Short-run	Long-run	Short-run	Long-run	Short-run	Long-run
Panel A: Income						
Low Income	−19.96	−24.23	−17.63	−21.14	−14.81	−17.12
	(1.01)	(1.01)	(0.97)	(0.96)	(1.06)	(1.09)
Medium Income	−20.69	−25.16	−18.18	−21.87	−15.10	−17.51
	(0.74)	(0.76)	(0.70)	(0.71)	(0.78)	(0.80)
High Income	−19.14	−23.45	−17.08	−20.72	−13.64	−15.90
	(0.88)	(0.89)	(0.86)	(0.87)	(0.91)	(0.94)
Panel B: Household size						
Household size ≤ 2	−18.41	−23.20	−16.29	−20.28	−12.71	−15.15
	(0.75)	(0.77)	(0.72)	(0.74)	(0.77)	(0.81)
Household size > 2	−21.48	−25.49	−18.97	−22.29	−16.19	−18.42
	(0.69)	(0.69)	(0.66)	(0.66)	(0.72)	(0.75)

Notes: Table shows the change in simulated purchase probabilities of sugary soft drinks after imposing the respective tax compared to the baseline simulation. In the simulations, the first decision of each consumer is taken as given and the subsequent decisions are forward simulated conditional on the observed state space and the choice probabilities of the previous period. Choice probabilities are averaged over consumers. Long-run price elasticities are measured ten weeks after the tax introduction. The tax is implemented in ten different weeks and elasticities are averaged over these ten weeks. Bootstrapped standard errors with 200 replications are in parentheses.

groups.[14] I focus on subgroups that have received particular attention in the policy debate: poor consumers and households with children.

First, I focus on the regressivity of the tax by analyzing differential tax responsiveness by income. Panel A of Table 7 shows that the reduction in purchase probability is relatively uniform across the income distribution. There is some indication that the richest consumers reduce their purchases the least but the differences are small and insignificant. These findings support the result from previous studies that price elasticities of poor consumers are at least as high as those of rich consumers (Dubois *et al.*, 2019; Wang, 2015). Thus, the regressivity of soft drink taxes is alleviated when taking into account that poor individuals have larger health improvements from improving their diet (Allcott *et al.*, 2019a).

Second, I differentiate the tax responsiveness into households that have more than two household members or not. I take that as a proxy for households with or without children (assuming that in most households with more than two members, there are children living in the household). Soft drink consumption by children is seen as particularly problematic as many consumption habits are formed early in life (Mennella *et al.*, 2016). Hence, soft drink taxes can be seen as well targeted if they lead households with children to reduce their propensity to consume soft drinks (Dubois *et al.*, 2019). Panel B of Table 7 shows that larger households respond to the taxes slightly more than smaller households. This suggests that

[14] On the one hand, the price coefficient in consumer's utility is allowed to differ by income group. On the other hand, I assign consumers to unobserved tastes based on their purchasing patterns. If consumers' preferences differ by demographic group, unobserved types will be distributed accordingly.

the tax is well targeted in reducing the sugar purchases of households with children. However, the difference in the responses between households is relatively small.

6 Conclusion

There is evidence from biology and economics that sugar and soft drink consumption is habit forming. Nevertheless, most demand models of soft drink consumption do not take such inter-temporal complementarities into account. This paper incorporates habit formation into the analysis of soft drink taxes. Using scanner data from the US, I find that there is strong reduced-form evidence for both habit formation and stockpiling in soft drink purchases. I estimate a discrete choice model that includes habit formation and stockpiling, the latter being proxied by purchases of price reduced products. Using the estimates from the model, I simulate the short- and long-run responses to different soft drink taxes.

The results show that long-run price elasticities are approximately 20 percent larger than short-run elasticities. In the tax simulations, I find that an excise tax on sugary beverages is more effective in reducing purchases than an *ad valorem* tax. Although both increase the price of the average product by the same percentage, an excise tax reduces purchases more since it gives less incentives to substitute to cheaper products. An excise tax on all products is the least effective since it discourages substitution to diet soft drinks. The reduction in purchases are relatively uniform across demographic groups.

These results qualify the findings in Wang (2015) who has shown that not taking into account stockpiling leads to overestimation of price elasticities. I show that not taking into account habit formation can in turn lead to underestimation of long-run elasticities. Hence, future models of soft drink demand should ideally take both positive and negative state dependence into account. Moreover, policy evaluations of soft drink taxes should consider a sufficiently long post-treatment period to capture the entire treatment effect.

In this paper, I employ a myopic model of habit formation. If consumers are instead forward-looking, they will anticipate the intertemporal complementarities of consuming soft drinks. It would be interesting to assess the implications of a dynamic model of habit formation, in which consumers take into account that instantaneous consumption affects their future utility. This questions is left for further research.

References

Ahmed, S. H., Guillem, K. and Vandaele, Y. (2013). Sugar addiction. *Current Opinion in Clinical Nutrition and Metabolic Care*, **16** (4), 434–439.

Allcott, H., Lockwood, B. B. and Taubinsky, D. (2019a). Regressive sin taxes, with an application to the optimal soda tax. *Quarterly Journal of Economics*, **134** (3), 1557–1626.

—, — and — (2019b). Should we tax sugar-sweetened beverages? An overview of theory and evidence. *Journal of Economic Perspectives*, **33** (3), 202–227.

Andreyeva, T., Long, M. W. and Brownell, K. 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.

Becker, G. S. and Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy*, **96** (4), 675–700.

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

Bronnenberg, B. J., Kruger, M. W. and Mela, C. F. (2008). Database paper: The IRI marketing data set. *Marketing Science*, **27** (4), 745–748.

Bucklin, R. E. and Gupta, S. (1992). Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *Journal of Marketing Research*, **29**, 201–215.

Capacci, S., Allais, O., Bonnet, C. and Mazzocchi, M. (2019). The impact of the French soda tax on prices and purchases. an ex post evaluation. *PLOS One*, **14** (10), e0223196.

Cawley, J., Frisvold, D., Hill, A. and Jones, D. (2019a). The impact of the philadelphia beverage tax on purchases and consumption by adults and children. *Journal of Health Economics*, **67**, 10225.

—, Thow, A. M., Wen, K. and Frisvold, D. (2019b). 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.

Colchero, M. A., Rivera-Dommarco, J., Popkin, B. M. and Ng, S. W. (2017). In Mexico, evidence of sustained consumer response two years after implementing a sugar-sweetened beverage tax. *Health Affairs*, **36** (3), 564–571.

Dubois, P., Griffith, R. and O'Connell, M. (2019). How well targeted are soda taxes? *TSE Working Papers*, 17-868.

Dubé, J.-P., Hitsch, G. J. and Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics*, **41** (3), 417–445.

Gordon, B. R., Goldfarb, A. and Li, Y. (2013). Does price elasticity vary with economic growth? a cross-category analysis. *Journal of Marketing Research*, **50** (1), 4–23.

— and Sun, B. (2015). A dynamic model of rational addiction: Evaluating cigarette taxes. *Marketing Science*, **34** (3), 452–470.

Guadagni, P. M. and Little, J. D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, **2** (3), 203–238.

Haan, P. (2010). A multi-state model of state dependence in labor supply: Intertemporal labor supply effects of a shift from joint to individual taxation. *Labour Economics*, **17**, 323–335.

Heckman, J. and Singer, B. (1984). A method for minimizing the distributional assumptions in econometric models for duration data. *Econometrica*, **52**, 271–320.

Heckman, J. J. (1981). Heterogeneity and state dependence. In S. Rosen (ed.), *Studies in Labor Markets*, Chicago: University of Chicago Press, pp. 91–140.

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

Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics*, **15** (3), 310–327.

Liem, D. and de Graaf, C. (2004). Sweet and sour preferences in young children and adults: role of repeated exposure. *Physiology & Behavior*, **83** (3), 421–429.

McFadden, D. (1978). Modelling the choice of residential location. In F. S. A. Karlqvist and J. Weibull (eds.), *Spatial Interaction Theory and Residential Location*, Amsterdam; New York: North Holland Pub. Co., pp. 75–96.

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

Nakamura, R., Mirelman, A. J., Cuadrado, C., Silva-Illanes, N., Dunstan, J. and Suhrcke, M. (2018). Evaluating the 2014 sugar-sweetened beverage tax in chile: An observational study in urban areas. *PLOS Medicine*, **15** (7), e1002596.

O'Connell, M. and Smith, K. (2020). Corrective tax design in oligopoly, `https://martinoconnell85.github.io/Mywebsite/CorrectiveTaxDesigninOligopoly.pdf` (retrieved 05/01/2020).

Pollak, R. A. (1970). Habit formation and dynamic demand functions. *Journal of Political Economy*, **78** (4), 745–763.

Powell, L. M., Chriqui, J. F., Khan, T., Wada, R. and Chaloupka, F. J. (2013). Assessing the potential effectiveness of food and beverage taxes and subsidies for improving public health: a systematic review of prices, demand and body weight outcomes. *Obesity Reviews*, **14** (2), 110–128.

Serse, V. (2019). Do sugar taxes affect the right consumers?, cORE DP 2019/17.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling*. Boca Raton, Florida: Chapman and Hall.

Train, K. (2003). *Discrete Choice Models using Simulation*. Cambridge, UK: Cambridge University Press.

Tuchman, A. E. (2019). Advertising and demand for addictive goods: The effects of e-cigarette advertising. *Marketing Science*, **38** (6), 913–1084.

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

WHO (2016). *Fiscal Policies for Diet and Prevention of Noncommunicable Diseases*. Technical meeting report, World Health Organization, Geneva.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, **20** (1), 39–54.

Zhen, C., Wohlgenant, M. K., Karns, S. and Kaufman, P. (2011). Habit formation and demand for sugar-sweetened beverages. *American Journal of Agricultural Economics*, **93** (1), 175–193.

A Additional tables

Table A.1: Descriptive statistics

	2013	2014
Avg income per capita (K)	21.600	21.590
	(14.552)	(14.577)
Avg household size	2.528	2.535
	(1.248)	(1.242)
Avg weekly volume purchased (ounces)	214.922	197.177
	(194.611)	(188.237)
Avg weekly dollars spent	4.281	4.161
	(3.859)	(3.892)
Number of Households	3062	3062

Notes: Tables shows descriptive statistics of the sample for the years 2013 and 2014. The income is calculated by assigning households the midpoint of its discrete income bracket. Standard deviations are in parentheses.

Table A.2: Average prices of soft drink products (in cents per ounce)

		Can		Bottle	
Brand	Type	$<$ 12 Cans	$\geq$ 12 Cans	$<$ 2 Liter	$\geq$ 2 Liter
Coca Cola	Sugary	4.18	2.35	5.77	1.87
		(0.69)	(0.46)	(0.50)	(0.31)
	Diet	4.36	2.39	5.68	1.96
		(0.70)	(0.46)	(0.45)	(0.32)
Pepsi	Sugary	4.36	2.31	5.46	1.80
		(0.76)	(0.45)	(0.77)	(0.29)
	Diet	4.30	2.34	5.48	1.87
		(0.78)	(0.45)	(0.73)	(0.31)
Dr Pepper	Sugary	4.52	2.50	5.44	1.89
		(0.60)	(0.43)	(0.45)	(0.27)
	Diet	3.21	2.51	5.48	1.83
		(1.51)	(0.43)	(0.41)	(0.32)
Mountain Dew	Sugary	4.20	2.31	5.64	1.85
		(0.73)	(0.45)	(0.57)	(0.27)
	Diet	3.55	2.43	5.45	1.89
		(1.39)	(0.46)	(0.45)	(0.30)
Sierra Mist	Sugary	4.09	2.35	5.54	1.78
		(0.82)	(0.53)	(0.97)	(0.34)
	Diet	3.36	2.41	5.33	1.75
		(1.19)	(0.53)	(0.86)	(0.37)
Sprite	Sugary	3.83	2.39	5.59	1.88
		(1.17)	(0.45)	(0.74)	(0.34)
	Diet	3.67	2.54	5.47	1.73
		(1.41)	(0.44)	(0.81)	(0.33)
Private Label	Sugary	2.63	1.59	1.59	1.04
		(0.63)	(0.21)	(0.68)	(0.09)
	Diet	2.15	1.59	0.77	1.11
		(0.56)	(0.21)	(0.05)	(0.12)
Other	Sugary	2.90	2.32	4.21	1.69
		(1.00)	(0.44)	(0.63)	(0.29)
	Diet	2.84	2.43	3.32	1.69
		(0.75)	(0.42)	(0.98)	(0.25)
Mean		3.67	2.29	4.82	1.73

Notes: Table shows average prices of products in cents per ounces. Products are differentiated by brand, packaging and sugar content. Standard deviations are given in parentheses.