Who Pays Sin Taxes? Understanding the Overlapping Burdens of Corrective Taxes*

Christopher Conlon[1], Nirupama Rao[2], and Yinan Wang[3]

[1] *Department of Economics, NYU Stern School of Business and NBER.*
[2] *Business Economics and Public Policy, University of Michigan Ross School of Business.*
[3] *Department of Economics, NYU Stern School of Business.*

July 11, 2022

Abstract

We find that sin-good purchases are highly concentrated, with 10% of households paying more than 80% of taxes on alcohol and cigarettes. Total sin-tax burdens are poorly explained by demographics (including income), but are well explained by eight household clusters defined by purchasing patterns. The two most taxed clusters comprise 8% of households, pay 63% of sin taxes, are older, less educated, and lower income. Taxes on sugary beverages broaden the tax base but add to the burdens of heavily taxed households. Efforts to increase sin taxes should consider the heavy burdens borne by few households.

Keywords: Excise tax, sin tax, tax burden

JEL Codes: L66, H22, H23, H25

*Disclaimer: Our own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein

1. Introduction

"Sin taxes" – or excise taxes on particular goods that society deems harmful – are popular in the United States. Federal, state, and local governments levy taxes on alcohol and tobacco with the dual and sometimes conflicting goals of curbing consumption and raising revenue. For many of these products, taxes represent a large share of the overall price. In New York City, a 1.75L bottle of vodka might sell for as little as $11.99 of which $7.97 is tax; and a $13.00 pack of cigarettes includes $6.86 in taxes.

To forward these goals, taxes on sin goods have grown in recent years. In 2009 the federal excise tax on a pack of cigarettes increased from $0.39 to $1.01. As part of the 2021 reconciliation package, House Democrats proposed doubling that to $2.00 per pack.[1] All but nine states have substantially raised their tobacco taxes in the last two decades, with the median tax on cigarettes more than quadrupling between 2000 and 2021 from $0.34 to $1.78. Meanwhile, tax revenues from alcoholic beverages have grown, due to both rising consumption and state tax rate increases.[2] Over the last decade several localities have also levied new taxes on sugar-sweetened beverages (SSBs), with dozens more considering such taxes. Relative to income taxes, general sales taxes, or excise taxes on gasoline, sin taxes enjoy broad public support across the political spectrum.[3]

One complaint about sin taxes is that they are regressive (Allcott et al., 2019b; Hirono and Smith, 2017; Sanders, 2016). One way to counter the regressivity would be to transfer some of the sin-tax revenue back to households through the income tax code (Hendren,

[1]See House Committee on Ways and Means (2021).

[2]For state sin tax revenues, please see Tax Policy Center (2021a) and Tax Policy Center (2021b). For federal sin tax revenues, see Tax Policy Center (2019).

[3]For example, in 2015 Kansas Governor Sam Brownback proposed raising alcohol and tobacco taxes to help close the state's $648 million budget shortfall (Lowry, 2015).

2020). This is difficult (and less effective) if the sin tax burdens of households with similar incomes vary drastically. It also requires understanding the combined burden across multiple sin taxes. Most studies focus on sin taxes for a single category in isolation, such as alcoholic beverages (Griffith et al., 2019; Conlon and Rao, 2019, 2020; Miravete et al., 2020, 2018), sugar-sweetened beverages (Dubois et al., 2020; Allcott et al., 2019a; Seiler et al., 2021; Bollinger and Sexton, 2018), or cigarettes (Adda and Cornaglia, 2006; Colman and Remler, 2008; Harding et al., 2012; Hansen et al., 2017; Friedson et al., 2021).

In this article, we draw on data describing household purchases of alcoholic beverages, tobacco, and sugar-sweetened beverages and provide new measures of the concentration of total sin-tax burdens. Because a relatively small set of households purchases large shares of multiple sin goods, sin-tax burdens are highly concentrated; the top 10% of sin-tax payers account for 80% of all sin taxes while the majority of households pay little to no sin tax. This extreme heterogeneity of burdens across the full population also holds among members of any income or demographic group, complicating both welfare analysis and measurements of distributional impacts. Relying on representative agent frameworks and single elasticity "sufficient statistics" based on averages–even averages within income or other demographic groups–will likely miss the extreme heterogeneity in sin-goods purchases across households and obscure the stark distributional impacts. Analysis of potential policy reforms, such as the Congressional Budget Office's revenue options report (CBO, 2018), that aim to assess the distributional effects of changing sin tax policy should consider moving beyond average impacts by income group to meaningfully capture the impact of reforms.

Our analysis begins with documenting the high concentration of beer, wine, spirits, and cigarette purchases. Just 10% of households account for more than 80% of alcoholic beverage purchases by volume, while the bottom half of the distribution nearly abstains totally from purchases of beer, wine, or spirits. For cigarettes, 8% of households are responsible for virtually all purchases. We also consider a hypothetical national penny-per-ounce tax on

SSBs, which would be more broadly borne since sugary beverages are purchased by three-quarters of households and the top 10% of purchasers account for only 55% of sales volume.

The burden of sin taxes is further concentrated because households tend to purchase multiple categories of sin goods or none at all. This is particularly true of smokers, who in addition to buying highly-taxed tobacco products also purchase larger quantities of SSBs as well as beer and spirits than the typical household. Heavy purchasers of wine, beer, or spirits also tend to purchase large quantities of the other alcoholic beverage categories. As a result, combined burdens are even more concentrated than sin taxes on individual categories with the top 20% of households to pay more than 90% of sin taxes.

These concentrated sin-tax burdens are not well-explained by demographics like income, education, age, and race, or even state-level tax rates; the correlation between cigarette taxes and income, for example, is only -0.06. We document far more heterogeneity in sin-good purchases within income groups than across them, and the median household at all income levels faces little or no exposure to sin taxes, rendering the overall progressivity or regressivity of sin taxes less meaningful. Though household demographics explain only a tiny fraction of the variation of sin-tax burdens across households, the burdens are well-explained by household purchase patterns (i.e. preferences), which appear to be relatively stable across time.

To account for both the multiple dimensions of dependence, and the extreme concentration in sin-good purchases, we discretize the heterogeneity using k-means clustering, and assign each household to one of eight mutually exclusive clusters. These clusters explain 80% of the overall variation in sin-tax burden, while demographics alone explain less than 4%. Two clusters, which we label *Everything* and *Smokers*, comprise 8% of the population but pay 63% of existing sin taxes, averaging approximately 2% of income. These households are disproportionately from the bottom income quintile, low education, and ages 55 to 64. Demographically, these households bear a striking similarity to those Case and Deaton (2020) describe as most susceptible to "deaths of despair." Because they also purchase more sugary beverages than any other clusters, *Everything* and *Smokers* would also bear a disproportionate share of new taxes on SSBs.

Households in the third most-taxed cluster, which we label *Heavy Drinkers*, on average purchase the equivalent of 11 alcoholic drinks per adult per week and make up 6.7% of the population. They are most likely to come from the highest education and income bins. Most previous studies (Conlon and Rao, 2019; Miravete et al., 2020) suggest that wealthier households are less price sensitive, and respond to price increases by switching to less expensive products rather than away from alcoholic beverages altogether. This suggests corrective taxes may be less effective at discouraging consumption among these households. If negative externalities are convex in alcohol consumption (Griffith et al., 2019; Cnossen, 2007), this group along with the *Everything* group would comprise 9% of the population yet be responsible for almost 60% of alcohol's external damage. At the same time, the effective alcohol tax rate faced by these groups is not particularly high when compared to *Moderate Spirits* purchasers.

Assigning households to clusters also allows us to explain the evolution of the sin tax burden from 2007-2020. Within a cluster, the sin-tax burden is relatively constant over time, even as some states changed statutory tax rates. Between 2007 and 2019 the *Everything* and *Smokers* clusters shrank by more than one-half, while the cluster of households purchasing little to no sin goods grew by nearly 70%. As a result, the share of sin taxes paid by the top 1% of sin-tax payers grew by 40%. These long-run trends reversed in 2020 with an uptick in the population shares of *Everything* and *Heavy Drinkers*. This reversal merits more investigation – it might be the COVID lockdowns led to higher purchases of alcoholic beverages, or it may be that consumption shifted from bars and restaurants. Our data cover only purchases in retail stores.

Our findings suggest that policymakers should carefully consider the distributional implications of raising tobacco, alcohol, or SSB taxes. A narrow set of households bears these taxes; unless policymakers believe that even higher taxes will lead them to smoke and drink substantially less, this small swath will bear much of the additional burden, too. Policy assessment will need to move beyond average impacts on demographic groups to account for differences in preferences and elasticities (and incomes) to accurately capture the welfare effects of different interventions.

2. Data

Our main data source is the Kilts NielsenIQ consumer panelist data for 2018. These data follow 61,384 households, who are compensated by NielsenIQ in exchange for recording all purchases of bar-coded products. This panel is designed by NielsenIQ (after weighting) to broadly represent the demographics of the United States. Whenever aggregating, we use the provided *projection factors*.

Since sin taxes are almost always volumetric, our main focus is the volume of purchases of sin goods (tobacco, beer, wine, and distilled spirits) and SSBs. We also include non-sin household staples, specifically yogurt and toilet tissue as comparisons. When we report consumer demographics, we report them in exhaustive mutually-exclusive bins (mostly) following NielsenIQ's definitions, rather than impute them as continuous values.[4] For example, NielsenIQ reports household income in 16 discrete bins which we group into 13 bins (by consolidating households earning below \$10k in income), and later into five "quintiles" (<\$25k, \$25k-\$45k, \$45k-\$70k, \$70k-\$99k, \$100k+).[5]

[4] We use four levels of household head education: HS or less, some college, college graduates, and postgraduates; four race categories: White, Black, Asian, and Other; 5 bins for the head's age; and indicators for whether the head is Hispanic and a child under 18 lives in the home.

[5] We also eliminate 23 "outlier" households from our analysis as detailed in Appendix C.

Our dataset differs from other datasets in some important ways. The most commonly-used dataset on alcoholic beverage consumption is the National Institute on Alcohol Abuse and Alcoholism (NIAAA)'s NESARC-III survey of 36,309 individuals on alcohol usage between 2012-2013. One advantage of the NielsenIQ data is that purchases are not merely self-reported but verified with receipts.[6] Another difference is that the NielsenIQ data track household-level purchases rather than individual consumption. Thus sin goods purchased but not consumed within the household (as a gift or as part of a large gathering) may be wrongly attributed to the household. Because our primary interest is the *tax burden* of sin goods across households, we are primarily concerned with the distribution of sin good purchases rather than consumption. A larger challenge is that the NielsenIQ dataset does not report sin goods purchased and consumed outside the home. This is unlikely to present a major issue for tobacco products, but means we do not observe alcoholic beverages or SSBs consumed on-premise at bars, restaurants, sporting events, etc.[7] Industry reports suggest on-premise sales of alcoholic beverages accounted for around 23% of beer, 18.5% of wine, and 21.2% of spirits sales in 2018 by total volume (Adams Media Inc., 2019).

Our product category definitions are meant to correspond to those used to calculate taxes on various products.We convert cigarette purchases into the equivalent number of packs, and liquids into liters. We exclude e-cigarettes and nicotine cartridges from our tobacco category because in many states those are either untaxed or taxed differently from cigarettes. To match the NIAAA, we apply a constant alcohol by volume percentage (ABV%) to beer (4.5%) and wine (12.9%) and use the observed ABV for spirits when possible (typically around 40%). When we compute sin tax paid by households, we apply the relevant combined federal and

[6]Naimi et al. (2016) find heavy drinkers in survey data drink similar amounts and are demographically similar to heavy drinkers in our results.

[7]Appendix A compares estimates of overall alcoholic beverage and tobacco consumption in the NielsenIQ data and other data sources.

state rates and assume consumers bear the full economic incidence of the taxes.[8] This is clearly an unrealistic simplification, but if the consumer share of the burden is similar across products and across consumers, our results will be proportional to the correct distributional effects.[9]

Likewise our sugar-sweetened beverage category is meant to mimic the set of products commonly subjected to taxes on SSBs. It includes sugary carbonated beverages (Coke and Pepsi) as well as sports drinks (Gatorade) and sweetened teas and juice drinks (Arizona Iced Tea, Hi-C, etc.), but does not include diet carbonated beverages (Diet Coke) or 100% juice products which are typically exempted. When we consider the tax burden, we apply a hypothetical penny-per-ounce tax meant to mimic existing laws and proposals.[10]

3. Empirical Analysis

3.1. The Concentration of Sin Good Purchases

We begin by documenting the concentration of household sin-good purchases. For each household, we compute the annual total liters purchased (or packs in the case of cigarettes). We then rank each household by its total purchases in each category. Because excise taxes

[8]In several "control states," rather than (or in addition to) excise taxes, the state controls the sale of distilled spirits directly. For these states, we impute the tax rate as the average from other states within the region.

[9]The Urban-Brookings Tax Policy Center employs similar assumptions in its distributional analysis of excise taxes on tobacco and alcohol (Rosenberg, 2015). For a list of tax rates, please see Appendix Table B1.

[10]See Tax Policy Center (2020) for details. Consistent with all enacted SSB taxes, we apply the penny-per-ounce tax equally based on volume, rather than on actual sugar content, which differs greatly across products.

on these items are based primarily on volume rather than expenditure, purchase volume (mostly) corresponds to tax burden.[11] Our goal is to describe the concentration of purchases such as the "80-20 Rule" of the Pareto distribution.

Panel A of Figure 1 plots the CDF of annual household purchases for various categories of sin goods and, for comparison, consumer staples. In Panel B, we zoom in on the purchases of the top decile of households. For household staples, the distribution of purchases is not particularly skewed: the top 10% of households purchase 32% of toilet tissue (Gini= 0.45) and 46% of yogurt (Gini= 0.64) products by volume.[12] For beer, wine, and spirits, we find that the top 10% of households account for about 80% of purchases (by volume) (Gini = 0.85), while the bottom half of households purchase little to no alcoholic beverages. For tobacco, the top 10% of households are responsible for virtually all of the purchases, producing a Gini coefficient of 0.90.[13] The distribution of SSB purchases does not resemble that of other sin goods. Over 75% of households purchase significant amounts of SSBs, and purchases are substantially less concentrated. The top 10% of households account for around 55% of purchases (Gini= 0.61) – more similar to that of yogurt (Gini= 0.64) than to alcohol or

[11]Tax burden will also vary by the state in which the purchases are made, and distilled spirits are taxed at the federal level by alcohol content, though the majority of spirits are bottled at 40% alcohol by volume.

[12]On a "per capita" basis the distribution for these staple goods is even less skewed, because much of the variation is explained by household size.

[13]The most recent CDC data suggest around 14% of adults smoke (Centers for Disease Control and Prevention, 2020). The most recent Tobacco Use Supplement to the CPS finds the share of households to be only 4%. Additionally, researchers have reported substantial assortative matching among couples by smoking status using CPS data (Chiappori et al., 2017), helping to explain the smaller share of smoking households.

cigarettes.

[Figure 1 here.]

These purchase distributions have important consequences. The first is that the majority of existing sin taxes are paid by a very small number of households, while many households don't purchase any sin goods. Panel C of Figure 1 plots the distribution of different sin taxes. The top 20% of households pay roughly 90% of all sin taxes, while more than half of households pay virtually no sin taxes. Following the purchase patterns, taxes on cigarettes are more concentrated than those on alcohol beverages. The second important implication is that taxes on SSBs would be much more broad-based than existing sin taxes on alcoholic beverages or tobacco. As Panel C and D of Figure 1 show, SSB taxes would be much more evenly distributed, with the top 20% of households paying about 60% of the tax (Gini= 0.61).

The next question is whether the same households who pay most of the cigarette taxes also pay much of alcohol or (hypothetical) SSB taxes. The obvious approach of examining correlations of annual purchase totals faces some challenges. First, 68% of households never purchase alcoholic beverages or tobacco, leading to a large number of zeros.[14] Second, Figure 1 indicates that nearly all of the consumption is the tails of the distribution, so measures of dependence that average over the entire distribution may miss the majority of purchases.

To address these concerns, we provide two measures of dependence in Figure 2. The first is the Spearman correlation using the rank of a household's purchases in two categories.[15] The second computes the upper tail dependence parameter for the bivariate copula: the

[14] A large number of zeros also complicates more general approaches, which rely on an inverse CDF transformation including copulas (Ibragimov and Prokhorov, 2017).

[15] Because of the large number of ties at zero, we rank households from least to most consumption, and break ties by choosing the lowest rank. The estimated correlations are

conditional probability that a household in the top decile for one category is also in the top decile for another category.[16] Both measures indicate a significant amount of dependence among purchases of beer, wine, and spirits, suggesting at least some households purchase alcohol in multiple forms. The tail dependence measure also detects overlap among households that purchase cigarettes and those in the top decile for purchases of beer or sugary beverages.

[Figure 2 here.]

A common concern regarding sin taxes is that they are regressive or that they disproportionately affect poorer households. In the right portion of Figure 2, we also compute the correlation between purchases in each category and the midpoint of the corresponding income bin for each household. The correlation between purchases of sin goods and income is extremely weak. Purchases of cigarettes and SSBs appear to be very slightly regressive, $\rho = -0.062$ and $\rho = -0.054$ respectively, while purchases in other categories appear to have a small positive correlation with income (such as wine $\rho = 0.06$). These correlations are so weak that income alone explains less than 1% of the variation in sin taxes across households, and household demographics explain less than 3%.

3.2. Discretizing Heterogeneity

In order to better understand these households at the extremes of the distribution, we discretize the heterogeneity in household purchases for six sin categories (beer, wine, spirits,

robust to these assumptions. We also obtain nearly identical results when we use Kendall's τ as an alternative measure of pairwise dependence.

[16]If there were no dependence at all we would expect this measure to be 0.1, the unconditional probability of falling in the top decile.

total ethanol, cigarettes, and SSBs) using k-means clustering.[17] We express the purchases of each household as a vector $\mathbf{z_i}$ and solve the following k-means clustering problem:

$$\left(\widehat{\boldsymbol{\mu}}(1), \ldots, \widehat{\boldsymbol{\mu}}(K), \widehat{k}_1, \ldots, \widehat{k}_N\right) = \underset{(\boldsymbol{\mu}(1),\ldots,\boldsymbol{\mu}(K),k_1,\ldots,k_N)}{\operatorname{argmin}} \sum_{i=1}^{N} \|\mathbf{z_i} - \boldsymbol{\mu}(k_i)\|^2 \tag{1}$$

Each household i is assigned to a group k_i, and assigned the group mean $\boldsymbol{\mu}(k)$. The idea is to minimize the Euclidean distance from each household's purchase vector to the mean of its assigned group.[18] We then assign each household to one of $K = 8$ clusters. Our cluster assignments explain 80% of the variation in the tax burden across households.

We select $K = 8$ clusters by calculating a battery of 30 indices from `NBClust` and taking the median (Charrad et al., 2014). As is often an issue with k-means, there is little agreement across indices in the "optimal" number of clusters. For example, the "elbow method" and "silhouette method" recommend $K = 2$ clusters, while the Gap statistic recommends $K = 12$, and the CH score recommends $K = 13$. We are using the clusters largely as a way to describe heterogeneity, so in Appendix E, we also report categorizations under $K = 7$ and $K = 9$ clusters.

After inspecting the purchasing patterns of each cluster, we assign it a name for expo-

[17] Recent work by Bonhomme et al. (2022) suggests that even when heterogeneity is not discrete, approximations by k-means can still be effective in a variety of settings.

[18] A well-known limitation of k-means is its sensitivity to transformations of $\mathbf{z_i}$. In order to deal with the skewness in the distribution of purchases and the large number of zeros, we first apply the inverse-hyperbolic-sine transformation: $arcsinh(x) = \ln(x + \sqrt{x^2 + 1})$ and then apply a Z-score to each column. This is similar to the $\log(\dot{})$ transform but maps zeros: $f(0) \to 0$. The $arcsin(x)$ transformation can make regression coefficients difficult to interpret, but these transformed variables never appear in a regression equation.

sitional purposes. This allows us to categorize households by the products they actually purchase, rather than merely demographics or location.

3.3. Results

We describe the results of our cluster assignments in three tables. In Table 1, we report summary statistics for each cluster of households; In Table 2, we report the demographic makeup of each cluster; and in Table 3 we apply our cluster assignments from 2018 to annual household purchases from other years in order to explain the evolution of the sin-tax burden over time.

Table 1 reports purchase distributions (in liters — except for cigarettes, which are measured in packs) and tax burdens by cluster and for the overall sample. The first panel of Table 1 are the centroids (weighted averages) that are used in assigning each household to the nearest centroid. In the second panel we report quantiles of the purchase distribution, which highlight the extent to which the clusters overlap (or do not overlap). The bottom panels describe ethanol consumption details, externalities and tax burden measures.

The two largest groups *SSB Only* (44% of households) and *Nothing* (17.8% of households) purchase negligible amounts of sin goods and are largely exempt from any sin tax burden. The main difference between the groups is that the former purchases approximately 1 liter of sugary beverages per week for each member of their household, and would face a significant portion (54%) of any SSB taxes. The size of the *SSB Only* group is what drives the base-broadening effect of hypothetical taxes on sugar-sweetened beverages, as this group would pay about $39 per year in new taxes under our hypothetical penny-per-ounce tax.

The two most important groups for understanding the burden of sin taxes are: *Everything* (2.5% of population) and *Smokers* (5.5% of population). On average, each week the *Everything* group purchases the equivalent of 2.91 packs of cigarettes and 17 alcoholic drinks (with a large percentage of alcoholic beverages coming in the form of beer and to a lesser extent spirits). Despite representing only 2.5% of the population, they pay 25.7% of all sin taxes or an average of \$512 per year, the highest burden of any group. The *Smokers* purchase similar amounts of cigarettes (around 2.5 packs per week), but negligible amounts of alcoholic beverages. However, because the taxes on tobacco are large relative to taxes on alcohol, they account for 37.3% of the sin tax burden and only 5.5% of the population or \$343 per year. Additionally, both the *Smokers* and *Everything* groups consume the largest amount of sugar-sweetened beverages of any of our clusters, around 1.5 and 1.3 liters per household member each week. This means that any new taxes on sugar-sweetened beverages, while broadening the overall base, would still fall disproportionately on the two most heavily-taxed groups increasing their sin tax burden by an average of \$50-\$60 per year.

The third most important group is the *Heavy Drinkers* (6.7% of population). This group purchases the equivalent of 20 drinks per week, including the largest amount of spirits and second largest amounts of beer and wine of any group. Despite representing only 6.7% of the population they are responsible for 47.5% of alcohol taxes and 20.2% of overall sin taxes (an average of \$159 per year). The remaining three groups compose 23.5% of the population, purchase moderate amounts of *Beer* (9.4%), *Wine* (7%), and *Spirits* (7.1%), and account for a combined 14.5% of the tax burden and 30.3% of ethanol externalities.

For alcoholic beverages, we can try to compare the average tax burden to the level of externalities generated. We draw on the approach from Griffith et al. (2017), which calibrates the external damage using an exponential function in ethanol consumption x_i as $ED(x_i) = \phi_0 \left(e^{\phi_1 z_i} - 1\right)$. We follow their calibration, where ϕ_1 is chosen to match the convexity of the external damage from Cnossen (2007), such that an individual who consumes 20 U.S. drinks generates 20× the external damage of an individual who consumes 4 U.S. drinks.[19]

[19] This calibration yields $\phi_1 = 0.0615$. Griffith et al. (2017) also report upper and lower bound estimates of $\phi_1 \in (0.0435, 0.0695)$, though qualitatively this has little impact on our

This enables us to report the average external damage from ethanol consumption by cluster, and the fraction of overall external damage attributable to each cluster without taking a stand on the overall level of external damage. The calculations imply that *Heavy Drinkers* and *Everything* households generate roughly 9× the external damage (per household) of *Moderate Spirits* and *Moderate Beer* drinkers, and about 3× the external damage of *Mostly Wine* drinkers.

Ideally, households with larger (marginal) external damage would face higher effective tax rates on ethanol. The good news is that *Heavy Drinkers* and *Everything* households pay $8.85 and $8.21 in taxes per liter of ethanol while *Moderate Beer* and *Mostly Wine* households pay $5.84 and $5.06 per liter. Thus the heaviest drinkers pay around 50% more than moderate drinkers, and are responsible for $2-3\times$ the external damage. However, due to higher statutory rates on distilled spirits, the *Moderate Spirits* group pays $15.53 per liter of ethanol, suggesting that either this group is significantly over-taxed or other groups are significantly under-taxed.[20]

[Table 1 here.]

Table 2 describes the demographic makeup of each cluster. For each demographic category (Race, Hispanic Origin, Children, Age, Income, Education) we divide the population into a set of mutually-exclusive bins, and report the overall probability of each demographic bin. For example, 27.4% of the overall sample completed high school or less. We calculate the probability of having completed high school or less for households assigned to the *Everything* cluster (35.6%) and report the ratio in Table 2 as $\frac{35.6}{27.4} = 1.30$. Resampling the

results. As we focus on relative damages rather than the level of external damage, ϕ_0 is irrelevant; we nonetheless plug in the provided value of $\phi_0 = 1.298$.

[20] A more comprehensive analysis would account for elasticities both within and across categories.

population of households many times, we recompute each ratio, and highlight cells where the bootstrapped ratio lies outside of $(0.9, 1.1)$. This gives us an easy way to understand which demographic groups are under- or over-represented within each cluster.[21]

The most important takeaway from Table 2 is that the *Everything* and *Smokers* groups are demographically similar to one another, but quite different from the population at large. They tend to be less educated (high school or less) and lower income (under $25,000) than the overall population. They are also less likely to be Asian or have children at home. The 55 to 64 year old age group is also significantly over-represented among these two most-taxed clusters.

The *Heavy Drinkers* also bear a significant share of the overall sin tax burden, but demographically look quite different. They are 53% more likely than a randomly selected household to earn over $100,000, and much less likely to be from the lowest income groups. They are also unlikely to be black or Asian, and more likely to have a college or post-graduate degree. The only other cluster that contains such a large fraction of high-income, high-education households is *Mostly Wine*, which consumes about a third as much ethanol as *Heavy Drinkers* and pays a lower effective tax rate, since wine is taxed at a lower rate per unit of ethanol relative to spirits.

[Table 2 here.]

In Table 3, we use our cluster centroids from the 2018 data and apply them to annual household purchases each year between 2007 and 2020. Over the 2007 to 2020 period, the average tax burden within each of our assigned clusters (which are based on quantity purchased) has remained relatively stable or has increased slightly as some states have raised

[21] As an alternative, in Appendix D we estimate a multinomial logit model, which gives qualitatively similar results but can be more difficult to interpret depending on choice of baseline cluster or household demographics.

taxes on tobacco and alcoholic beverages, and our clusters explain 80% of the total tax burden across households and time.

Several related trends attributable to declining tobacco use and more widespread moderate alcohol purchasing emerge over the 2007 to 2019 period. Between 2007 and 2019, the *Everything* and *Smokers* clusters shrink from a combined share of 17.2% to 8.1% of the population, largely due to reductions in tobacco consistent with widely-documented declines in smoking behavior. Meanwhile, *Nothing* households rise from 11.2% of the population in 2007 to 19.0% in 2019, leading to even more concentrated sin-tax burdens. We also observe increases in the shares of *Heavy Drinkers* and *Moderate Spirits*, and a decline in the share of *Moderate Beer* between 2007 and 2019 in line with overall industry trends away from beer and towards spirits (Adams Media Inc., 2019).

The onset of the COVID-19 pandemic in 2020 marked a striking increase in number of households purchasing substantial quantities of alcohol. *Heavy Drinkers* grew from 6.6 to 9.1 percent that year, while *Moderate Spirits* and *Moderate Beer* households grew more moderately from a combined 16.6 to 19.0 percent of the population. Most of this growth came at the expense of the *Nothing* and *SSB Only* categories which shrank from 61.8% of households in 2019 to 56.3% in 2020. Without data from subsequent years, we cannot determine whether this increase in “drinking types” is due to a temporary shift from consumption that was taking place at bars and restaurants, or a more permanent shift in consumption patterns.

The change in consumption in 2020 reflects a departure from the long run trends in the concentration of the sin-tax burden. Between 2007 and 2019, the share of households paying less than $10 per year in sin taxes grew from 60.0% to 65.2%, and the share of households paying over $100 per year declined from 15.5% to 10.1%. This overall decline in the purchase of sin goods led to an increase in the concentration of the sin-tax burden, with the share of sin taxes paid by the top 1% of households increasing from 17% to 24.4%. In 2020, these trends reversed with the number of households paying less than $10 per year in sin taxes falling to 60.2%, while the share paying $100 per year or more rose to 12.2%. More widespread alcohol purchasing reduced the concentration of sin-tax burdens in 2020, with the share paid by the top percentile falling from 24.4% to 22.0% and the top 5% paying 55.1% of sin taxes in 2020 rather than the 58.9% they paid in 2019.

[Table 3 here.]

4. Discussion

Taken together, our findings suggest some important patterns for those seeking to better understand the distributional impacts of sin taxes. First, focusing on average impacts of sin taxes is likely to be unhelpful. The purchase of sin goods is extremely concentrated among a small number of households with 10% of households paying 80% of sin taxes (Figure 1). The concentration of the burden may help explain (in part) the popularity of these taxes as not only ways to address externalities, but also as a source of government revenue, as these are taxes that few households expect to pay in significant amounts.

The second takeaway is that saying "sin taxes are regressive" or "sin taxes are progressive" largely misses the point. There is much more variation among households *within* income groups than *across* them in purchases of sin goods. Even among the lowest-income groups, the majority of households pay negligible amounts of sin taxes, and there are heavy smokers and heavy drinkers at all levels of education and income. This means attempting to correct regressivity of sin taxes using transfers within the tax-code might be more difficult than more broadly-based taxes (such as gasoline or carbon taxes). It also means that statements about aggregate consumer welfare impacts, or attempts to construct policies that are a "Pareto improvement" may be misleading, unless they specifically address the impacts on these small groups of heavy consumers.

At the same time, our results suggest that the two most heavily taxed groups, *Every-*

thing and *Smokers*, are much more likely come from certain demographic groups than the overall population. These groups tend to earn incomes below \$25,000 per year, have lower educational attainment, and are more likely to be between ages 55-64 when compared to the overall population. Policy discussions around additional sin taxes should address whether these groups will elastically adjust consumption in response to additional taxes, or whether additional taxes are simply a transfer from these highly-taxed households facing difficult circumstances.

We also identify a cluster of non-smoking *Heavy Drinkers* (6.7% of population), who disproportionately come from the highest income and education groups. If negative externalities (drunk-driving, liver damage, domestic abuse, etc.) are increasing in ethanol consumption, then this group (along with *Everything*) should be the source of the bulk of external damage. Our *Heavy Drinkers* consume ethanol from a variety of sources, while our *Everything* tends to consume primarily beer and spirits. These groups are relevant for understanding "tagging," or increasing sin taxes on products preferred by the highest externality individuals. We show that the groups that consume the most ethanol tend not to face the highest tax rates per unit of ethanol. This is because our heaviest drinkers consume both distilled spirits and beer, the latter of which is taxed at one-third to one-half the amount per unit of ethanol relative to the former.

The most important insight of our analysis is that when researchers model the welfare effects of sin taxes, it is crucial to take into account both the extreme concentration of existing sin taxes, and the sometimes overlapping burdens across multiple sin taxes. Representative agent frameworks and single elasticity "sufficient statistics" to calculate aggregate average welfare are unlikely to accurately capture the extreme heterogeneity in the underlying purchase distribution and will miss the extreme heterogeneity in the distributional impacts of sin tax policy.

Exhibits

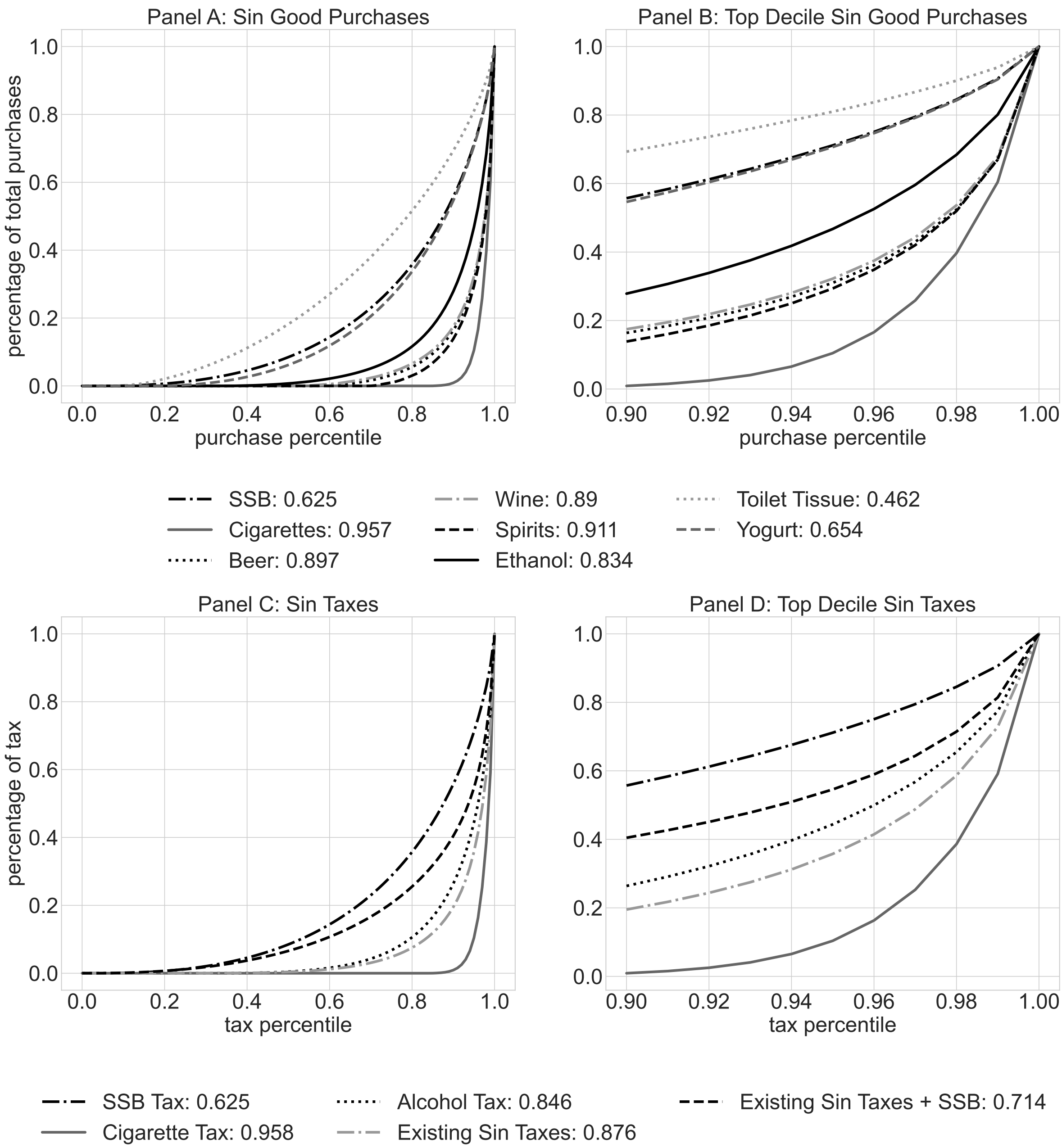

Figure 1: CDFs of Sin Good Purchases and Sin Taxes

Note: Each observation is a household, and households are ranked by annual consumption. The 45 degree line would constitute equal consumption by all households. Legend reports corresponding Gini coefficients.

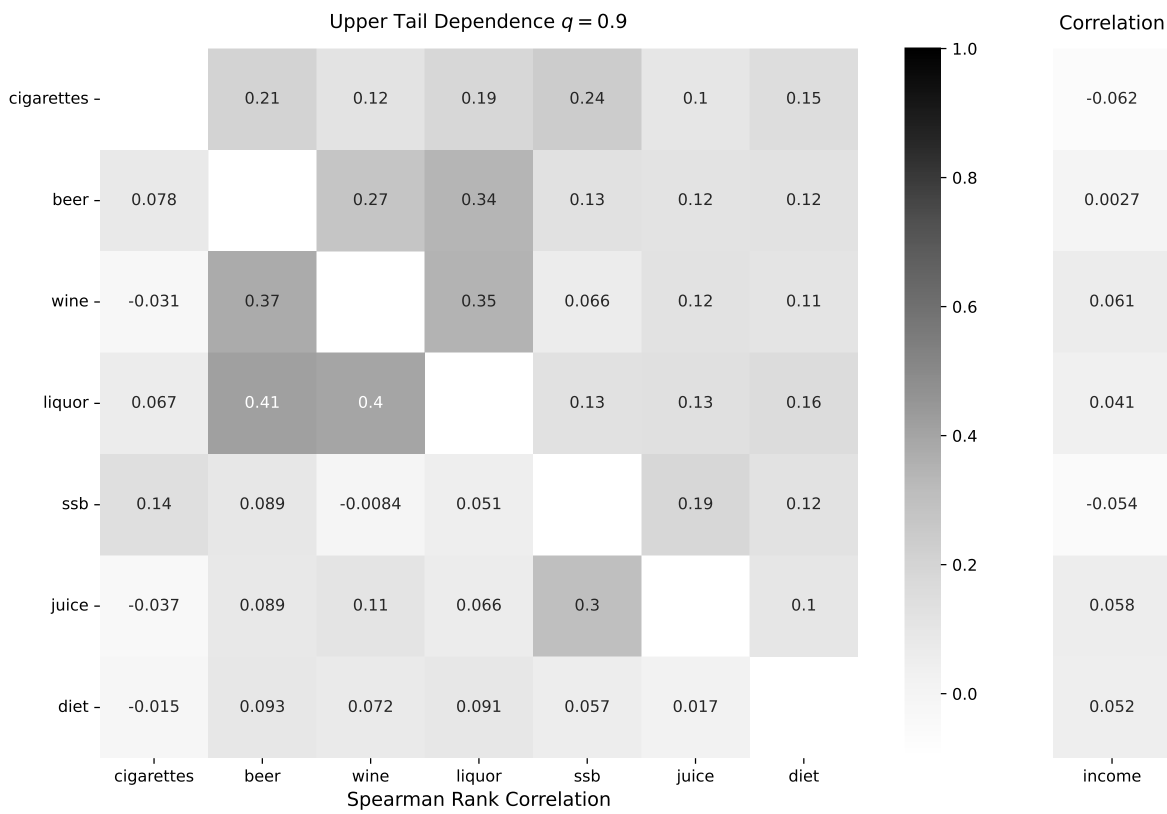

Figure 2: Correlation and Tail Dependence in Purchase Behavior and Income

Note: Each observation is a household and households are ranked by 2018 purchases in each category.
Correlation with income uses midpoint of household's income bin from NielsenIQ Panelist data. Top bin is coded at $150K.
Lower Triangle: Spearman Correlation computes correlation $\text{Corr}(R(x_i), R(y_i))$ among the ranks from smallest to largest, and assigns ties the lowest rank within the set.
Upper Triangle: Upper tail dependence computes $\lambda^U(q) = \Pr\left[Y > F_Y^{-1}(q) \mid X > F_X^{-1}(q)\right] = \frac{\Pr\left[Y < F_Y^{-1}(q), X < F_X^{-1}(q)\right] + 1 - 2q}{1-q}$ for $q = 0.9$ using empirical copula/bivariate ECDF.

	Everything	Smokers	Heavy Drinkers	Moderate Spirits	Mostly Wine	Moderate Beer	SSB only	Nothing
Beer (mean)	150.08	2.28	99.73	7.42	19.17	69.17	0.71	2.43
Wine (mean)	21.03	0.62	37.43	2.56	42.39	1.87	0.79	0.99
Spirits (mean)	19.23	0.42	27.50	9.59	1.37	0.63	0.12	0.32
Tobacco (mean)	151.18	132.18	0.34	0.39	0.41	0.39	0.10	0.21
SSB (mean)	152.09	175.69	92.94	110.44	49.17	114.92	111.19	3.59
Ethanol (mean)	15.66	0.30	18.53	3.33	6.75	3.52	0.17	0.33
Beer 50%	47.54	0.00	40.46	2.88	4.26	25.55	0.00	0.00
Beer 75%	187.38	1.42	105.76	9.23	17.00	61.75	0.00	0.00
Beer 95%	589.20	12.78	414.64	30.90	72.40	281.73	4.26	12.78
Wine 50%	2.25	0.00	15.00	1.50	21.00	0.75	0.00	0.00
Wine 75%	12.00	0.00	37.25	3.79	43.50	3.00	0.75	0.75
Wine 95%	107.74	3.75	153.61	9.50	144.75	7.25	4.50	5.25
Spirits 50%	5.25	0.00	15.76	6.00	0.38	0.00	0.00	0.00
Spirits 75%	17.79	0.00	33.78	10.50	1.88	0.75	0.00	0.00
Spirits 95%	94.17	2.57	91.00	26.97	5.25	3.00	0.90	1.75
Tobacco 50%	80.00	60.00	0.00	0.00	0.00	0.00	0.00	0.00
Tobacco 75%	215.40	179.68	0.00	0.00	0.00	0.00	0.00	0.00
Tobacco 95%	521.17	470.42	2.00	2.33	1.00	2.00	0.05	0.00
SSB 50%	87.90	101.11	49.10	63.99	23.69	69.12	64.31	2.59
SSB 75%	209.95	243.74	122.07	140.58	59.00	148.96	138.40	6.27
SSB 95%	519.74	579.84	330.67	348.41	180.39	380.03	359.13	10.29
SSB (L) per Person/Week	1.35	1.55	0.77	0.91	0.43	0.93	0.99	0.04
Drinks per Week	16.97	0.32	20.08	3.61	7.32	3.82	0.18	0.35
Drinks per Adult	10.12	0.19	11.45	2.25	4.56	2.24	0.10	0.25
Effective Ethanol Tax/L	8.21	4.15	8.85	15.53	5.06	5.84	2.98	3.37
Ethanol Externality	2.18	0.01	2.21	0.24	0.70	0.27	0.01	0.02
Externality Share	18.35	0.26	48.88	5.59	16.31	8.36	1.12	1.13
Total Tax Share	25.73	37.35	20.19	6.05	4.60	3.86	1.17	1.06
Alcohol Tax Share	14.56	0.66	47.48	13.95	10.55	8.68	2.18	1.94
Tobacco Tax Share	33.89	64.17	0.23	0.28	0.25	0.33	0.42	0.42
SSB Tax Share	4.21	10.52	6.77	8.58	3.78	11.83	53.60	0.70
Tax Burden/Income (%)	2.33	1.88	0.39	0.29	0.11	0.18	0.14	0.01
# Households	1394	2808	4362	4247	4972	5609	26413	11527
Share of Households	2.53	5.47	6.65	7.10	7.03	9.41	44.03	17.78

Table 1: Annual Household Purchases by Cluster

Source: NielsenIQ Consumer Panelist Data (2018). All averages and quantiles are projection factor weighted.
Beer, wine, spirits, total ethanol, and SSBs are all measured in liters (per year). Cigarettes are measured in packs of 20.
Reported as percentage of income (averaged over households).
Alcohol external damage calculation from Griffith et al. (2017): $ed_i(z_i) = \phi_0 * (e^{z_i * \phi_1} - 1)$ with $\phi_0 = 1.298$ and $\phi_1 = .0615$ where z_i is liters of ethanol at the household level.

	Everything	Smokers	Heavy Drinkers	Spirits	Wine	Beer	SSB	Nothing
Baseline probability	0.025	0.055	0.067	0.071	0.070	0.094	0.440	0.178
Race: White (74.9%)	1.05	1.10	1.07	0.93	1.06	1.01	0.96	1.04
Race: Black (12.5%)	0.95	0.83	0.74	1.26	0.80	0.83	1.24	0.62
Race: Asian (4.4%)	0.57	0.39	0.68	0.89	0.81	0.72	0.98	1.68
Race: Other (8.2%)	0.87	0.65	0.96	1.34	0.83	1.27	1.05	0.80
Hispanic: No (86.8%)	1.03	1.06	1.01	0.98	1.00	0.93	0.99	1.04
Hispanic: Yes (13.2%)	0.82	0.62	0.95	1.13	1.02	1.47	1.04	0.76
Children: Yes (31.3%)	0.70	0.95	0.82	1.10	0.86	1.20	1.18	0.58
Children: No (68.7%)	1.14	1.02	1.08	0.96	1.06	0.91	0.92	1.19
Age: < 35 (12.9%)	0.62	0.64	0.78	1.09	0.98	1.11	1.03	1.08
Age: 35 to 44 (18.0%)	0.74	0.93	0.85	1.06	0.93	1.14	1.10	0.79
Age: 45 to 54 (21.8%)	1.07	1.15	1.05	1.17	0.89	1.08	1.02	0.80
Age: 55 to 64 (22.7%)	1.45	1.21	1.11	0.97	0.96	0.96	0.94	1.02
Age: > 65 (24.6%)	0.92	0.91	1.08	0.78	1.20	0.81	0.94	1.28
Income: < 24,999 (20.4%)	1.38	1.88	0.42	0.74	0.52	0.75	1.07	1.15
Income: 25,000 - 44,999 (17.7%)	1.20	1.20	0.70	0.89	0.75	0.92	1.09	1.00
Income: 45,000-69,999 (18.2%)	1.00	0.96	0.93	0.98	0.93	1.06	1.03	0.96
Income: 70,000-99,999 (15.5%)	0.92	0.57	1.22	1.12	1.16	1.20	0.97	0.91
Income: > 100,000 (28.1%)	0.64	0.50	1.53	1.20	1.46	1.09	0.89	0.97
Edu: High School or less (27.4%)	1.30	1.64	0.75	0.83	0.66	1.01	1.06	0.89
Edu: Some College (31.4%)	1.31	1.15	1.00	1.11	0.84	1.01	1.02	0.87
Edu: Graduated College (26.3%)	0.70	0.55	1.15	1.07	1.21	1.05	0.96	1.08
Edu: Post College Grad (14.9%)	0.33	0.30	1.19	0.96	1.59	0.88	0.90	1.34

Table 2: Relative Risk by Demographic Group: $\frac{Pr(h \in Demog | h \in Cluster)}{Pr(h \in Demog)}$

A value of 1 indicates that conditional on being in a particular demographic bin (row) households are equally likely to belong to the given cluster (column) as a randomly chosen household.
Light denotes demographic bins with values greater than 1.1 for a 95% (bootstrapped) CI.
Dark denotes demographic bins with values less than 0.9 for a 95% (bootstrapped) CI.

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Fraction of Each Type														
Everything	5.5	5.1	4.9	4.5	4.2	3.7	3.5	3.1	3.0	2.9	2.6	2.5	2.4	2.9
Smokers	11.7	11.1	10.8	9.5	8.6	8.1	8.0	7.5	6.7	6.6	6.1	5.5	5.7	5.8
Heavy Drinkers	5.7	5.9	5.7	5.7	6.1	5.7	5.9	6.1	6.0	6.3	6.2	6.7	6.6	9.1
Moderate Spirits	5.1	5.0	4.9	5.7	5.9	5.6	6.1	6.4	6.5	6.7	6.8	7.1	7.6	8.5
Mostly Wine	6.3	6.1	6.2	6.3	6.3	6.1	6.2	6.5	6.6	6.8	6.8	7.0	6.9	6.9
Moderate Beer	11.6	11.5	11.6	11.4	11.3	11.2	9.9	10.1	10.3	10.7	10.2	9.4	9.0	10.5
SSB only	42.9	43.4	44.0	44.5	44.4	45.6	45.1	44.5	44.5	44.2	44.3	44.0	42.8	39.5
Nothing	11.2	11.8	11.8	12.4	13.1	14.0	15.3	15.9	16.2	15.8	17.0	17.8	19.0	16.8
Average Taxes Per Year														
Everything	500.8	497.1	501.5	505.6	533.8	515.1	517.4	518.4	501.7	504.9	477.3	512.2	488.2	532.9
Smokers	332.3	337.0	329.1	322.5	329.0	326.8	321.7	315.3	313.3	332.4	339.0	343.8	321.8	335.5
Heavy Drinkers	147.0	145.3	152.5	150.4	148.0	151.7	157.7	157.2	152.6	149.6	150.8	152.9	154.2	158.3
Moderate Spirits	43.7	43.1	43.7	42.4	41.1	44.5	45.8	43.5	42.8	43.3	44.3	42.9	45.0	42.4
Mostly Wine	31.6	32.5	32.5	31.6	32.4	31.5	32.3	31.7	31.1	32.0	31.8	33.0	33.3	33.6
Moderate Beer	19.6	19.2	20.6	20.6	20.6	19.8	21.1	21.0	20.3	19.9	20.2	20.7	20.6	21.2
SSB only	1.3	1.2	1.3	1.3	1.3	1.3	1.4	1.3	1.3	1.4	1.3	1.3	1.4	1.4
Nothing	2.9	2.9	3.0	3.1	2.9	3.2	3.2	3.1	3.0	3.2	3.0	3.0	3.1	3.4
Hypothetical SSB Taxes Per Year														
Everything	62.5	64.5	62.6	62.0	64.2	57.2	55.9	55.0	52.8	55.7	54.4	51.4	53.1	58.6
Smokers	69.0	70.2	68.4	66.9	65.7	63.2	62.4	63.7	61.2	61.0	60.2	59.4	59.1	65.9
Heavy Drinkers	38.6	40.2	39.2	36.8	38.2	35.8	36.5	33.4	32.1	32.0	32.2	31.4	30.3	33.2
Moderate Spirits	42.4	42.5	47.3	44.0	41.2	40.6	40.4	38.8	35.0	37.0	36.4	37.3	34.7	38.0
Mostly Wine	21.6	21.4	21.0	20.8	22.1	19.3	18.1	18.3	17.6	18.2	16.4	16.6	16.6	18.0
Moderate Beer	50.3	49.4	49.9	47.5	47.9	46.2	43.1	42.0	39.5	40.6	37.5	38.9	37.4	39.6
SSB only	47.0	46.5	45.2	44.7	44.5	42.1	41.0	40.5	39.9	40.0	38.9	37.6	36.8	40.7
Nothing	1.3	1.3	1.3	1.3	1.3	1.3	1.2	1.2	1.3	1.3	1.3	1.2	1.2	1.2
Share of Taxes Paid by														
Top 1%	17.0	17.7	18.2	18.6	20.4	21.0	21.6	23.1	23.3	24.8	25.6	25.1	24.4	22.0
Top 5%	51.2	51.9	51.8	53.0	55.0	57.3	57.4	58.1	59.3	60.0	60.6	59.7	58.9	55.1
Top 10%	71.3	71.3	71.1	72.6	72.7	74.9	74.7	75.6	76.2	76.3	76.7	76.5	75.1	71.9
Top 15%	81.8	81.8	81.5	82.4	82.4	84.1	83.7	84.3	84.7	84.6	85.0	84.7	83.8	81.0
Fraction of Households By Annual Burden														
Less than $10	60.0	61.0	61.5	62.3	62.7	64.7	64.2	64.6	65.2	64.8	66.0	65.9	65.2	60.2
Between $10-$25	10.6	10.3	10.1	10.6	10.6	10.5	10.6	10.8	10.9	11.0	10.8	10.6	11.1	11.9
Between $25-$100	13.9	13.8	13.6	13.6	13.8	13.0	13.5	13.3	13.2	13.7	13.2	13.4	13.6	15.7
Between $100-$250	6.8	6.6	6.9	6.3	6.1	5.7	5.7	5.7	5.4	5.3	5.1	5.2	5.4	6.5
Greater than $250	8.7	8.4	7.9	7.2	6.7	6.2	6.1	5.6	5.3	5.2	4.9	4.9	4.7	5.7

Table 3: Variation in Heterogeneous Types and Tax Burden Over Time

All calculations use NielsenIQ projection factor as weights.
Types are labeled using the k-means centroids from 2018.
Annual Burden and Share of Taxes Paid, assume 100% of incidence is on consumers and exclude (hypothetical) SSB taxes.

References

Adams Media Inc. (2019) "Liquor Handbook 2019," in *Adams/Jobson's liquor handbook.*, New York, NY :: Adams Media Inc.

Adda, Jérôme and Francesca Cornaglia (2006) "Taxes, Cigarette Consumption, and Smoking Intensity," *American Economic Review*, 96 (4), 1013–1028, 10.1257/aer.96.4.1013.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky (2019a) "Regressive Sin Taxes, with an Application to the Optimal Soda Tax," *The Quarterly Journal of Economics*, 134 (3), 1557–1626, 10.1093/qje/qjz017.

Allcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky (2019b) "Should We Tax Sugar-Sweetened Beverages? An Overview of Theory and Evidence," *Journal of Economic Perspectives*, 33 (3), 202–27, 10.1257/jep.33.3.202.

Bollinger, Bryan and Steven Sexton (2018) "Local Excise Taxes, Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax,"Technical report, Working Paper.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2022) "Discretizing Unobserved Heterogeneity," *Econometrica*, 90 (2), 625–643.

Case, Anne and Angus Deaton (2020) *Deaths of Despair and the Future of Capitalism*: Princeton University Press, `https://deathsofdespair.princeton.edu/`.

CBO (2018) "Options for Reducing the Deficit: 2019 to 2028," CBO Publication 54667, `www.cbo.gov/publication/54667`.

Centers for Disease Control and Prevention (2020) "Current Cigarette Smoking Among Adults in the United States," `https://www.taxpolicycenter.org/briefing-book/how-do-state-and-local-soda-taxes-work`.

Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs (2014) "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set," *Journal of Statistical Software*, 61 (6), 1–36, 10.18637/jss.v061.i06.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque (2017) "Bidimensional Matching with Heterogeneous Preferences: Education and Smoking in the Marriage Market," *Journal of the European Economic Association*, 16 (1), 161–198, 10.1093/jeea/jvx012.

Cnossen, Sijbren (2007) "Alcohol taxation and regulation in the European Union," *International Tax and Public Finance*, 14 (6), 699–732, `https://EconPapers.repec.org/RePEc:kap:itaxpf:v:14:y:2007:i:6:p:699-732`.

Colman, Gregory J. and Dahlia K. Remler (2008) "Vertical Equity Consequences of Very High Cigarette Tax Increases: If the Poor Are the Ones Smoking, How Could Cigarette Tax Increases Be Progressive?" *Journal of Policy Analysis and Management*, 27 (2), 376–400, http://www.jstor.org/stable/30162849.

Conlon, Christopher T. and Nirupama L. Rao (2019) "The Price of Liquor is Too Damn High," Working Paper.

——— (2020) "Discrete Prices and the Incidence and Efficiency of Excise Taxes," *American Economic Journal: Economic Policy*, 12 (4), 111–43, 10.1257/pol.20160391.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell (2020) "How Well Targeted Are Soda Taxes?" *American Economic Review*, 110 (11), 3661–3704, 10.1257/aer.20171898.

Friedson, Andrew I, Moyan Li, Katherine Meckel, Daniel I Rees, and Daniel W Sacks (2021) "Cigarette Taxes, Smoking, and Health in the Long-Run," Working Paper 29145, National Bureau of Economic Research, 10.3386/w29145.

Griffith, Rachel, Martin O'Connell, and Kate Smith (2017) "Design of optimal corrective taxes in the alcohol market," IFS Working Paper W17/02, Institute for Fiscal Studies, https://ifs.org.uk/uploads/publications/wps/WP201702.pdf.

——— (2019) "Tax design in the alcohol market," *Journal of Public Economics*, 172, 20 – 35, https://doi.org/10.1016/j.jpubeco.2018.12.005.

Hansen, Benjamin, Joseph J. Sabia, and Daniel I. Rees (2017) "Have Cigarette Taxes Lost Their Bite? New Estimates of the Relationship between Cigarette Taxes and Youth Smoking," *American Journal of Health Economics*, 3 (1), 60–75, 10.1162/AJHE_a_00067.

Harding, Matthew, Ephraim Leibtag, and Michael F. Lovenheim (2012) "The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data," *American Economic Journal: Economic Policy*, 4 (4), 169–98, 10.1257/pol.4.4.169.

Hendren, Nathaniel (2020) "Measuring economic efficiency using inverse-optimum weights," *Journal of Public Economics*, 187, 104198, https://doi.org/10.1016/j.jpubeco.2020.104198.

Hirono, Katie and Katherine Smith (2017) "Tobacco tax hikes are great, so long as you're not a poor smoker," https://theconversation.com/tobacco-tax-hikes-are-great-so-long-as-youre-not-a-poor-smoker-75211.

House Committee on Ways and Means (2021) "Subtitle I – Responsibly Funding Our Priorities," https://waysandmeans.house.gov/sites/democrats.waysandmeans.house.gov/files/documents/SubtitleISxS.pdf.

Online Appendix for: Who Pays Sin Taxes? (Conlon, Rao, Wang)

A. Comparison of Nielsen Panelists to Government Surveys

A.1. Alcoholic Beverages

An important question is whether the NielsenIQ Consumer Panelist purchase data capture similar patterns to government survey data on alcohol and tobacco consumption. In Table A1 we compare the deciles of weekly alcohol purchases according to our Nielsen Consumer Panelist data and consumption data from the NIAAA's National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). The NESARC deciles are based on the widely-cited tabulations of Cook (2007) with a key adjustment. Because aggregate alcohol consumption in the NESARC survey is roughly half of apparent consumption as measured by shipments reported by manufacturers and tracked by the Alcohol and Tobacco Tax and Trade Bureau of the U.S. Department of the Treasury (TTB), Cook (2007) inflates NESARC survey consumption by roughly a factor of two. We adjust these tabulations, deflating them so that we can compare direct survey data from these two sources. Specifically, we deflate the deciles reported in Cook (2007) by a factor of 1.97. These adjusted deciles are reported in Table A1 under NESARC.

NESARC surveys individuals and the data describes weekly drinks at the individual level. Nielsen on the other hand is panel of households. We provide two measures of weekly purchases from the Nielsen data: drinks per adult in the household and per household. The adjusted NESARC deciles are quite similar to the deciles of the NielsenIQ data. Even at the 80th and 90th percentiles the NESARC data falls between the household level and per adult level measures using the NielsenIQ data. The two main rationales for these descrepancies are: (a) the NESARC survey includes "on premise" consumption (bars and restaurants) while our data look at purchases for "off premise" consumption only; (b) we don't know who within a household consumes the drinks. This means that our per household calculations lie somewhat above the NESARC data, and our per adult calculations lie somewhat below (as this assumes equal consumption within the household). We break out purchases per adult (rather than per household) as measured in "standard drinks" per week in Figure A1. The figure reports the quantiles of "drinks per adult per week". This suggests that alcohol purchase (by ethanol units) are increasing in income, but still dominated by a small number of very heavy drinkers.

While Table A1 compares weekly purchases from the Nielsen panelist data to weekly consumption described by the NESARC survey, Table A2 compares annual average alcoholic beverage volume and liters of ethanol per household by beverage category in the Nielsen data to NIAAA data on apparent consumption, that is alcohol sales, which come from TTB data and shipments reported by manufacturers. It is well-known that aggregate consumption totals from survey responses do not match the quantity of alcohol sold in the U.S.; survey responses generally account for only half of the alcohol sold (Cook, 2007). We sum total annual beer, wine and spirits consumption in the NIAAA apparent consumption data and

	NESARC	Nielsen per adult	Nielsen Households
10%	0	0	0
20%	0	0	0
30%	0	0	0
40%	0.01	0.05	0.10
50%	0.07	0.16	0.31
60%	0.32	0.36	0.69
70%	1.10	0.76	1.42
80%	3.17	1.67	3.11
90%	7.76	4.62	8.47
max	37.49	78.68	154.56

Table A1: Alcoholic Drinks Per Week

Note: The table above reports the average number of drinks per adult aged 18 years or older per week and the average number of drinks per household per week. The number of drinks is calculated according to `https://pubs.niaaa.nih.gov/publications/practitioner/PocketGuide/pocket_guide2.htm` where a standard drink is any drink that contains about 14 grams of pure alcohol (about 0.6 fluid ounces or 1.2 tablespoons). The first column, NIAAA, is based on `https://www.washingtonpost.com/news/wonk/wp/2014/09/25/think-you-drink-a-lot-this-chart-will-tell-you/` but we divided numbers by 1.97 to recover original data. The second and third column are averages from the Nielsen data at the individual adult and household level where we use the same formula to determine standard drinks.

	Nielsen Volume	Nielsen Ethanol	NIAAA Volume	NIAAA Ethanol	Ethanol Discrepancy(%)	On-Premise (%)
Beer	19.68	0.89	188.09	8.46	89.53	23.00
Wine	6.92	0.89	26.98	3.48	74.35	18.50
Spirits	3.29	1.35	16.95	6.97	80.63	21.20
Total	29.89	3.13	232.02	18.91	83.45	

Table A2: Volume and Ethanol Consumption per Household, Nielsen versus NIAAA Data

Note: All units are in liters. To convert volume liters to ethanol liters we use ABV values of 0.045 for beer, 0.129 for wine, and 0.411 for spirits. On-premise shares reported in the final column are from Adams Media Inc. (2019). NIAAA data are from https://pubs.niaaa.nih.gov/publications/surveillance115/CONS18.htm.19 and Census data can be found at `https://fred.stlouisfed.org/series/TTLHH25`.

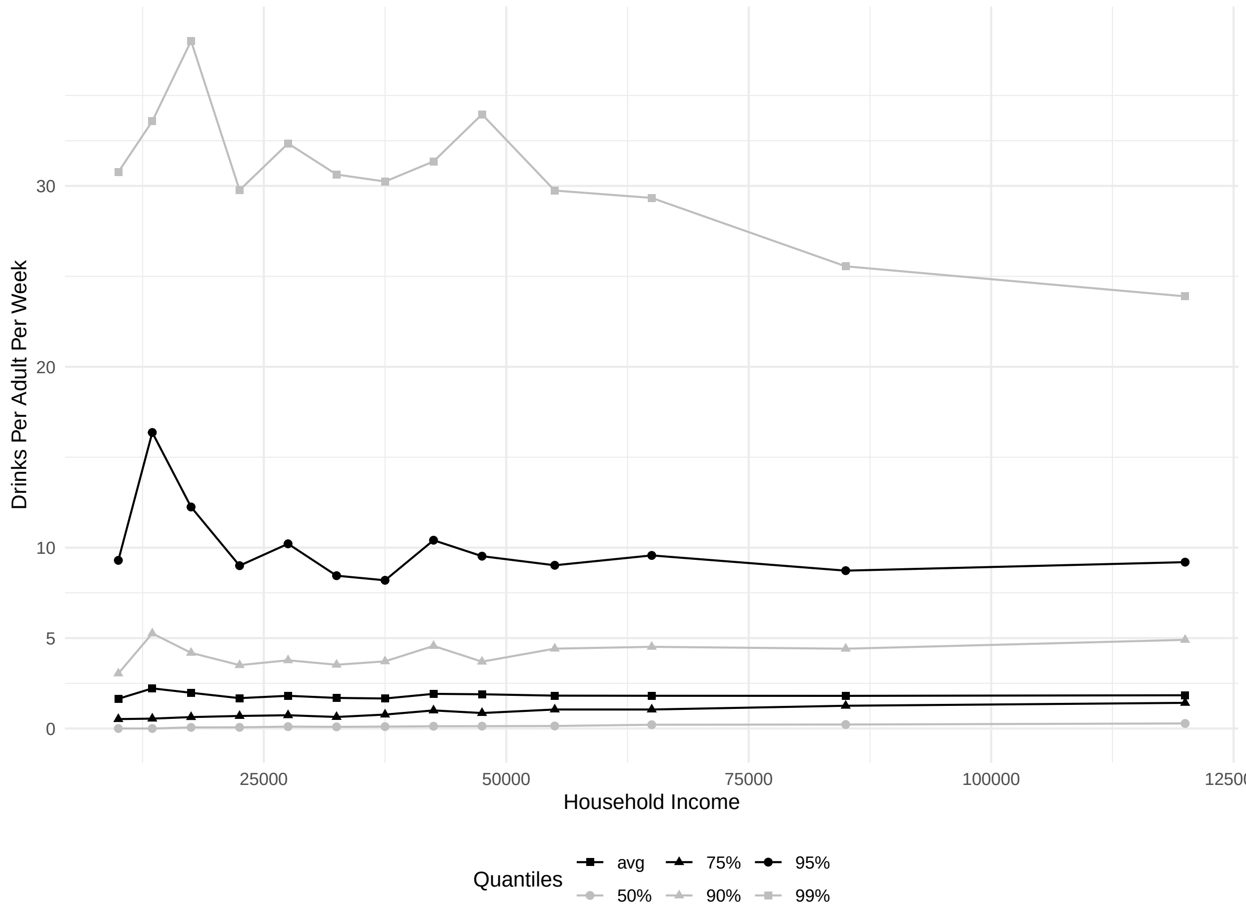

Figure A1: Alcoholic Drinks Per Adult Per Week

Source: NielsenIQ Panelist Data and authors' calculations.
These are computed on a per-household basis and then divided (equally) by household size. We convert liters of ethanol to "standard drinks" (17.7mL of ethanol).

divide by the number of U.S. households in 2018 according to the U.S. Census[1]. As NIAAA data are reported in gallons, we convert these values into liters. Similarly, we sum total annual purchases of beer, wine and spirits in the Nielsen data and scale by the number of households. We convert these volumes into ethanol liters per household using standardized alcohol by volume measures (ABV) consistent with the NIAAA data: 0.045 for beer, 0.129 for wine and 0.411 for spirits.

As we would expect, average household alcoholic beverage consumption from the NIAAA data exceeds our tabulations of average household purchases from NielsenIQ in terms of both volume and ethanol. This is especially true for beer, where average ethanol purchased per household according to the NiselsenIQ data is 89.48% lower than apparent consumption

[1]We use 127,586,000 households in all of our calculations. `https://fred.stlouisfed.org/series/TTLHH`

reported in the NIAAA data. This discrepancy is in part explained by the inclusion of on-premise consumption in the NIAAA data. Nielsen data tracks retail purchases and excludes on-premise purchases. As the last column of Table A2 shows, however, industry reports suggest that on-premise sales account for less than a quarter of alcoholic beverage sales (by volume) in any category.

A.2. Cigarettes

Finally, we compare cigarette purchases as recorded in the Nielsen data to consumption data from the Current Population Survey Tobacco Use Supplement (TUS). The TUS surveys individuals rather than households as Nielsen does. Of the 142,577 records in the TUS, 108 are invalid responses and 137,964 or 96% report not smoking at all. In the Nielsen data, 88% of households make zero annual cigarette purchases. We restrict our attention to the top few percentiles due to the small number of purchasers in both datasets. Table A3 reports individual daily cigarette consumption from the TUS data, as well as average cigarette purchases per day from Nielsen at the per adult and per household level.

In part because on-premise sales are less common for cigarettes, purchases and consumption track each other closely. Because we can only measure household purchases in the Nielsen data and not all adults in a household may smoke, our cigarettes per adult per day measure understates true daily consumption by smoking adults. As such the TUS averages generally lie between the Nielsen per adult and per household averages.

	TUS	Nielsen (Per Adult)	Nielsen (Per Household)
95%	0	0.767	1.425
96%	0	1.458	2.740
97%	4	2.822	5.205
98%	10	5.317	9.260
99%	20	10.339	16.969
max	40	52.0	197.26

Table A3: Cigarette Consumption Per Capita, Nielsen versus CPS

Note: This table compares average daily cigarette consumption according to the Current Population Survey Tobacco Use Supplement (TUS) `https://cancercontrol.cancer.gov/brp/tcrb/tus-cps/questionnaires-data#2018` to average daily Nielsen purchases per adult and household. The TUS surveys individuals while Nielsen tracks the purchases of households. The table presents the number of cigarettes smoked per day where one pack of cigarettes contains 20 cigarettes. We report only the top few percentiles of consumption because the vast majority of respondents in both dataset do not purchase cigarettes. The 2018 contains 142,577 records of which 108 respondents do not provide any valid answer and 137,964 people (about 96%) report not smoking at all. In the Nielsen data 88% households made zero (annual) cigarette purchases.

B. Tax Information

B.1. Sin Tax Rates

In the United States, alcoholic beverages and tobacco are taxed by the federal government as well as most states. Different rates typically apply to beer, wine and spirits. The difference in rates often means that spirits are taxed at a higher rate per unit of ethanol, relative to wine and beer. For example, the federal government's tax rates for beer wine and spirits of \$0.15, \$0.28 and \$3.57 per liter translate into tax rates of \$3.33, \$2.17 and \$8.93 per liter of ethanol, respectively. Table B1 lists the federal and state tax rates for beer, wine and spirits in 2018 in terms of dollars per liter in the left panel and per liter of ethanol in the next panel. This list includes control states that often do not levy a specific sin tax on alcohol but charge a markup on alcoholic beverages which are exclusively sold in state-run stores. There is substantially heterogeneity in alcohol tax rates across states and also within states in the rates they apply to different beverages categories. For example, tax rates on beer range from \$0.01 to \$0.34 per liter. As on the federal level, state taxes per unit of ethanol are very different across categories even within a state. In most cases tax per unit ethanol is highest for spirits and lowest for beer, but there are exceptions. In Tennessee, for example, beer bears the highest state tax at \$7.57 per ethanol liter while spirits enjoy a relatively low tax \$2.91 per liter. Perhaps unsurprisingly, Tennessee is known for whiskey production.

Cigarettes are also subject to federal taxes as well as additional state taxes in all states. These taxes range from \$0.17 per pack (of 20 cigarettes) in Missouri to \$4.35 per pack in Connecticut and New York.

The final panel of Table B1 reports average sin taxes per household in each state. These averages reflect both state consumption patterns and state tax rates on alcohol and tobacco. Average sin taxes per household range considerably across states, particularly because some control states do not impose explicit sin taxes but instead apply substantial markups at state-run monopoly stores.

It is worth noting that alcohol excise tax rates in the United Kingdom, which have the been the subject of a number of recent studies, are much higher than in the United States. We provide a comparison of tax rates in the UK to those in New York and California in Table B2. The tax rates on beer are are roughly comparable, but rates on wine are around $10\times$ as large in the UK and rates on distilled spirits are around $3-4\times$ as large in the UK. Tobacco taxes in the UK are around \$2.00 per pack higher than in New York.

B.2. SSB Tax Nielsen Modules

We have to determine which products would be subjected to a potential tax on sugar-sweetened beverages. As far as we can tell, all of the implemented SSB taxes apply to total volume (rather than sugar content). Thus a 20oz bottle of moderatly sweetened ice tea is taxed at the same rate as 20oz bottle of full-sugar soda. Jurisdictions differ in which products they exempt. For example Philadelphia and Washington, DC both include diet soda in the tax, whereas other cities (Berkeley, Oakland, San Francisco, and Seattle) do not.[2]

We try as best we can to match the Berkeley, CA or Seattle, WA definitions. We include non-diet soda, sports and energy drinks, and all "juice drinks" that are not 100% juice. This means we treat the following `product_module_code`'s as being subjected to our hypothetical SSB tax: 1030, 1041,1042, and 1484, but exclude any brand name that includes the terms

[2]See `https://www.taxpolicycenter.org/briefing-book/how-do-state-and-local-soda-taxes-work` for more details.

‘diet’ or ‘zero’. Such excluded brand code include 541289, 541308, 688343, 620855, 620862, 754017, 721725, 754017, and 688073. Untaxed SSB modules are 1553 and any module in product group 507 except for 1030, 1041,1042, and 1484 unless the brand name includes ‘diet’ or ‘zero’.

State	Tax Rate (per L) Beer	Wine	Spirits	Cigarette	Tax Rate (per Ethanol L) Beer	Wine	Spirits	Tax/HH
FED	0.15	0.28	2.85	1.01	3.40	2.19	7.13	61.95
AL	0.28	0.45	0.72*	0.68	6.16	3.48	1.80*	63.91
AZ	0.04	0.22	0.79	2.00	0.94	1.72	1.98	75.13
AR	0.06	0.20	0.66	1.15	1.41	1.54	1.65	72.42
CA	0.05	0.05	0.87	2.87	1.17	0.41	2.18	50.24
CO	0.02	0.08	0.60	0.84	0.47	0.65	1.51	56.97
CT	0.06	0.19	1.43	4.35	1.41	1.47	3.57	85.23
DE	0.04	0.43	1.19	2.10	0.94	3.34	2.97	81.63
FL	0.13	0.59	1.72	1.34	2.82	4.61	4.29	78.33
GA	0.27	0.62	1.22	0.37	5.93	4.81	3.05	56.16
ID	0.04	0.12	0.95*	0.57	0.88	0.92	2.38*	55.59
IL	0.06	0.37	2.26	1.98	1.36	2.85	5.65	72.16
IN	0.03	0.12	0.71	1.00	0.68	0.96	1.77	60.94
IA	0.05	0.46	0.89*	1.36	1.12	3.58	2.22*	69.52
KS	0.05	0.08	0.66	1.29	1.06	0.61	1.65	38.97
KY	0.02	0.13	0.51	1.10	0.47	1.02	1.27	55.71
LA	0.11	0.20	0.80	1.08	2.37	1.55	2.00	56.21
ME	0.09	0.16	1.43*	2.00	2.05	1.23	3.57*	82.48
MD	0.02	0.11	0.40	2.00	0.53	0.82	0.99	58.93
MA	0.03	0.15	1.07	3.51	0.62	1.13	2.67	61.56
MI	0.05	0.13	0.89*	2.00	1.17	1.04	2.22*	59.54
MN	0.04	0.08	1.33	3.04	0.86	0.61	3.32	55.68
MS	0.11	0.09	0.72*	0.68	2.51	0.72	1.80*	47.30
MO	0.02	0.11	0.53	0.17	0.35	0.86	1.32	54.62
MT	0.04	0.28	0.95*	1.70	0.80	2.17	2.38*	70.06
NE	0.08	0.25	0.99	0.64	1.82	1.95	2.48	65.37
NV	0.04	0.18	0.95	1.80	0.94	1.43	2.38	80.12
NH	0.08	0.16*	1.43*	1.78	1.76	1.23*	3.57*	92.57
NJ	0.03	0.23	1.45	2.70	0.70	1.79	3.63	51.36
NM	0.11	0.45	1.60	1.66	2.41	3.48	4.00	93.62
NY	0.04	0.08	1.70	4.35	0.82	0.61	4.25	54.96
NC	0.16	0.26	0.72*	0.45	3.62	2.04	1.80*	56.04
ND	0.04	0.13	0.66	0.44	0.94	1.02	1.65	53.68
OH	0.05	0.08	0.89*	1.60	1.06	0.66	2.22*	62.77
OK	0.10	0.19	1.47	2.03	2.33	1.47	3.67	71.37
OR	0.02	0.18	0.95*	1.33	0.48	1.37	2.38*	55.21
PA	0.02	0.16*	1.43*	2.60	0.46	1.23*	3.57*	72.06
RI	0.03	0.37	1.43	4.25	0.62	2.87	3.57	66.10
SC	0.20	0.29	0.72	0.57	4.51	2.21	1.80	63.01
SD	0.07	0.25	1.04	1.53	1.58	1.90	2.60	61.41
TN	0.34	0.32	1.16	0.62	7.57	2.48	2.91	60.63
TX	0.05	0.05	0.63	1.41	1.12	0.42	1.59	57.71
UT	0.11	0.22*	0.95*	1.70	2.39	1.72*	2.38*	31.40
VT	0.07	0.15	1.43*	3.08	1.56	1.13	3.57*	94.37
VA	0.07	0.40	0.72*	0.30	1.51	3.09	1.80*	48.71
WA	0.07	0.23	3.77	3.02	1.53	1.78	9.42	76.22
WV	0.05	0.26	0.72*	1.20	1.04	2.05	1.80*	65.64
WI	0.02	0.07	0.89	2.52	0.38	0.51	2.22	83.29
WY	0.01	0.22*	0.95*	0.60	0.12	1.72*	2.38*	51.29
DC	0.02	0.08	0.40	2.50	0.53	0.61	0.99	66.59

Table B1: Federal and State Tax Rates on Sin Goods

Source: Tax Policy Center. cigarettes:
`https://www.taxpolicycenter.org/statistics/state-cigarette-tax-rates`,alcohol:`https://www.taxpolicycenter.org/statistics/state-alcohol-excise-tax-rates`.
The table reports federal and state alcohol and cigarette taxes. Alcohol taxes are generally applied in terms of dollars per gallon; we convert these figures into dollars per liter terms. Cigarette taxes are reported in terms of dollars per 20-cigarette pack. Tax rates are from Tax Policy Center. For control states (where the alcohol tax rate is "n.a."), we used the regional median where the four regions are based on Census definitions

Jurisdiction	Beer	Wine	Spirits	Cigarette
CA+Federal	0.20	0.33	3.72	2.87
NY+Federal	0.19	0.36	4.55	4.35
UK	0.23	3.60	13.96	6.39

Table B2: Comparison of US (combined state and local) and UK Sin Taxes

Source: `https://www.gov.uk/tax-on-shopping/alcohol-tobacco` and authors' calculations. Tobacco in the UK also includes a 16.5% *ad valorem* tax (not included).

C. Sample Selection and Outliers

In our analysis we drop a total of 23 households because their reported per capita (per adult) purchases are higher than seems plausible for an adult to regularly consume. We drop households whose per adult purchases exceed either 10 standard drinks or 3 pack of cigarettes per day (over the course of an entire year). One standard drink contains 20mL of ethanol, meaning that 10 drinks per day totals 73L of ethanol per year. Of the 61,384 households in the sample, per capita purchases exceed these thresholds in 23 cases. These observations are excluded from the analysis. Summary stats for those households are provided in Appendix C.

Panel A: Alcohol Outliers - 20

	Beer	Spirits	Wine	Ethanol	Cigarette	Household Size	Adult	SSB	Income
Mean	1394.58	93.42	187.78	121.52	88.88	1.20	1.20	73.58	34849.50
Std	2466.86	121.36	311.52	94.79	167.22	0.41	0.41	90.10	20211.45
Min	0.00	0.00	0.00	74.74	0.00	1.00	1.00	0.00	2499.50
25%	6.39	0.00	0.56	79.39	0.00	1.00	1.00	12.04	24999.50
50%	166.51	20.25	26.63	95.77	0.00	1.00	1.00	38.31	32499.50
75%	1812.94	190.81	193.56	114.00	62.90	1.00	1.00	107.01	42499.50
Max	10830.81	407.15	925.00	505.94	461.00	2.00	2.00	330.69	84999.50

Panel B: Tobacco Outliers - 3

	Beer	Spirits	Wine	Ethanol	Cigarette	Household Size	Adult	SSB	Income
Mean	0.71	1.67	0.96	0.42	1275.00	1.00	1.00	90.58	34499.50
Std	1.23	2.89	1.66	0.39	47.70	0.00	0.00	76.83	27672.19
Min	0.00	0.00	0.00	0.00	1245.00	1.00	1.00	3.55	10999.50
25%	0.00	0.00	0.00	0.23	1247.50	1.00	1.00	61.39	19249.50
50%	0.00	0.00	0.00	0.47	1250.00	1.00	1.00	119.24	27499.50
75%	1.06	2.50	1.44	0.62	1290.00	1.00	1.00	134.10	46249.50
Max	2.13	5.00	2.88	0.78	1330.00	1.00	1.00	148.96	64999.50

Table C1: Distribution of Outliers

Note: The table above describes the 19 households that we drop from our analysis due to their implausibly high per capita purchases. We define an outlier as households whose per capita consumption exceeds 10 standard drinks or 3 packs cigarettes per day. On average, one drink contains 20 ml ethanol, and 10 drinks per day sums to 73L per year. Only 23 households of the 61,384 households in our sample exceed these purchase thresholds and are removed from the main sample.

D. Details for Calculations in Paper

This section contains alternative version of tables and figures in the main text, as well as calculations referenced in the text.

In Table D1, we calculate the average tax paid for different categories of sin goods by household income. We also report the ratio of average sin tax paid by the top income bin ($> \$100K$) to the average sin tax paid in the bottom income bin ($< \$25K$). We see that beer taxes are pretty evenly distributed across income groups. The highest income group purchase fewer SSBs and would only pay 77% as much as most other income groups (including the very poorest). This is driven mostly by the much higher purchases of diet sodas by the highest income groups. Taxes on cigarettes are very regressive with the poorest households paying almost 3× as much as the richest. However, taxes on wine and distilled spirits appear to be strongly progressive with the richest households paying $1.87-2\times$ as much in tax as the poorest. It is important to note that this is not driven by more expensive wine and spirits purchases, as the taxes apply to volume, not revenue.

Income Bins	Beer Tax	Wine Tax	Spirits Tax	Cigarette Tax	SSB Tax	Existing Sin Taxes
*Ratio**	1.03	2.40	1.87	0.38	0.79	0.65
<24,999	4.94	2.07	8.99	60.33	37.91	76.32
25,000 - 44,999	5.27	2.53	11.13	48.31	40.01	67.23
45,000-69,999	5.50	3.12	12.59	38.75	38.54	59.96
70,000-99,999	5.58	3.81	15.08	30.43	35.52	54.90
> 100,000	5.11	4.97	16.85	22.83	30.01	49.77

Table D1: Average Tax Burden by Category and Income Level

Source: NielsenIQ Panelist data and authors' calculations.
All units are dollars per household per year.
*Ratio** divides the tax burden for households whose income >$100k by the burden for those < $25k.

We want to be careful about interpreting Table D1 as the definitive information regarding the progressivity or regressivity, which is why it is not the focus of our analysis but provided here for comparison. In Figure D1 we show why average taxes paid by income are not necessarily an ideal comparison. We plot the average annual total sin taxes paid by households against the 13 income levels provided by Nielsen.[3] If we compare the average sin taxes paid for each income level, sin taxes look highly regressive and the correlation coefficient is $\rho = -0.765$. However, if we plot the log of sin taxes paid (plus one dollar) we find that the correlation is strongly positive $\rho = 0.857$.[4] The problem is the extreme heteroskedasticity where the standard deviation of sin taxes paid is more than $210 for lower income households and only $152 for the highest income households, this and the extreme skewness of the distribution explain the discrepancy. The $\log(x)$ transform implies that going from $100 (90th percentile) to $1000 (99th percentile) in sin tax spending is a change of 2.3 log points, which is the same as the median household going from one dollar of sin tax spending to $10 (around 60th percentile).

Table D2 reports key percentiles of the tax burden distribution by race and ethnicity for existing sin taxes as well existing sin taxes and a penny-per-ounce tax on SSBs. As panels A and C show, the burden of existing sin taxes falls more heavily on white households and non-Hispanic households. Taxing SSBs increases taxes on all groups, but also alters the relative tax burdens of racial and ethnic groups. When SSBs are taxed Black households at or below

[3]We consolidate all income bins below $12,000, so that we have 13 bins instead of 16. These only constitute 6% of the population.

[4]The $\log(x+1)$ vs. $\log(x)$ transformation is not driving the result. Even after dropping the zeros, or using $arcsinh(x)$ the result is similar.

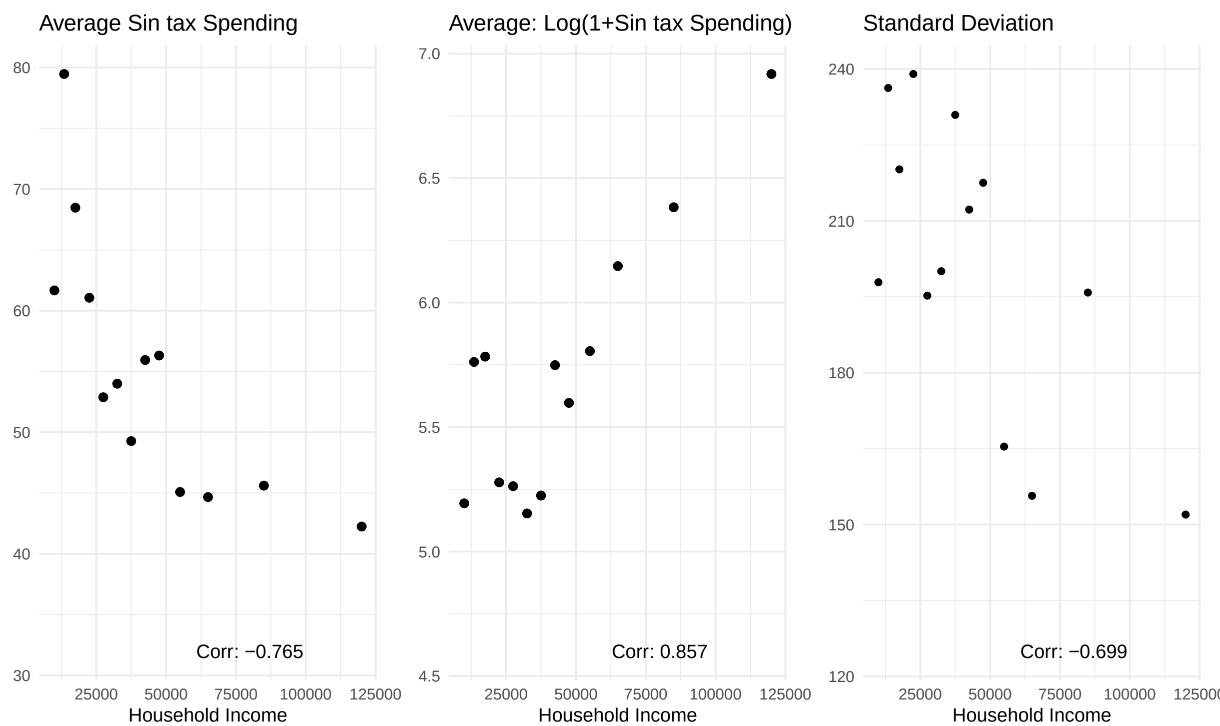

Figure D1: Sin Taxes Paid by Income (Levels vs. Logs)

Source: NielsenIQ Panelist data and authors' calculations.
Income is reported in 13 bins with all incomes below $10,000 consolidated into the first bin.

the median pay higher sin taxes than their white counterparts (Panel B). A similar pattern holds in panel C, which compares the taxes of Hispanic and non-Hispanic households when SSB taxes are included. At higher points in the tax distribution white households pay more sin taxes than Black (and non-Hispanic households pay more than Hispanic households), but the gap is consistently narrower than when just existing sin taxes are considered.

In the text we report the summary results of the regression:

$$\log(1 + \text{Sin Tax}_i) = \beta x_i + \gamma(k_i) + \lambda_{s_i} + \varepsilon_i$$

Where x_i are household demographics and λ_{s_i} are state-fixed effects. The full set of estimates is below in Table D3. The differences are the cluster fixed effects $\gamma(k_i)$ (in columns 2 and 4) and whether the hypothetical SSB tax is included in the overall sin tax burden (columns 3 and 4). The main takeaway is that the explanatory power of demographic variables (Within R^2) is very weak. The explanatory value of the state fixed effects ($R^2 = 0.0359$) is also very weak, especially considering that the statutory tax rates are fully explained by state fixed effects. On the other hand including the $k = 8$ grouped fixed effects for our clusters increases the overall $R^2 = 0.80$ in column 2.

Panel A: Existing Sin Taxes on Alcohol and Tobacco						
	tax 25%	tax 50%	tax 75%	tax 90%	tax 95%	mean
White	0.00	3.36	30.43	152.84	365.49	68.57
Black	0.00	2.10	18.43	92.47	213.97	42.13
Asian	0.00	1.19	8.68	38.35	86.41	21.40
Other	0.00	3.33	23.10	91.60	222.90	48.12

Panel B: Existing Sin Taxes + SSB Taxes						
	tax 25%	tax 50%	tax 75%	tax 90%	tax 95%	mean
White	11.52	35.35	92.19	226.69	434.93	104.33
Black	16.06	39.36	86.09	178.85	292.66	83.22
Asian	5.92	17.68	45.80	98.51	154.60	45.00
Other	14.58	38.22	86.35	178.43	310.71	87.74

Panel C: Existing Sin Taxes on Alcohol and Tobacco						
	tax 25%	tax 50%	tax 75%	tax 90%	tax 95%	mean
Hispanic	0.22	3.56	21.53	82.94	192.50	42.51
Nonhispanic	0.00	2.97	27.76	142.03	340.73	64.66

Panel D: Existing Sin Taxes + SSB Taxes						
	tax 25%	tax 50%	tax 75%	tax 90%	tax 95%	mean
Hispanic	14.87	37.34	83.14	166.18	275.86	80.87
Nonhispanic	11.52	34.92	89.76	216.94	411.08	100.65

Table D2: Sin Tax Paid by Race and Ethnicity

The obvious conclusion is that purchase patterns rather than demographics or geography are the primary source of heterogeneity. There are several demographic bins that are significant, though it is important to remember that significance is relative to our baseline household (White, Non-Hispanic, College Graduate, No Children, \$45,000-\$69,000). Also note that the income coefficients are monotonically increasing in nearly all specifications suggesting that sin taxes are progressive. This is an artifact of the $\log(\cdot)$ transform and the large amount of heteroskedasticity we discuss in Figure D1.

In Table D4 we compute the sin tax burden as a share of income for each household and then compute the appropriately weighted quantiles or mean for each of our $k = 8$ clusters. We report the weighted mean in Table 1 in the main text of the paper. The two panels differ by whether or not the hypothetical SSB taxes are included in the overall sin tax burden.

Since we observe income in bins, we use the midpoint of each bin as the income for every household within that income bin. For the lowest bin we use a value of \$10,000 and for the highest income households we use a value of \$120,000. (These are meant to approximate the median of these bins after fitting a lognormal to the overall income distribution).

As one might expect, the ratio of sin taxes to income is highly skewed both because purchases of sin taxes are highly concentrated among a small number of households and because household income is also quite skewed. Other than *Smokers, Everything* and *Heavy Drinkers* most households pay a negligible amount of income in sin taxes. Even households in the top 5% of the other clusters rarely pay more than 0.25% of income in sin taxes. The *Everything* and *Smokers* pay a much larger share (often more than 1-2%) both because they are poorer than average and because they face much larger tax burdens – particularly since cigarette taxes are such a large share of the overall sin tax burden.

In Table D5 we report the conditional distribution of demographics given that a household belongs to a particular cluster. For example, a household assigned to our *Everything* cluster has a 5% chance of having received a postgraduate degree. In the main text Table 2, we report the ratio of 5% divided by the overall rate of postgraduates in the data (15%) for 0.33. The ratios are likely to facilitate quicker comparisons and Table 2 could be constructed from Table D5 and vice versa, so this is purely for convenience.

As an alternative to the relative risk we calculate in Table 2, we also fit a multinomial logit regression where we predict cluster assignment as a function of demographics. This has advantages and disadvantages. The main disadvantage is that we need to specify both a baseline set of consumer demographics, and a baseline cluster. This makes the results a little harder to interpret. The main advantage over the results we report in Table 2 is that it better handles the fact that many demographics are highly correlated (such as education and income) so those effects are more moderated.

In Table D7 we provide the bootstrapped confidence intervals for the relative risk in Table 2. The bootstrap procedure is quite straightforward:

1. For each household, assign it to a cluster following (1) from the main text.
2. Re-sample $N = 61{,}332$ households with replacement.
3. Compute $Pr(h \in Demog | h \in Cluster)$ as in Table D5
4. Compute $Pr(h \in Demog)$ and the ratio.
5. Repeat (2)-(4) 500 times and report the $\alpha = 0.025$ and $1 - \alpha$ quantiles.
6. Highlight cells in Table 2 if the confidence interval is strictly above 1.1 or strictly below 0.9.

	Log(Sin Tax)		Log(Sin Tax+SSB Tax)	
	(1)	(2)	(3)	(4)
Income: < 24,999	-0.156	-0.055	-0.097	-0.032
	(0.045)	(0.029)	(0.027)	(0.036)
Income: 25,000-44,999	-0.129	-0.043	-0.055	-0.014
	(0.034)	(0.017)	(0.026)	(0.022)
Income: 70,000-99,999	0.110	0.019	0.036	-0.015
	(0.035)	(0.023)	(0.029)	(0.011)
Income: > 100,000	0.254	0.026	0.109	-0.019
	(0.029)	(0.034)	(0.019)	(0.025)
Race: Asian	-0.541	-0.111	-0.519	-0.170
	(0.033)	(0.033)	(0.035)	(0.022)
Race: Black	-0.210	-0.042	0.112	0.032
	(0.047)	(0.040)	(0.036)	(0.022)
Race: Other	-0.041	0.015	0.029	0.006
	(0.035)	(0.005)	(0.037)	(0.013)
Hispanic: Yes	-0.081	-0.052	-0.008	-0.040
	(0.041)	(0.040)	(0.041)	(0.021)
Children: Yes	-0.180	-0.060	0.294	0.132
	(0.033)	(0.036)	(0.023)	(0.053)
Edu: High School or Less	0.132	-0.006	0.362	0.141
	(0.034)	(0.024)	(0.021)	(0.052)
Edu: Some College	0.165	0.029	0.246	0.074
	(0.036)	(0.011)	(0.023)	(0.021)
Edu: Post College Grad	-0.132	-0.020	-0.272	-0.109
	(0.032)	(0.025)	(0.020)	(0.025)
Age: Under 35	-0.001	-0.010	-0.027	-0.041
	(0.043)	(0.100)	(0.031)	(0.087)
Age: 35-44	0.062	-0.010	0.131	0.023
	(0.046)	(0.085)	(0.029)	(0.067)
Age: 45-54	0.179	0.009	0.277	0.082
	(0.040)	(0.062)	(0.030)	(0.058)
Age: 55-64	0.203	0.037	0.231	0.069
	(0.035)	(0.028)	(0.026)	(0.022)
Standard-Errors	State	State & Cluster Assignment	State	State & Cluster Assignment
Observations	61,332	61,332	61,332	61,332
R^2	0.0359	0.8064	0.0532	0.6716
Within R^2	0.0171	0.0048	0.0399	0.0167
State FE	✓	✓	✓	✓
Cluster Assignment FE		✓		✓

Table D3: Sin Tax Burden with and without Grouped Fixed Effects

Note: The omitted base demographics are for a household that is Race: White, Education: Graduated College, Hispanic: No, Age: over 65, Children at Home: No; Income: 45,000-69,999.
Observations are weighted by Nielsen projection factors.

Panel A: Existing Sin Taxes on Alcohol and Tobacco

	Everything	Smokers	Heavy Drinkers	Moderate Spirits	Mostly Wine	Moderate Beer	SSB only	Nothing
min	0.018	0.005	0.009	0.003	0.001	0.001	0.000	0.000
ratio 25%	0.278	0.128	0.068	0.023	0.013	0.008	0.000	0.000
ratio 50%	0.730	0.444	0.140	0.045	0.028	0.017	0.000	0.000
ratio 75%	1.860	1.483	0.332	0.100	0.065	0.046	0.003	0.005
ratio 90%	4.138	3.752	0.740	0.238	0.155	0.125	0.010	0.018
ratio 95%	6.756	6.159	1.225	0.436	0.279	0.240	0.019	0.039
mean	1.882	1.541	0.347	0.123	0.082	0.063	0.005	0.011
max	155.155	114.796	39.463	19.633	11.259	22.706	1.113	4.412

Panel B: Existing Sin Taxes Plus New Taxes on SSBs

	Everything	Smokers	Heavy Drinkers	Moderate Spirits	Mostly Wine	Moderate Beer	SSB only	Nothing
min	0.022	0.007	0.013	0.006	0.002	0.004	0.001	0.000
ratio 25%	0.374	0.251	0.094	0.047	0.025	0.034	0.022	0.001
ratio 50%	0.895	0.680	0.186	0.096	0.050	0.075	0.054	0.004
ratio 75%	2.109	1.862	0.411	0.212	0.112	0.179	0.140	0.011
ratio 90%	4.542	4.402	0.878	0.462	0.245	0.406	0.341	0.027
ratio 95%	7.282	7.021	1.429	0.788	0.417	0.687	0.594	0.051
mean	2.099	1.861	0.426	0.236	0.127	0.192	0.164	0.015
max	155.155	124.276	39.825	55.554	12.777	25.266	59.279	4.412

Panel A1: Existing Sin Taxes with potential $1 increase in tobacco tax

	Everything	Smokers	Heavy Drinkers	Moderate Spirits	Mostly Wine	Moderate Beer	SSB only	Nothing
min	0.022	0.009	0.009	0.003	0.001	0.001	0.000	0.000
ratio 25%	0.357	0.188	0.068	0.023	0.014	0.008	0.000	0.000
ratio 50%	0.967	0.656	0.140	0.045	0.028	0.018	0.000	0.000
ratio 75%	2.514	2.180	0.334	0.101	0.066	0.047	0.003	0.005
ratio 90%	5.613	5.539	0.743	0.241	0.158	0.130	0.011	0.019
ratio 95%	9.134	8.892	1.231	0.443	0.287	0.251	0.022	0.041
mean	2.533	2.239	0.348	0.124	0.084	0.065	0.005	0.012
max	200.404	185.370	39.463	19.633	13.678	22.706	1.393	6.173

Panel B1: Existing Sin Taxes Plus New Taxes on SSBs with potential $1 increase in tobacco tax

	Everything	Smokers	Heavy Drinkers	Moderate Spirits	Mostly Wine	Moderate Beer	SSB only	Nothing
min	0.026	0.012	0.013	0.006	0.002	0.004	0.001	0.000
ratio 25%	0.459	0.322	0.094	0.048	0.025	0.034	0.022	0.001
ratio 50%	1.131	0.896	0.186	0.096	0.050	0.076	0.055	0.004
ratio 75%	2.764	2.567	0.413	0.214	0.113	0.181	0.141	0.011
ratio 90%	6.022	6.119	0.882	0.465	0.248	0.411	0.342	0.028
ratio 95%	9.738	9.723	1.431	0.791	0.421	0.693	0.595	0.053
mean	2.750	2.560	0.427	0.237	0.129	0.194	0.165	0.016
max	200.404	194.850	39.825	55.554	14.977	25.266	59.279	6.173

Table D4: Distribution of Sin Tax to Income Ratios (%)

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing
Race: White (74.9%)	0.78	0.83	0.80	0.69	0.80	0.76	0.72	0.78
Race: Black (12.5%)	0.12	0.10	0.09	0.16	0.10	0.10	0.16	0.08
Race: Asian (4.4%)	0.03	0.02	0.03	0.04	0.04	0.03	0.04	0.07
Race: Other (8.2%)	0.07	0.05	0.08	0.11	0.07	0.10	0.09	0.07
Hispanic: No (86.8%)	0.89	0.92	0.87	0.85	0.87	0.81	0.86	0.90
Hispanic: Yes (13.2%)	0.11	0.08	0.13	0.15	0.13	0.19	0.14	0.10
Children: Yes (31.3%)	0.22	0.30	0.26	0.34	0.27	0.37	0.37	0.18
Children: No (68.7%)	0.78	0.70	0.74	0.66	0.73	0.63	0.63	0.82
Age: $<$ 35 (12.9%)	0.08	0.08	0.10	0.14	0.13	0.14	0.13	0.14
Age: 35 to 44 (18.0%)	0.13	0.17	0.15	0.19	0.17	0.21	0.20	0.14
Age: 45 to 54 (21.8%)	0.23	0.25	0.23	0.26	0.19	0.24	0.22	0.17
Age: 55 to 64 (22.7%)	0.33	0.28	0.25	0.22	0.22	0.22	0.21	0.23
Age: $>$ 65 (24.6%)	0.23	0.22	0.26	0.19	0.30	0.20	0.23	0.31
Income: $<$ 24,999 (20.4%)	0.28	0.38	0.09	0.15	0.11	0.15	0.22	0.23
Income: 25,000 - 44,999 (17.7%)	0.21	0.21	0.12	0.16	0.13	0.16	0.19	0.18
Income: 45,000-69,999 (18.2%)	0.18	0.18	0.17	0.18	0.17	0.19	0.19	0.18
Income: 70,000-99,999 (15.5%)	0.14	0.09	0.19	0.17	0.18	0.19	0.15	0.14
Income: $>$ 100,000 (28.1%)	0.18	0.14	0.43	0.34	0.41	0.31	0.25	0.27
Edu: High School or less (27.4%)	0.36	0.45	0.21	0.23	0.18	0.28	0.29	0.24
Edu: Some College (31.4%)	0.41	0.36	0.31	0.35	0.26	0.32	0.32	0.27
Edu: Graduated College (26.3%)	0.18	0.14	0.30	0.28	0.32	0.28	0.25	0.28
Edu: Post College Grad (14.9%)	0.05	0.04	0.18	0.14	0.24	0.13	0.13	0.20

Table D5: Share of households with certain demographics by cluster: $Pr(h \in Demog | h \in Cluster)$

	cluster = Everything		cluster = Smokers		cluster = Heavy Drinkers		cluster = Moderate Spirits	
	coef	std err	coef	std err	coef	std err	coef	std err
Intercept	-2.4002	0.086	-1.8103	0.066	-0.6951	0.051	-1.1415	0.054
Edu: High School or Less	0.8314	0.082	0.9944	0.061	0.1193	0.059	0.2703	0.059
Edu: Post College Grad	-0.8504	0.108	-0.7952	0.082	-0.4063	0.048	-0.4951	0.051
Edu: Some College	0.6750	0.072	0.6262	0.056	0.1495	0.047	0.2596	0.047
Race: Black	0.4988	0.093	0.2452	0.076	0.0971	0.068	0.7027	0.059
Race: Asian	-0.9562	0.206	-0.9968	0.153	-1.2282	0.112	-0.6916	0.095
Race: Other	0.3199	0.133	-0.0704	0.113	-0.0023	0.093	0.2332	0.087
Hispanic: Yes	-0.0212	0.130	-0.1186	0.103	0.1342	0.081	0.1986	0.077
Age: 35 - 44	0.3184	0.111	0.3551	0.080	-0.2591	0.068	0.2891	0.066
Age: 45 - 54	0.6845	0.085	0.6626	0.064	-0.1025	0.056	0.4795	0.056
Age: 55 - 64	0.5662	0.072	0.5006	0.055	-0.0474	0.045	0.2353	0.049
Age: Under 35	-0.1903	0.155	-0.3062	0.113	-0.4749	0.086	0.1844	0.078
Children: Yes	0.2122	0.086	0.7563	0.062	0.2949	0.056	0.6250	0.051
Income: < 24,999	-0.3971	0.088	0.0491	0.063	-1.2542	0.078	-0.7966	0.067
Income: 25,000 - 44,999	-0.3815	0.080	-0.2252	0.060	-0.6441	0.058	-0.5645	0.056
Income: 70,000 - 99,999	-0.2857	0.086	-0.6282	0.072	0.1508	0.053	-0.0484	0.054
Income: > 100,000	-0.4351	0.099	-0.7032	0.081	0.4627	0.054	0.1373	0.055

	cluster = Mostly Wine		cluster = Moderate Beer		cluster = SSB only	
	coef	std err	coef	std err	coef	std err
Intercept	-0.4361	0.048	-0.8420	0.049	0.6703	0.034
Edu: High School or Less	-0.2491	0.061	0.5403	0.051	0.4373	0.036
Edu: Post College Grad	-0.1335	0.044	-0.5765	0.047	-0.3508	0.031
Edu: Some College	-0.0756	0.046	0.2792	0.043	0.2252	0.031
Race: Black	0.2077	0.063	0.2606	0.060	0.7081	0.042
Race: Asian	-0.9303	0.093	-0.8447	0.089	-0.4679	0.053
Race: Other	-0.0358	0.089	-0.0020	0.081	0.1751	0.059
Hispanic: Yes	0.2465	0.076	0.5328	0.067	0.1179	0.053
Age: 35 - 44	-0.3206	0.063	0.2543	0.060	0.1258	0.042
Age: 45 - 54	-0.3113	0.054	0.3333	0.052	0.2574	0.036
Age: 55 - 64	-0.2832	0.044	0.2048	0.044	0.1126	0.030
Age: Under 35	-0.4764	0.077	0.1047	0.071	-0.0928	0.049
Children: Yes	0.3685	0.053	0.7894	0.047	0.9368	0.035
Income: < 24,999	-0.9614	0.070	-0.9177	0.062	-0.3890	0.039
Income: 25,000 - 44,999	-0.4958	0.054	-0.5729	0.051	-0.2883	0.035
Income: 70,000 - 99,999	0.0789	0.051	0.0169	0.049	-0.1624	0.036
Income: > 100,000	0.2744	0.052	0.0728	0.051	-0.2022	0.038

Table D6: Multinomial Logit Regression

	Everything	Smokers	Heavy Drinkers	Moderate Spirits	Mostly Wine	Moderate Beer	SSB only	Nothing
Race: White (74.9%)	(1.01, 1.08)	(1.08, 1.13)	(1.04, 1.09)	(0.9, 0.95)	(1.04, 1.09)	(0.99, 1.04)	(0.95, 0.96)	(1.03, 1.06)
Race: Black (12.5%)	(0.77, 1.14)	(0.72, 0.96)	(0.65, 0.84)	(1.15, 1.39)	(0.7, 0.9)	(0.75, 0.92)	(1.2, 1.27)	(0.57, 0.67)
Race: Asian (4.4%)	(0.3, 0.84)	(0.25, 0.55)	(0.5, 0.85)	(0.71, 1.09)	(0.64, 1.0)	(0.58, 0.87)	(0.92, 1.04)	(1.54, 1.81)
Race: Other (8.2%)	(0.65, 1.14)	(0.52, 0.79)	(0.8, 1.13)	(1.16, 1.51)	(0.69, 0.97)	(1.13, 1.42)	(1.0, 1.1)	(0.72, 0.89)
Hispanic: No (86.8%)	(1.0, 1.06)	(1.04, 1.07)	(0.99, 1.03)	(0.96, 1.0)	(0.98, 1.01)	(0.91, 0.95)	(0.99, 1.0)	(1.03, 1.05)
Hispanic: Yes (13.2%)	(0.63, 1.03)	(0.51, 0.75)	(0.82, 1.07)	(1.01, 1.26)	(0.91, 1.15)	(1.35, 1.58)	(1.0, 1.07)	(0.7, 0.83)
Children: Yes (31.3%)	(0.6, 0.79)	(0.87, 1.02)	(0.77, 0.89)	(1.03, 1.16)	(0.8, 0.92)	(1.14, 1.26)	(1.17, 1.2)	(0.55, 0.61)
Children: No (68.7%)	(1.09, 1.18)	(0.99, 1.06)	(1.05, 1.11)	(0.93, 0.99)	(1.04, 1.09)	(0.88, 0.94)	(0.91, 0.92)	(1.17, 1.21)
Age: $<$ 35 (12.9%)	(0.45, 0.8)	(0.52, 0.76)	(0.67, 0.9)	(0.97, 1.21)	(0.87, 1.1)	(0.99, 1.22)	(1.0, 1.07)	(1.01, 1.16)
Age: 35 to 44 (18.0%)	(0.6, 0.88)	(0.82, 1.03)	(0.76, 0.94)	(0.97, 1.15)	(0.84, 1.01)	(1.07, 1.23)	(1.08, 1.13)	(0.74, 0.83)
Age: 45 to 54 (21.8%)	(0.93, 1.21)	(1.06, 1.25)	(0.97, 1.13)	(1.09, 1.26)	(0.81, 0.96)	(1.01, 1.14)	(1.0, 1.05)	(0.76, 0.84)
Age: 55 to 64 (22.7%)	(1.3, 1.6)	(1.13, 1.31)	(1.04, 1.19)	(0.91, 1.05)	(0.89, 1.02)	(0.9, 1.02)	(0.92, 0.97)	(0.98, 1.06)
Age: $>$ 65 (24.6%)	(0.81, 1.03)	(0.83, 0.98)	(1.01, 1.14)	(0.72, 0.85)	(1.14, 1.27)	(0.75, 0.86)	(0.92, 0.96)	(1.23, 1.32)
Income: $<$ 24,999 (20.4%)	(1.21, 1.53)	(1.76, 1.99)	(0.36, 0.48)	(0.67, 0.82)	(0.47, 0.59)	(0.68, 0.82)	(1.04, 1.09)	(1.09, 1.2)
Income: 25,000 - 44,999 (17.7%)	(1.06, 1.35)	(1.09, 1.3)	(0.63, 0.78)	(0.8, 0.97)	(0.68, 0.82)	(0.85, 0.99)	(1.06, 1.11)	(0.95, 1.05)
Income: 45,000-69,999 (18.2%)	(0.87, 1.14)	(0.87, 1.06)	(0.85, 1.01)	(0.9, 1.07)	(0.86, 1.01)	(0.99, 1.13)	(1.01, 1.06)	(0.91, 1.01)
Income: 70,000-99,999 (15.5%)	(0.78, 1.08)	(0.5, 0.66)	(1.13, 1.32)	(1.03, 1.21)	(1.07, 1.26)	(1.11, 1.28)	(0.94, 1.0)	(0.86, 0.97)
Income: $>$ 100,000 (28.1%)	(0.53, 0.75)	(0.44, 0.57)	(1.45, 1.6)	(1.14, 1.28)	(1.39, 1.53)	(1.03, 1.15)	(0.87, 0.91)	(0.93, 1.01)
Edu: High School or less (27.4%)	(1.17, 1.43)	(1.55, 1.72)	(0.69, 0.82)	(0.76, 0.9)	(0.6, 0.72)	(0.94, 1.06)	(1.04, 1.09)	(0.86, 0.94)
Edu: Some College (31.4%)	(1.21, 1.43)	(1.08, 1.23)	(0.93, 1.05)	(1.05, 1.17)	(0.78, 0.9)	(0.95, 1.06)	(1.01, 1.04)	(0.83, 0.9)
Edu: Graduated College (26.3%)	(0.6, 0.8)	(0.49, 0.61)	(1.08, 1.22)	(1.0, 1.13)	(1.14, 1.28)	(1.0, 1.11)	(0.94, 0.98)	(1.03, 1.12)
Edu: Post College Grad (14.9%)	(0.24, 0.41)	(0.24, 0.36)	(1.09, 1.3)	(0.86, 1.05)	(1.49, 1.7)	(0.8, 0.96)	(0.87, 0.93)	(1.28, 1.4)

Table D7: Relative Risk - Bootstrap Result

E. Alternative Specifications

In Table E1, and Table E2 we consider robustness to our choice of cluster assignments in the main text $k = 7$ and $k = 9$ (instead of $k = 8$). We did not perform "hierarchical clustering" where cluster assignments are nested inside one another by construction. Instead, we perform standard k-means with a different choice of k. However, we can see that with $k = 7$ we lose the ability to separate the *Moderate Spirits* drinkers and with $k = 9$ we further separate an additional group of non-smoking *Heavy Beer* drinkers (mostly taken from the *Heavy Drinkers* and *Mostly Beer* drinkers).

	Everything	Heavy Drinker	Moderate Beer	Mostly Wine	Nothing	SSB only	Smoker
Beer (mean)	149.08	77.83	60.08	20.05	2.63	0.71	2.27
Wine (mean)	21.34	29.01	1.64	38.22	0.99	0.80	0.62
Spirits (mean)	18.45	26.67	1.66	1.70	0.45	0.33	0.47
Tobacco (mean)	146.88	0.31	0.35	0.37	0.20	0.11	131.10
SSB (mean)	151.84	95.02	121.98	53.12	3.80	111.94	175.90
Ethanol (mean)	15.42	16.04	3.32	6.34	0.37	0.22	0.31
Beer 50%	46.84	25.55	21.87	4.26	0.00	0.00	0.00
Beer 75%	187.19	78.78	51.10	15.55	0.59	0.00	1.42
Beer 95%	569.72	325.03	240.61	77.71	12.78	4.26	12.78
Wine 50%	2.25	9.62	0.75	17.52	0.00	0.00	0.00
Wine 75%	12.22	27.50	2.62	38.25	0.75	0.75	0.00
Wine 95%	114.57	128.02	6.37	135.75	5.25	4.50	3.75
Spirits 50%	5.31	16.00	0.00	0.75	0.00	0.00	0.00
Spirits 75%	17.59	31.30	2.50	2.63	0.00	0.00	0.00
Spirits 95%	93.02	86.54	7.14	6.57	2.53	2.25	2.97
Tobacco 50%	75.00	0.00	0.00	0.00	0.00	0.00	60.00
Tobacco 75%	205.00	0.00	0.00	0.00	0.00	0.00	178.00
Tobacco 95%	500.00	2.00	2.00	1.00	0.00	0.20	464.95
SSB 50%	87.58	53.26	73.40	26.79	2.84	65.25	101.21
SSB 75%	207.76	126.57	154.69	64.82	6.62	139.45	245.02
SSB 95%	519.69	328.53	393.71	189.63	10.74	360.12	573.56
SSB per Person/Week	1.35	0.80	0.98	0.47	0.05	0.99	1.55
Drinks per Week	16.71	17.38	3.60	6.87	0.40	0.23	0.34
Drinks per Adult	9.98	10.05	2.10	4.26	0.28	0.13	0.19
Effective Ethanol Tax/L	8.19	10.10	7.57	5.56	3.62	3.68	4.24
Total Tax Share	25.82	23.17	4.97	5.36	1.29	1.99	37.40
Alcohol Tax Share	14.60	54.51	11.29	12.31	2.49	4.09	0.71
Tobacco Tax Share	34.02	0.26	0.36	0.27	0.41	0.46	64.22
SSB Tax Share	4.34	8.64	14.92	4.85	0.77	55.85	10.63
Tax Burden/Income (%)	2.28	0.39	0.23	0.11	0.01	0.14	1.87
# Households	1433	5437	6622	5784	11958	27262	2836
Share of Households	2.61	8.30	11.17	8.34	18.47	45.58	5.52

Table E1: 7 clusters

	Everything	Heavy Beer	Heavy Drinker	Moderate Beer	Moderate Spirits	Mostly Wine	Nothing	SSB only	Smoker
Beer (mean)	147.48	180.24	78.70	11.36	8.20	14.94	2.23	0.10	2.41
Wine (mean)	21.56	2.77	45.60	0.99	2.51	38.42	0.98	0.83	0.63
Spirits (mean)	19.88	2.33	29.22	0.25	11.08	1.21	0.34	0.14	0.42
Tobacco (mean)	158.09	0.66	0.44	0.20	0.44	0.37	0.22	0.10	131.81
SSB (mean)	150.88	104.97	87.78	114.91	115.27	52.64	3.02	107.76	176.39
Ethanol (mean)	15.81	9.14	19.28	0.70	3.86	5.99	0.32	0.15	0.30
Beer 50%	46.72	98.66	33.48	8.23	4.12	6.10	0.00	0.00	0.00
Beer 75%	181.48	204.41	84.47	14.91	11.88	17.03	0.00	0.00	1.42
Beer 95%	572.29	585.84	317.10	33.36	31.91	55.36	12.78	0.89	13.84
Wine 50%	3.00	0.88	19.88	0.00	1.50	18.75	0.00	0.00	0.00
Wine 75%	13.13	3.75	47.42	1.50	3.75	38.25	0.75	0.75	0.00
Wine 95%	108.09	10.50	171.65	4.50	9.00	135.00	5.25	4.50	3.75
Spirits 50%	5.88	0.75	17.63	0.00	6.51	0.20	0.00	0.00	0.00
Spirits 75%	18.57	3.00	36.31	0.00	11.40	1.75	0.00	0.00	0.00
Spirits 95%	94.61	9.50	92.75	1.75	33.25	4.72	1.88	1.12	2.62
Tobacco 50%	84.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	60.00
Tobacco 75%	230.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	178.87
Tobacco 95%	532.28	4.00	2.00	1.00	3.00	1.00	0.00	0.00	467.87
SSB 50%	89.44	58.87	45.92	69.36	67.77	26.00	2.00	60.80	102.19
SSB 75%	206.53	139.28	116.58	146.33	146.54	63.57	5.39	132.71	244.94
SSB 95%	521.87	376.07	307.95	373.51	358.78	193.81	8.81	354.67	579.20
SSB per Person/Week	1.36	0.86	0.74	0.94	0.94	0.45	0.04	0.97	1.54
Drinks per Week	17.13	9.91	20.89	0.76	4.18	6.50	0.35	0.16	0.33
Drinks per Adult	10.21	5.96	11.85	0.42	2.59	4.04	0.24	0.09	0.19
Effective Ethanol Tax/L	8.28	5.97	8.85	6.02	15.35	5.10	3.31	2.65	4.16
Total Tax Share	25.36	4.78	18.28	1.10	6.78	4.27	1.00	1.00	37.43
Alcohol Tax Share	14.06	10.94	42.92	2.30	15.63	9.79	1.80	1.88	0.67
Tobacco Tax Share	33.62	0.28	0.27	0.21	0.31	0.23	0.42	0.36	64.30
SSB Tax Share	3.97	5.14	5.54	14.82	8.74	4.25	0.54	46.38	10.62
Tax Burden/Income (%)	2.40	0.27	0.40	0.13	0.31	0.10	0.01	0.14	1.88
# Households	1337	2754	3853	6814	4164	5119	10657	23815	2819
Share of Households	2.40	4.47	5.76	11.78	6.93	7.38	16.46	39.32	5.50

Table E2: 9 clusters

References

Adams Media Inc. (2019) "Liquor Handbook 2019," in *Adams/Jobson's liquor handbook.*, New York, NY :: Adams Media Inc.

Cook, Philip J. (2007) *Paying the Tab: The Costs and Benefits of Alcohol Control*, Princeton, NJ: Princeton University Press.

Ibragimov, Rustam and Artem Prokhorov (2017) *Heavy Tails and Copulas*: World Scientific, 10.1142/9644.

Lowry, Bryan (2015) "Gov. Sam Brownback's proposed budget fix slows income tax cuts, hikes taxes on liquor, tobacco," http://www.kansas.com/news/politics-government/article6952787.html.

Miravete, Eugenio J., Katja Seim, and Jeff Thurk (2018) "Market Power and the Laffer Curve," *Econometrica*, 86 (5), 1651–1687, https://doi.org/10.3982/ECTA12307.

——— (2020) "One Markup to Rule Them All: Taxation by Liquor Pricing Regulation," *American Economic Journal: Microeconomics*, 12 (1), 1–41, 10.1257/mic.20180155.

Naimi, Timothy, James Daley, Ziming Xuan, Jason Blanchette, Frank Chaloupka, and David Jernigan (2016) "Who Would Pay for State Alcohol Tax Increases in the United States?" *Preventing Chronic Disease*, 13, 10.5888/pcd13.150450.

Rosenberg, Joseph (2015) "The Distributional Burden of Federal Excise Taxes," https://www.taxpolicycenter.org/sites/default/files/alfresco/publication-pdfs/2000365-the-distributional-burden-of-federal-excise-taxes.pdf.

Sanders, Bernie (2016) "A Soda Tax Would Hurt Philly's Low-Income Families," https://www.phillymag.com/citified/2016/04/24/bernie-sanders-soda-tax-op-ed/.

Seiler, Stephan, Anna Tuchman, and Song Yao (2021) "The Impact of Soda Taxes: Pass-Through, Tax Avoidance, and Nutritional Effects," *Journal of Marketing Research*, 58 (1), 22–49, 10.1177/0022243720969401.

Tax Policy Center (2019) "Types of Federal Excise Taxes," https://www.taxpolicycenter.org/statistics/types-federal-excise-taxes.

——— (2020) "The State of State (and Local) Tax Policy," https://www.taxpolicycenter.org/briefing-book/how-do-state-and-local-soda-taxes-work.

——— (2021a) "State and Local Alcohol Tax Revenue," https://www.taxpolicycenter.org/statistics/state-and-local-alcohol-tax-revenue.

——— (2021b) "State and Local Tobacco Tax Revenue," https://www.taxpolicycenter.org/statistics/state-and-local-tobacco-tax-revenue.