NBER WORKING PAPER SERIES

WHO PAYS SIN TAXES? UNDERSTANDING THE OVERLAPPING BURDENS OF CORRECTIVE TAXES

Christopher Conlon
Nirupama L. Rao
Yinan Wang

Working Paper 29393
http://www.nber.org/papers/w29393

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2021

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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Who Pays Sin Taxes? Understanding the Overlapping Burdens of Corrective Taxes
Christopher Conlon, Nirupama L. Rao, and Yinan Wang
NBER Working Paper No. 29393
October 2021
JEL No. H22,H23,H25,L66

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 68% 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.

Christopher Conlon
Stern School of Business
New York University
44 W. 4th St
New York, NY 10012
and NBER
cconlon@stern.nyu.edu

Yinan Wang
Stern School of Business
New York University
yw1633@stern.nyu.edu

Nirupama L. Rao
Ross School of Business
University of Michigan
701 Tappan Street, Room- R4312
Ann Arbor, MI 48109-1234
nirurao@umich.edu

A data appendix is available at http://www.nber.org/data-appendix/w29393

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. In all, combined federal and state taxes on alcohol and tobacco raise nearly $40 billion annually – an amount comparable to the total federal income tax paid by the bottom half of the income distribution.[2] Over the last decade several localities have also levied new taxes on sugar-sweetened beverages (SSBs), with dozens more considering introducing 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]

The main argument against sin taxes, made by both researchers and politicians, is that

[1] See https://waysandmeans.house.gov/sites/democrats.waysandmeans.house.gov/files/documents/SubtitleISxS.pdf.

[2] For state sin tax revenues, please see https://www.taxpolicycenter.org/statistics/state-and-local-tobacco-tax-revenue and https://www.taxpolicycenter.org/statistics/state-and-local-alcohol-tax-revenue. For federal sin tax revenues, see https://www.taxpolicycenter.org/statistics/types-federal-excise-taxes. For income tax statistics, please consult https://www.irs.gov/pub/irs-soi/17in02etr.xls

[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. For more details see http://www.kansas.com/news/politics-government/article6952787.html

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, 2020). This becomes more difficult (and less effective) if the sin-tax burden is shared unequally among households within a given income band. 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 paper we instead analyze the combined burdens of sin taxes on alcoholic beverages and tobacco, as well as potential taxes on SSBs, and provide new measures of the concentration of these burdens.

Our analysis begins with documenting the high concentration of beer, wine, spirits, and cigarettes 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 from beer, wine and spirits purchases. For cigarettes, 8% of households are responsible for virtually all purchases. We also consider a hypothetical national penny-per-ounce tax on SSBs, for which purchases are far less concentrated. Because sugary beverages are purchased by three quarters of households and the top 10% of purchasers account for only 55% of sales volume, such a tax would be more broadly distributed.

We also find that heavy purchasers of one sin good (those in the top decile) are likely to purchase larger amounts of other goods subject to corrective taxation. This phenomenon is particularly strong among households with smokers, who tend to also consume larger amounts of SSBs as well as beer and spirits. Households that purchase large amounts of any category of alcoholic beverages (beer, wine, or spirits) also heavily purchase products in the

remaining two categories of alcoholic beverages. These patterns imply that the combined burden tends to be even more concentrated than sin taxes on individual categories, leading the top 20% of households to pay more than 90% of all sin taxes.

The potential regressivity or progressivity of sin taxes is a more complicated story. Population averages largely confirm the belief that cigarette taxes are highly regressive, and fall disproportionately on lower-income households. Those earning below $25,000 per year pay roughly three times as much in cigarette taxes as households earning over $100,000 per year. On the other hand, taxes on wine and spirits appear to be quite progressive; households earning over $100,000 per year pay around 70% more on average than households earning below $25,000 per year. However, household demographics (including income) explain only a tiny fraction of the heterogeneity of sin-tax burdens across households, and income is extremely weakly correlated with tax burden (cigarettes $\rho = -0.06$). We document far more heterogeneity within income groups in sin good purchases than across them, and the median household at all income levels faces little or no exposure to sin taxes.

To deal with both the multiple dimensions of dependence, and the extreme concentration in sin good purchases, we discretize the heterogeneity by 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 3%.

We identify a tiny group of households (2.5% of the population), whom we label *Everything* and who consume extremely large amounts of nearly all sin goods and purchase more sugary beverages than any other group. They also tend to be disproportionately from the lowest levels of income and education, white, and older. These households bear a striking similarity to those Case and Deaton (2020) describe as most susceptible to "deaths of despair." A group of non-drinking *Smokers* are demographically similar and make up an additional 5.5% of the population. These groups pay 68% of existing sin taxes, which average approximately 2% of income, and would pay a disproportionate amount of new taxes on

SSBs.

The third most-taxed group we label *Heavy Drinkers* (6.8% of population). They purchase the equivalent of 11 alcoholic drinks per adult per week from a variety of sources (beer, wine, and spirits). If negative externalities are convex in alcohol consumption (Griffith et al., 2019), this group along with the *Everything* group would be responsible for the bulk of alcohol's external damage. These households are somewhat older and much more likely to be from the highest levels of education and income. 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.

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. Attempts to compensate households for a larger sin-tax burden, such as through the tax code, would need to be laser-focused on these small segments.

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 for 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 consolidate into 13 bins by combining three small bins of households earning under $10k per year into a single bin, and later into five "quintiles" (<$25k, $25k-$45k, $45k-$70k, $70k-$99k, $100k+). We eliminate 19 "outlier" households from our analysis that purchase more than four packs of cigarettes or 10 standard drinks per day; these purchases may be driven by sin goods consumed by individuals outside the household.[5]

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] A major distinction 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

[4]We use four levels of household head education: HS or less, some college, college graduates, and post graduates; 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]Table E1 details these outliers.

[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

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. For all liquids categories, we convert purchases into liters. We convert cigarette purchases into the equivalent number of packs. We exclude e-cigarettes and nicotine cartridges from our tobacco category because in many states those are either untaxed or taxed differently from cigarettes. Consistent with the NIAAA, we apply a constant alcohol by volume percentage (ABV%) to beer and wine, but the Nielsen product information is sufficiently rich that we are able to use the actual ABV% for distilled spirits purchases. When we compute sin tax paid by households, we apply the relevant combined federal and state rates and assume consumers bear the full economic incidence of the taxes. 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.[8]

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. When we consider the tax burden, we apply a hypothetical penny-per-ounce tax meant to mimic existing laws and proposals.[9]

data and other data sources.

[8]The Urban-Brookings Tax Policy Center employs similar assumptions in its distributional analysis of excise taxes on tobacco and alcohol `https://www.taxpolicycenter.org/sites/default/files/alfresco/publication-pdfs/2000365-the-distributional-burden-of-federal-excise-taxes.pdf`. For a list of tax rates, please see Table B1.

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

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 on these items are based primarily on volume rather than expenditure, purchase volume (mostly) corresponds to tax burden.[10] Our goal is to describe the concentration of purchases such as the Pareto distribution's "80-20 Rule."

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 28% of toilet tissue (Gini= 0.46) and 42% of yogurt (Gini= 0.65) products by volume.[11] For beer, wine, and spirits, we find that the top 10% of households account for about 80% of purchases (by volume) and Gini$\approx$ 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 $g = 0.90$.[12] The distribution of SSB purchases does not resemble that of other sin goods. For one, 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= 0.62) – more similar to that of yogurt (Gini= 0.64) than

[10]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.

[11]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.

[12]The most recent CDC data suggest around 14% of adults smoke `https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm`. The most recent Tobacco Use Supplement to the CPS finds that number 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.

to alcohol or cigarettes.

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= 0.62). It is important to mention that Figure 1 measures the share of the overall burden, and not whether highly-taxed households are more or less exposed to additional taxes on SSBs (because the overall amount of taxes increases substantially when hypothetical SSB taxes are introduced).

The next question is whether the same households who pay most of the cigarette taxes also pay much of alcohol or (hypothetical) SSB taxes. Simple correlations of annual purchase totals are not helpful in describing this relationship because 68% of households never purchase alcoholic beverages or tobacco.[13]

Because the distribution of purchases is highly skewed, we condition on the top decile of households for each category (top 5% for tobacco), as in Panel B of Figure 1. We then plot quantiles of purchasing for these "heavy users" against their quantiles in the population at large in Figure 2. If no relationship between purchases of the two categories existed, we would expect them to follow the 45-degree line (black). The way to read panel A is that among heavy beer drinkers, households at the 50th percentile of spirits purchases would

[13]A large number of zeros also complicates more general approaches, which rely on an inverse CDF transformation, such as modeling the tail dependence with copulas (Ibragimov and Prokhorov, 2017).

represent the 80th percentile of spirits purchases for the overall population.

In Panels A, B, and C of Figure 2, we see the heaviest purchasers of beer, wine, and spirits also tend to be above-average purchasers of other alcoholic beverages (above the 45-degree line). We also see that heavy smokers (Panel D) tend to purchase larger amounts of SSBs, and to a lesser extent more beer and distilled spirits (but slightly less wine). Heavy purchasers of SSBs also tend to purchase less wine than average, and slightly more cigarettes. Finally, we see that heavy purchasers of diet sodas tend to purchase average amounts from other categories. Taxing diet drinks alongside SSBs, as some municipalities have done, would broaden the tax base, since heavy users of other taxed sin goods do not disproportionately purchase diet drinks (though the public health motivation is less clear here).

3.2. The Distributional Impacts of Sin Taxes

A limitation of the previous approach is that we only measure the dependence between pairs of categories. To better understand households at the extremes of the distribution (and the progressivity or regressivity of combined sin taxes), within each income bin we compute the quantiles of the distribution of sin taxes. We then plot the conditional quantiles of taxes as a bin-scatter plot in Figure 3. As a note of caution, the bin-scatter quantile estimates generally suggest a much stronger relationship between income and taxes than is present in the underlying data (Cattaneo et al., 2021).[14]

In panel A we see alcohol taxes appear progressive at most quantiles of purchasers, and around half of households purchase little to no alcohol. High-income households in the top decile spend around $52 per year on taxes, while low-income households in the top decile spend only around $32. Households in the top 5% spend more than twice these amounts.[15]

[14]For example, the correlation among the bin-scatter quantile estimates suggests a correlation between income and cigarette taxes paid of $\rho = -0.9$ at some quantiles, while the overall correlation $|\rho| \leq 0.06$ between income and any of our sin taxes.

[15]Recall from Figure 1 the top decile of the population accounts for 70% of all ethanol purchased. The top quartile of households accounts for nearly all purchases and the 75th percentile ranges from around $5 in tax at the low-end of the income distribution to $16 in tax at the high end.

This comes from two sources. First, high-income households purchase more ethanol (around 35% more) than low-income households at nearly all quantiles of the distribution. However, drinks per adult are roughly constant across quantiles, so extra purchases are mostly explained by larger household size. Second, distilled spirits comprise a higher share of drinks for higher-income households, and are taxed at a higher rate *per unit of ethanol* (or standard drink) than beer.[16]

Cigarette taxes, on the other hand, appear regressive in panel B of Figure 3, though the overall correlation between cigarette taxes and income is still quite weak, $\rho \approx -0.06$. Regressivity is driven largely by the extensive margin. At lower levels of income, more than 10% of households smoke, while this is not the case at higher income levels. The lowest-income households in the top 5% of the cigarette tax distribution spend more than $300 per year on these taxes, while the top 1% of households spend more than $1000 per year. This is consistent with the fact from Figure 1 that the top 1.5% of households account for 60% of all cigarette taxes. These heaviest smokers drive both the average burden and the overall regressivity of the tax.

Panel E of Figure 3 examines distributional impacts of a hypothetical SSB tax. Consistent with previous work (Allcott et al., 2019a), taxes on SSBs would be moderately regressive, though the overall correlation between household income and SSB consumption is only $\rho \approx -0.05$. The main distinction is that higher-income households purchase fewer SSBs at all quantiles of the distribution, but the purchase-income relationship widens in the extreme quantiles. For example, the median SSB purchaser would pay around $15 per year at the low end of the income distribution and about $13 per year at the high end. At the 75th percentile, this would widen to $42 at the low end and $32 at the high end; at the 90th percentile, this would be $90 and $66, respectively. The gap continues to widen into the extreme quantiles with the top 5% of households paying $130 per year in SSB taxes at the

[16]We provide details on the tax rates in Appendix B.1 and of weekly drinks in Figure A1.

low end of the income distribution and $95 per year at the high end.

Panels C and D of Figure 3 examine the combined burden of sin taxes both with and without a hypothetical penny-per-ounce tax on SSBs. At the 90th percentile of the sin-tax distribution (by income), lower-income households spend around $150 per year while high-income households spend $69 per year. The gap widens at the top 5%, where low-income households spend $340 per year, while high-income households spend $136 per year. The overall regressivity is driven by three factors: (1) smoking behavior is decreasing with income; (2) a subset of mostly lower-income households purchase both alcohol and cigarettes; and (3) while alcohol taxes are slightly progressive, the very highest quantiles in panel (a) are relatively flat.

Perhaps more striking is the broadening of the overall base in panels D and E. Under the existing sin-tax structure, 75% of households pay around $25 or less per year in sin taxes (across all income levels). The addition of an SSB tax would increase the average amount of sin taxes by around $30 per household per year, and half of households would pay more than $50 in overall sin taxes. Households at the 75th percentile of the sin-tax distribution would see their burdens rise $40-$60; households at the 90th and 95th percentile would see their burdens rise $60-$80. In both cases, the poorest households would be towards the top of this range, and the richest households towards the bottom. Across quantiles, the addition of the SSB tax would make the overall scheme of sin taxes somewhat more regressive.

3.3. 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, total ethanol, cigarettes, and SSBs) using k-means clustering.[17] We express the purchases

[17]Recent work by Bonhomme et al. (2021) suggests that even when heterogeneity is not discrete, approximations by k-means can still be effective in a variety of settings. This approach is increasingly common in matched worker-firm data and *grouped fixed effects* (Bonhomme and Manresa, 2015; Bonhomme et al., 2019) and in market definition Zheng (2016).

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. After inspecting the purchasing patterns of each cluster, we assign it a name for expositional purposes. This allows us to categorize households by the products they actually purchase, rather than merely demographics or location.

To validate our cluster assignments, we perform (but do not report) the following predictive regression for annual sin-tax expenditures by household i:

$$\log(1 + tax\ burden_i) = \beta X_i + \gamma_{k_i} + \lambda_{s_i} + \varepsilon_i \tag{2}$$

where γ_{k_i} are fixed effects for our cluster assignments, λ_{s_i} are state fixed effects, and X_i are the mutually exclusive binned demographics described in Section 2 such as *Edu: Some College* or *Income:* $> \$100k$. We find that state fixed-effects (which explain 100% of variation in tax rates) and demographics explain only about 3.5% of the variation in sin taxes paid by households. We find that including our $k = 8$ cluster assignment dummies increases this to 80% of the variation in sin taxes. Conditional on cluster assignment, we find that demographics provide little additional explanatory power (less than 1% of variation). Our cluster assignments are, encouragingly, picking up the relevant variation in the data, though since

[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 $\log()$ 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.

Figure 3 suggests focusing on the conditional mean (as linear regression does) is probably misleading, we should take these comparisons with a grain of salt.[19]

3.4. Results

We describe the results of our cluster assignments in two tables. In Table 1 we provide some basic summary statistics for each cluster (and for the overall sample). In Table 2 we describe 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. Table 2 reports the baseline probability of each demographic bin. For example, 27.4% of the sample completed high school or less. We then 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 $\frac{35.6}{27.4} = 1.30$. We resample the population of households many times, recompute each ratio, and highlight cells where the bootstrapped ratio varies from 1.00 by more than 10%. This gives us an easy way to understand which demographic groups are under- or over-represented within each cluster.[20]

For each cluster, Table 1 reports the 50th, 75th, and 95th percentile of the distribution of purchases (in liters, except for cigarettes which are measured in packs). The two largest clusters, which we label *Nothing* and *SSB Only*, comprise around 20% and 42% of the population respectively and purchase nearly no alcohol or cigarettes. Households in the *SSB Only* cluster purchase 116 liters of SSBs on average per year (around 1L of sugary soda per person per week). Even within the cluster, purchases of SSBs are not homogeneous, as evidenced by the reported quantiles. We also see the base broadening effect: the 42% of the

[19]For example estimated coefficients for (2) are monotone in income, suggesting sin taxes are in fact progressive. This is simply an artifact of the extreme heteroskedasticity and skewness in the data and the $\log(x)$ transformation. The same regression run in levels suggests regressivity. See Appendix C for more details.

[20]As an alternative, in Appendix C 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.

population classified as *SSB Only* currently pay negligible amounts of sin taxes but would pay $22.10 per adult under our hypothetical (penny-per-ounce) SSB tax. Table 2 gives us some insights into who these households are. We see the *SSB Only* group looks mostly like the overall population, except they are 20% more likely to have children and 24% more likely to be Black. The roughly 20% of the population we classify as *Nothing* are substantially less likely to be Black or have children at home, and more likely to be Asian, older, and have post-graduate degrees.

The most heavily-taxed clusters in our sample are the *Everything* cluster, which comprises only 2.5% of the population yet pays 27% of all existing sin taxes (around $288 per adult per year); and *Smokers*, who comprise 5.5% of the population and pay 41% of all existing sin taxes (around $211 per adult per year). The groups are quite similar to one another, except that the *Everything* group purchases a large amount of alcoholic beverages (primarily beer and spirits for 9.4L of ethanol per adult per year or 10.2 standard drinks per week), while the *Smokers* purchase almost none. Both groups purchase more sugary beverages than any other cluster (157-170L per household per year or about 1.5L per person per week). Both groups also purchase about a half pack of cigarettes per day, but with substantial dispersion across households (the top 20% of households in these groups purchase at least one pack per day). Demographically, these two groups look similar to one another: they are older, lower-income, lower-education, less likely to have children or belong to racial or ethnic minority groups. They are most likely to be between the ages of 55-64 and least likely to be below age 35.[21] Because many of these households are in the lowest-income bin ($< $25,000 per year), their overall sin-tax burden as a share of income can be significant. It averages 2% of income for the *Everything* group and 1.5% of income for the *Smokers* group. The hypothetical SSB

[21] One possible and sad explanation for the lower prevalence of 65+ households in these clusters is that the life expectancy for this group is low relative to the overall population. Demographically these clusters are similar to households who are most susceptible to the "Deaths of Despair" described by Case and Deaton (2020).

tax would add another 0.2%-0.3% of income on top of that.[22]

The next most-taxed cluster is the *Heavy Drinkers*. They comprise 6.8% of the population and pay 16% of sin taxes and 43% of alcohol taxes (around $61 per household per year – only about a quarter of what the groups of cigarette smokers pay). They consume slightly more ethanol than the *Everything* cluster (10.3L of ethanol per adult per year, or 11 standard drinks per week). They tend to consume ethanol in substantial amounts from all three categories (beer, wine, and spirits). They tend to drink an average amount of sugar-sweetened beverages (41L per person per year). Demographically, they are less likely to be Black or Asian than the overall population. Otherwise, they look quite different from the *Everything* and *Smokers* clusters. The *Heavy Drinkers* tend to come disproportionately from the higher-income and education groups (the > $100,000 per year group in particular). For this reason, as a share of income, their overall sin tax burden is generally quite modest (< 0.3% of income).

The remaining three clusters consist of moderate drinkers of *Beer* (9.3%), *Wine* (5.8%), and *Spirits* (8.5%). On average they pay less than $23 per year in sin taxes and less than 0.2% of income. The most striking fact is that the wine drinkers tend to purchase much more ethanol than other moderate groups (almost 5L per adult per year instead of around 2L per adult per year) yet face a much lower effective tax rate on the ethanol they purchase ($4.42/L vs $8.86/L for the *Moderate Spirits* cluster). This is due to much lower tax rates per unit of ethanol on wine (particularly in states with many wine drinkers, such as California). *Mostly Wine* drinkers are the most likely to be high education and high income, as well as over 65 years old, as well as not Black. *Moderate Spirits* drinkers also tend to be high income, 45-54 years old, and Black; while *Moderate Beer* drinkers are more likely to be Other Race or Hispanic and to earn between $70,000-$99,999 per year.

[22]There is still a good deal of within-cluster dispersion. The median household in the *Everything* cluster pays around 0.65% of income in sin taxes (0.78% with an SSB tax). These calculations are where assuming 100% incidence on the consumer side is substantive. We provide detailed calculations in Table C4.

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 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 (Figure 3). 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).

By discretizing the heterogeneity in household purchases, we are able to isolate three key clusters most burdened with sin taxes (Table 1). Two of these three clusters, *Everything* (5.5% of the population) and *Smokers* (2.5% of the population), contain essentially all cigarette smokers who bear the lion's share of the total sin tax burden because cigarettes are so highly taxed. Households in the *Everything* cluster also consume substantial amounts of beer and spirits as well, leading them to have the highest tax burden per adult. Households in these two clusters already pay 1.5-2.0% of income on average in sin taxes, and would disproportionately bear any new taxes on sugar-sweetened beverages. They are also much more likely to be from the lowest levels of income and educational attainment, as well as older (Age: 55-64). 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 households facing the most difficult circumstances.

We also identify a cluster of non-smoking *Heavy Drinkers* (6.8% of population), who disproportionately come from the highest income and education groups. If negative externalities (drunk-driving, liver damage, domestic abuse, etc.) are convex in ethanol consumption, then this group (along with *Everything*) should be the source of most 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. The diversity of ethanol sources purchased by these heavy consumers suggests that the U.S. policy of taxing distilled spirits at much higher rates per unit of ethanol than beer and wine may not address the externality as effectively as it would in jurisdictions where heavy consumers focus on one source (see Griffith et al. (2019) for the contrasting UK experience).

Our most important takeaway 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 distribution and miss the extreme heterogenity in the distributional impact of sin tax policy.

Exhibits

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing	Overall
Beer 50%	43.24	0.00	39.90	4.20	0.65	26.10	0.00	0.00	0.00
Beer 75%	172.12	0.00	104.38	14.70	6.30	59.17	0.00	0.00	6.30
Beer 95%	552.30	11.34	420.00	69.06	18.20	264.50	4.20	10.06	83.54
Beer mean	141.79	1.96	99.68	17.96	4.23	67.65	0.62	1.69	18.78
Wine 50%	2.25	0.00	14.25	25.25	2.00	0.75	0.00	0.00	0.00
Wine 75%	12.00	0.00	35.50	51.75	5.25	3.00	0.75	0.75	3.00
Wine 95%	114.75	3.75	146.25	160.50	12.63	8.00	4.50	5.25	29.00
Wine mean	21.41	0.70	35.72	49.76	3.52	2.01	0.84	0.93	6.91
Spirits 50%	8.25	0.00	17.56	0.75	6.71	0.00	0.00	0.00	0.00
Spirits 75%	25.46	0.38	36.39	2.80	11.56	1.75	0.00	0.00	1.78
Spirits 95%	105.82	4.30	93.53	8.68	31.88	5.25	1.75	2.49	19.25
Spirits mean	23.70	0.76	29.77	2.09	10.79	1.25	0.22	0.43	3.99
Cigarettes 50%	75.00	60.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cigarettes 75%	223.00	178.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cigarettes 95%	530.00	460.00	2.00	1.00	2.00	2.00	0.20	0.00	39.00
Cigarettes mean	150.13	131.58	0.26	0.35	0.27	0.31	0.11	0.13	11.21
SSB 50%	89.00	96.07	52.54	18.65	69.22	58.80	68.76	3.44	42.88
SSB 75%	212.71	237.69	126.70	47.08	147.04	136.36	143.40	7.79	114.14
SSB 95%	573.47	557.82	330.14	149.37	350.42	362.03	368.01	13.00	340.01
SSB mean	157.63	169.88	95.03	40.74	112.49	105.46	115.78	4.50	90.84
Ethanol(L) p.a.	9.41	0.20	10.28	4.93	1.83	2.16	0.10	0.20	1.67
Cigarette packs p.a.	89.38	81.96	0.14	0.21	0.16	0.18	0.06	0.10	6.87
SSBs(L) p.p.	72.21	77.93	40.97	19.47	47.31	44.67	53.38	2.83	41.09
Total tax share	0.27	0.41	0.16	0.04	0.05	0.04	0.01	0.01	1.00
Alcohol tax share	0.15	0.01	0.43	0.12	0.13	0.11	0.03	0.02	1.00
Cigarette tax share	0.34	0.65	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SSB tax share	0.04	0.10	0.07	0.03	0.10	0.11	0.53	0.01	1.00
Total tax p.a.	288.40	211.21	61.27	22.44	16.53	12.35	0.77	1.58	27.58
SSB tax p.a.	28.52	32.25	16.53	7.83	19.96	18.81	22.10	1.05	16.97
Effective Ethanol Tax Rate $/L	6.17	7.53	5.89	4.42	8.86	5.55	5.95	6.13	5.88
Combined Taxes / Income (%)	2.24	1.85	0.30	0.11	0.19	0.20	0.15	0.01	0.17
Number of households	1407	2824	4454	4189	5015	5649	25125	12702	61365
Weighted share (%)	2.50	5.50	6.80	5.80	8.50	9.30	41.80	19.70	100.00

Table 1: Annual Household Purchases by Cluster Assignment

Source: Nielsen 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 (20 cigarettes).

p.a. (per adult over 18); p.p. (per person including children).

Combined Taxes include hypothetical penny-per-ounce SSB tax. Reported as percentage of income (averaged over households).

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing
Baseline probability	0.025	0.055	0.068	0.058	0.085	0.093	0.418	0.197
Race: White (74.8%)	1.05	1.10	1.07	1.08	0.93	1.01	0.95	1.04
Race: Black (12.5%)	0.98	0.85	0.72	0.71	1.33	0.83	1.24	0.66
Race: Asian (4.5%)	0.48	0.43	0.66	0.78	0.83	0.79	0.98	1.62
Race: Other (8.2%)	0.91	0.63	1.01	0.79	1.23	1.29	1.06	0.81
Hispanic: Yes (13.2%)	0.77	0.64	0.96	0.98	1.10	1.48	1.05	0.78
Hispanic: No (86.8%)	1.03	1.06	1.01	1.00	0.98	0.93	0.99	1.03
Children: Yes (31.3%)	0.72	0.94	0.85	0.77	1.17	1.15	1.20	0.61
Children: No (68.7%)	1.13	1.03	1.07	1.10	0.92	0.93	0.91	1.18
Age: $<$ 35 (12.9%)	0.58	0.65	0.82	0.89	1.10	1.14	1.03	1.08
Age: 35 to 44 (18.0%)	0.78	0.92	0.85	0.85	1.13	1.10	1.12	0.79
Age: 45 to 54 (21.8%)	1.13	1.14	1.06	0.86	1.18	1.05	1.03	0.80
Age: 55 to 64 (22.7%)	1.38	1.22	1.13	0.99	0.91	0.98	0.94	1.02
Age: $>$ 65 (24.6%)	0.91	0.91	1.04	1.29	0.77	0.82	0.93	1.27
Income: $<$ 24,999 (20.4%)	1.40	1.85	0.41	0.53	0.73	0.74	1.06	1.16
Income: 25,000 - 44,999 (17.7%)	1.20	1.19	0.70	0.74	0.90	0.90	1.09	1.01
Income: 45,000-69,999 (18.2%)	1.00	0.97	0.95	0.92	1.00	1.05	1.04	0.95
Income: 70,000-99,999 (15.5%)	0.92	0.60	1.20	1.13	1.09	1.24	0.98	0.91
Income: $>$ 100,000 (28.1%)	0.63	0.51	1.54	1.48	1.21	1.09	0.89	0.96
Edu: High School or less (27.4%)	1.30	1.61	0.75	0.67	0.84	0.97	1.07	0.91
Edu: Some College (31.4%)	1.29	1.17	0.99	0.83	1.09	1.01	1.02	0.87
Edu: Graduated College (26.3%)	0.73	0.55	1.17	1.20	1.06	1.07	0.96	1.07
Edu: Post College Grad (14.9%)	0.31	0.31	1.18	1.63	0.99	0.91	0.89	1.31

Table 2: Relative Odds 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.

Blue denotes demographic bins with values greater than 1.1 for a 95% (bootstrapped) CI.

Red denotes demographic bins with values less than 0.9 for a 95% (bootstrapped) CI.

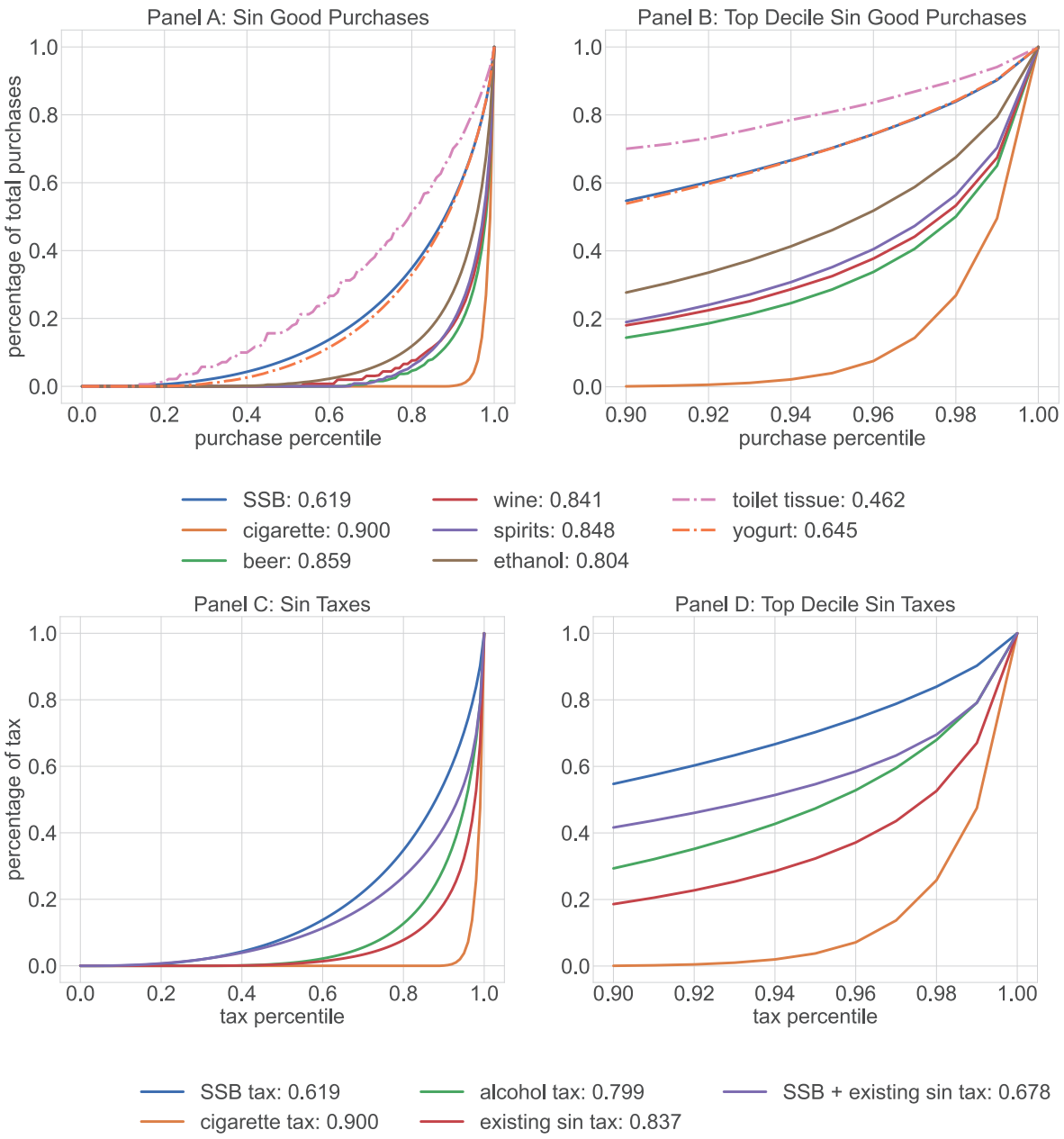

Figure 1: CDFs of Sin Good Purchases and Sin Taxes

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.

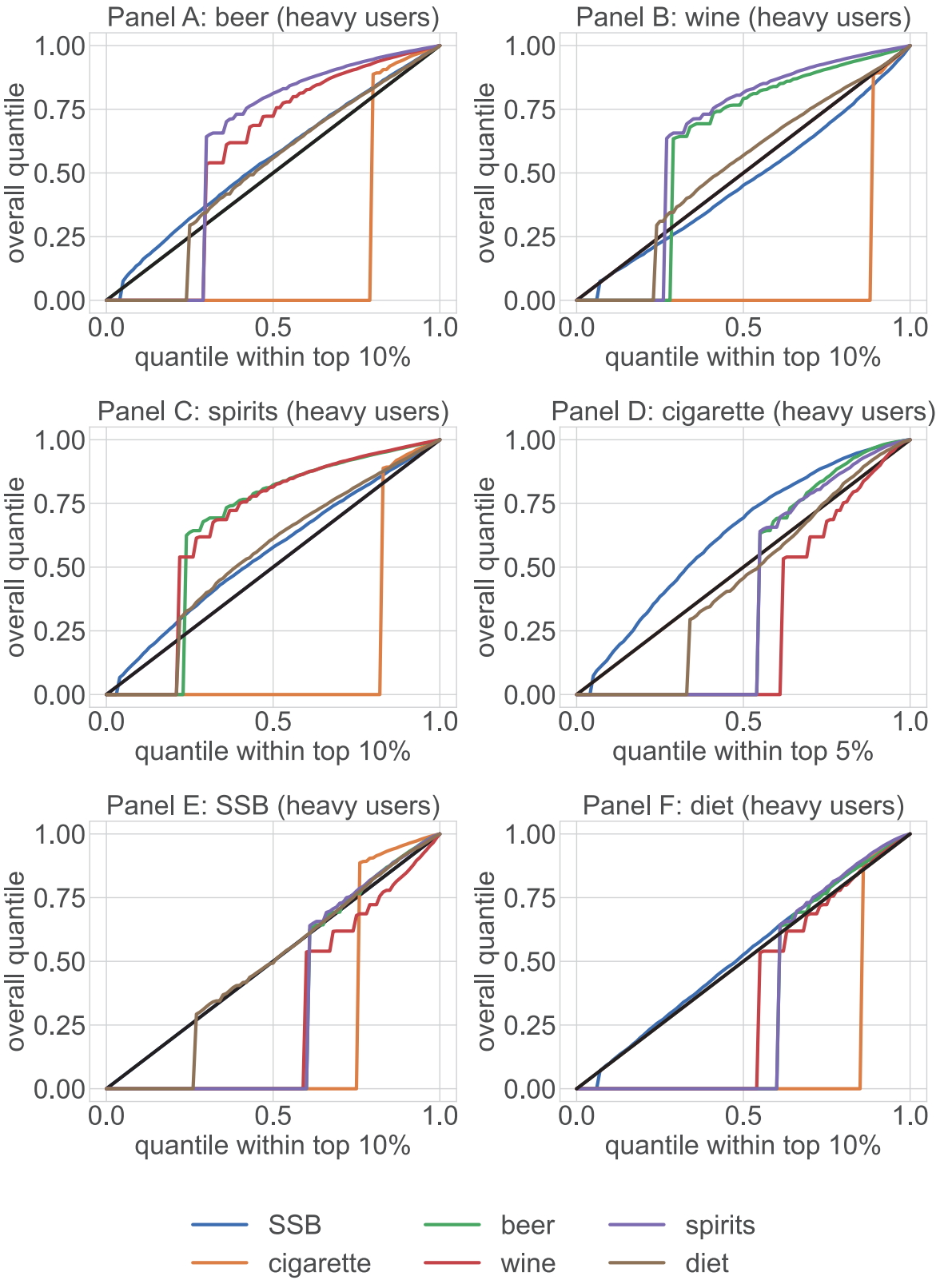

Figure 2: Pairwise Dependence of Purchases by Heavy Users

Note: Each observation is a household and households are ranked by annual purchases. Each panel conditions on the top-decile of purchasing for that category (tobacco conditions on top 5% because fewer than 10% of households purchase cigarettes). The 45 degree line would constitute no pairwise dependence among heavy users of one category and consumption for the second category. Above (below) the 45 degree line would constitute positive (negative) dependence.

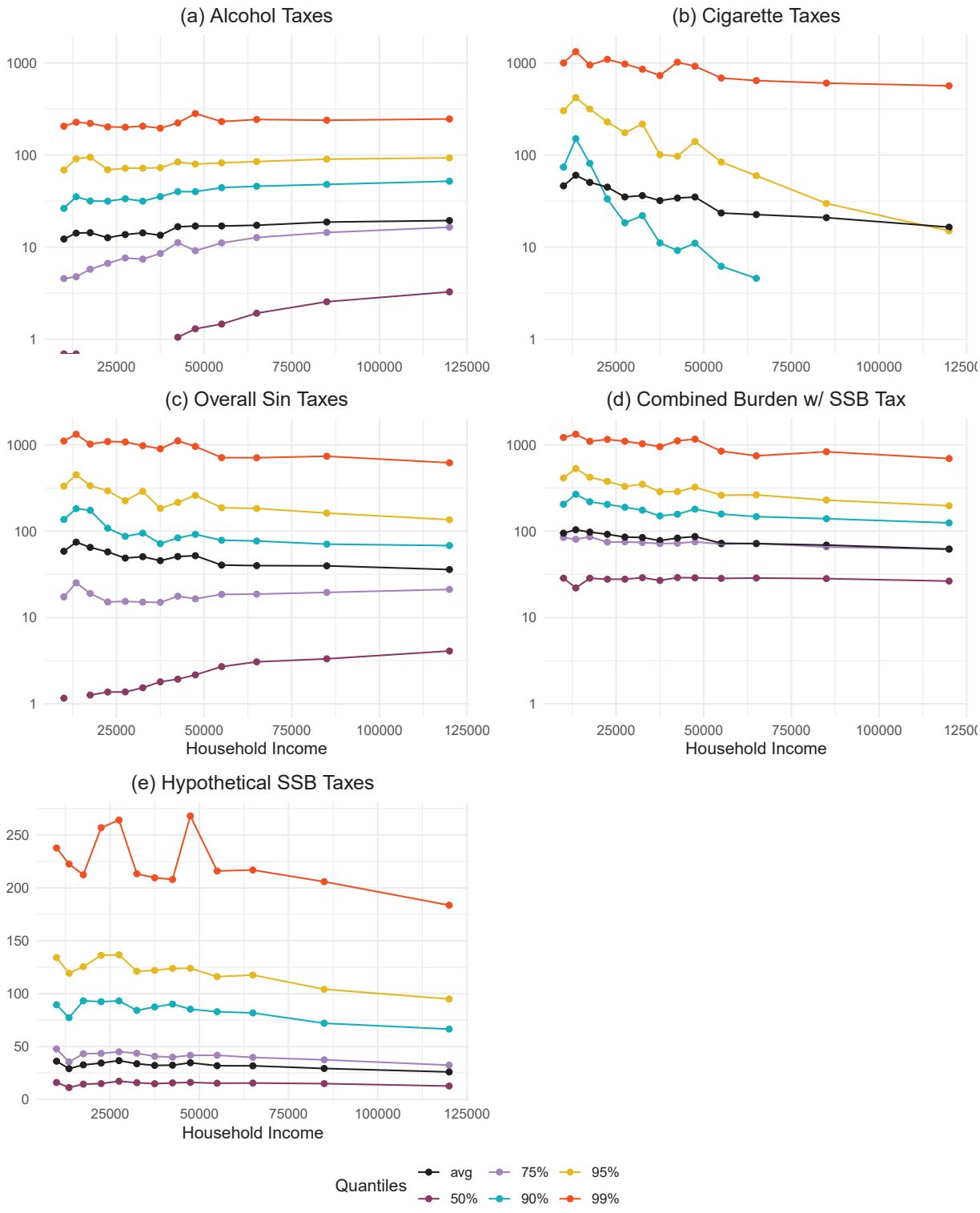

Figure 3: Quantiles of Annual Sin Good Taxes by Income

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

References

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

Adda, J. and F. Cornaglia (2006): "Taxes, Cigarette Consumption, and Smoking Intensity," *American Economic Review*, 96, 1013–1028.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019a): "Regressive Sin Taxes, with an Application to the Optimal Soda Tax," *The Quarterly Journal of Economics*, 134, 1557–1626.

——— (2019b): "Should We Tax Sugar-Sweetened Beverages? An Overview of Theory and Evidence," *Journal of Economic Perspectives*, 33, 202–27.

Bollinger, B. and S. Sexton (2018): "Local Excise Taxes, Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax," Tech. rep., Working Paper.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87, 699–739.

——— (2021): "Discretizing Unobserved Heterogeneity," .

Bonhomme, S. and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83, 1147–1184.

Case, A. and A. Deaton (2020): *Deaths of Despair and the Future of Capitalism*, Princeton University Press.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2021): "On Binscatter," .

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2017): "Bidimensional Matching with Heterogeneous Preferences: Education and Smoking in the Marriage Market," *Journal of the European Economic Association*, 16, 161–198.

Colman, G. J. and D. 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, 376–400.

Conlon, C. T. and N. 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, 111–43.

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

Dubois, P., R. Griffith, and M. O'Connell (2020): "How Well Targeted Are Soda Taxes?" *American Economic Review*, 110, 3661–3704.

Friedson, A. I., M. Li, K. Meckel, D. I. Rees, and D. W. Sacks (2021): "Cigarette Taxes, Smoking, and Health in the Long-Run," Working Paper 29145, National Bureau of Economic Research.

Griffith, R., M. O'Connell, and K. Smith (2019): "Tax design in the alcohol market," *Journal of Public Economics*, 172, 20–35.

Hansen, B., J. J. Sabia, and D. 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, 60–75.

Harding, M., E. Leibtag, and M. F. Lovenheim (2012): "The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data," *American Economic Journal: Economic Policy*, 4, 169–98.

Hendren, N. (2020): "Measuring economic efficiency using inverse-optimum weights," *Journal of Public Economics*, 187, 104198.

Hirono, K. and K. Smith (2017): "Tobacco tax hikes are great, so long as you're not a poor smoker," .

Ibragimov, R. and A. Prokhorov (2017): *Heavy Tails and Copulas*, World Scientific.

Miravete, E. J., K. Seim, and J. Thurk (2018): "Market Power and the Laffer Curve," *Econometrica*, 86, 1651–1687.

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

Naimi, T., J. Daley, Z. Xuan, J. Blanchette, F. Chaloupka, and D. Jernigan (2016): "Who Would Pay for State Alcohol Tax Increases in the United States?" *Preventing Chronic Disease*, 13.

Sanders, B. (2016): "A Soda Tax Would Hurt Philly's Low-Income Families," .

Seiler, S., A. Tuchman, and S. Yao (2021): "The Impact of Soda Taxes: Pass-Through, Tax Avoidance, and Nutritional Effects," *Journal of Marketing Research*, 58, 22–49.

Zheng, F. (2016): "Spatial Competition and Preemptive Entry in the Discount Retail Industry," Working Paper.

A. Appendix 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" as we did for taxes in Figure 3 in the main text. 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 divide by the number of U.S. households in 2018 according to the U.S. Census[23]. As NIAAA

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

	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	0.17	0.31
60%	0.32	0.37	0.69
70%	1.10	0.76	1.42
80%	3.17	1.67	3.09
90%	7.76	4.65	8.45
max	37.49	79.77	158.42

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.

Category	Nielsen Volume	Nielsen Ethanol	NIAAA Volume	NIAAA Ethanol	Ethanol Discrepancy(%)	On-Premise (%)
Beer	50.09	2.25	188.09	8.46	73.40	23.00
Wine	15.24	1.97	26.98	3.48	43.49	18.50
Spirits	10.93	4.49	16.95	6.97	35.58	21.20
Total	76.26	8.71	232.02	18.91	53.94	

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`.

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 73.4% lower than apparent consumption 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

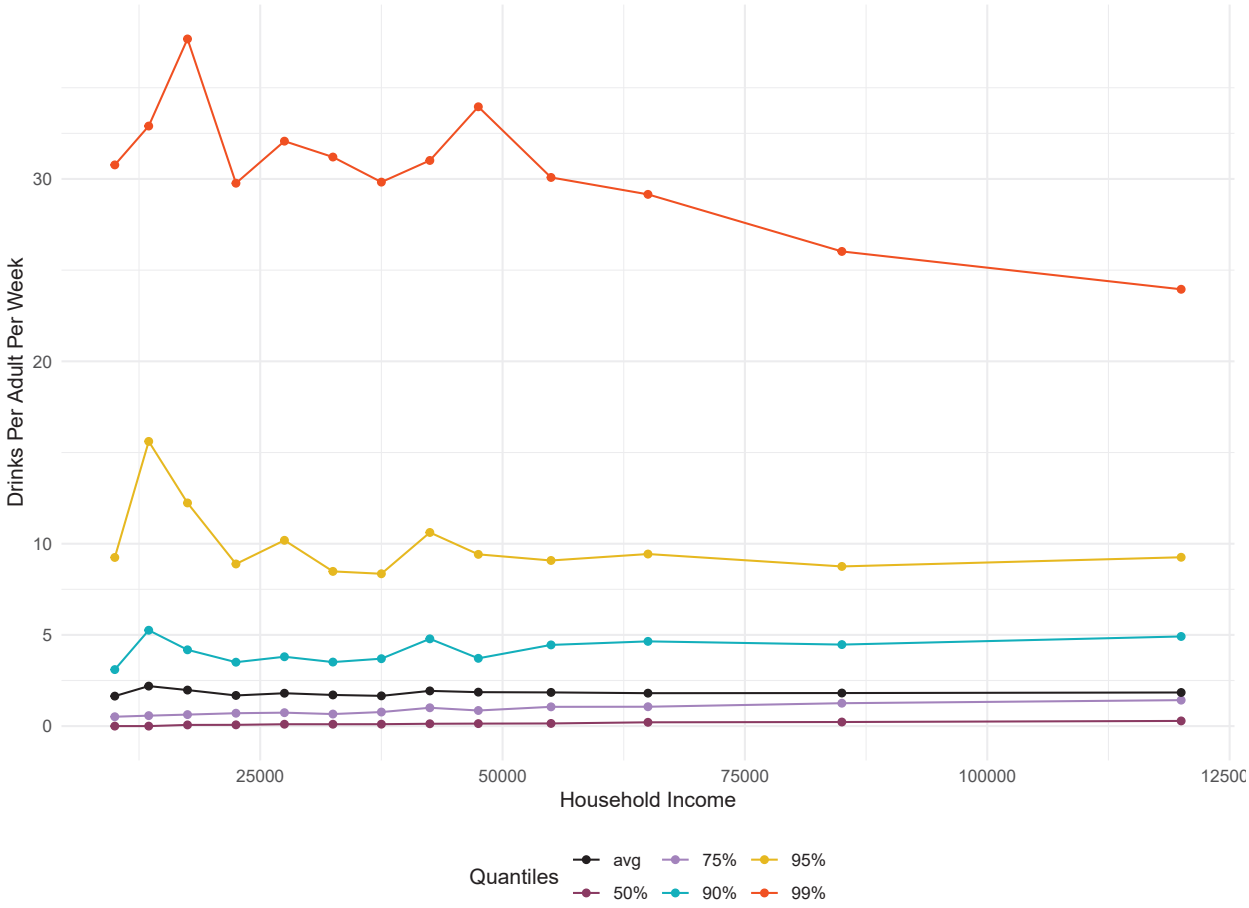

Figure A1: Alcoholic Drinks Per Adult Per Week

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).

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 consump-

tion 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	0.712	1.315
96%	0	1.37	2.521
97%	4	2.712	4.877
98%	10	5.068	8.767
99%	20	9.863	16.359
max	40	49.315	197.260

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. Appendix 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 \$3.26, \$2.57 and \$8.92 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 \$7.23 per ethanol liter while spirits enjoy a relatively low tax \$2.95 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.

B.2. SSB Tax Nielsen Modules

We have to make some decisions as to 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.[24]

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 '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'.

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

State	Tax Rate (per L)				Tax Rate (per Ethanol L)			
	Beer	Wine	Spirits	Cigarette	Beer	Wine	Spirits	Tax/HH
FED	0.15	0.28	3.57	1.01	3.26	2.57	8.92	45.40
AL	0.28	0.45	4.83	0.68	5.90	4.08	12.07	50.91
AZ	0.04	0.22	0.79	2.00	0.90	2.02	1.98	62.75
AR	0.09	0.38	2.04	1.15	1.92	3.46	5.11	60.18
CA	0.05	0.05	0.87	2.87	1.12	0.48	2.18	34.80
CO	0.02	0.08	0.60	0.84	0.45	0.76	1.51	31.45
CT	0.06	0.19	1.43	4.35	1.31	1.73	3.57	68.84
DE	0.07	0.43	1.19	2.10	1.48	3.91	2.97	57.32
FL	0.13	0.59	1.72	1.34	2.70	5.40	4.29	55.28
GA	0.27	0.40	1.00	0.37	5.68	3.64	2.50	34.33
ID	0.04	0.12	2.89	0.57	0.84	1.08	7.23	56.44
IL	0.06	0.37	2.26	1.98	1.30	3.34	5.65	55.43
IN	0.03	0.12	0.71	1.00	0.65	1.13	1.77	40.57
IA	0.05	0.46	3.45	1.36	1.07	4.20	8.63	57.95
KS	0.05	0.08	0.66	1.29	1.01	0.72	1.65	29.13
KY	0.22	0.92	2.08	0.60	4.78	8.33	5.19	45.58
LA	0.11	0.20	0.80	1.08	2.27	1.82	2.00	52.78
ME	0.09	0.16	1.54	2.00	1.97	1.44	3.85	44.74
MD	0.14	0.36	1.32	2.00	2.98	3.27	3.30	46.66
MA	0.03	0.15	1.07	3.51	0.60	1.32	2.67	41.17
MI	0.05	0.13	3.17	2.00	1.14	1.22	7.92	43.65
MN	0.13	0.32	2.37	3.04	2.78	2.88	5.92	34.89
MS	0.11	0.00	2.15	0.68	2.40	0.00	5.38	40.09
MO	0.02	0.11	0.53	0.17	0.34	1.01	1.32	38.48
MT	0.04	0.28	2.58	1.70	0.78	2.55	6.46	63.01
NE	0.08	0.25	0.99	0.64	1.74	2.28	2.48	43.14
NV	0.04	0.18	0.95	1.80	0.90	1.68	2.38	42.37
NH	0.08	0.00	0.00	1.78	1.69	0.00	0.00	68.01
NJ	0.03	0.23	1.45	2.70	0.67	2.10	3.63	33.80
NM	0.11	0.45	1.60	1.66	2.30	4.08	4.00	40.20
NY	0.04	0.08	1.70	4.35	0.79	0.72	4.25	45.14
NC	0.16	0.26	3.86	0.45	3.47	2.39	9.66	43.82
ND	0.10	0.28	1.23	0.44	2.19	2.55	3.08	29.38
OH	0.05	0.08	2.61	1.60	1.01	0.77	6.52	47.48
OK	0.11	0.19	1.47	1.03	2.27	1.73	3.67	40.42
OR	0.02	0.18	6.01	1.33	0.47	1.61	15.02	39.99
PA	0.02	0.00	1.91	2.60	0.45	0.00	4.78	56.42
RI	0.03	0.37	1.43	4.25	0.70	3.36	3.57	32.97
SC	0.20	0.29	1.43	0.57	4.32	2.59	3.58	49.33
SD	0.07	0.35	1.23	1.53	1.54	3.15	3.08	58.03
TN	0.34	0.34	1.18	0.62	7.23	3.06	2.95	43.68
TX	0.05	0.05	0.63	1.41	1.11	0.49	1.59	41.22
UT	0.11	0.00	4.06	1.70	2.32	0.00	10.16	31.61
VT	0.07	0.15	2.04	3.08	1.49	1.32	5.10	54.01
VA	0.07	0.40	5.26	0.30	1.44	3.64	13.16	35.73
WA	0.07	0.23	8.59	3.02	1.47	2.08	21.48	71.64
WV	0.05	0.26	2.03	1.20	1.00	2.40	5.07	65.70
WI	0.02	0.07	0.86	2.52	0.36	0.60	2.15	49.79
WY	0.01	0.00	0.00	0.60	0.11	0.00	0.00	14.95
DC	0.18	0.46	1.63	2.50	3.93	4.15	4.07	56.15

Table B1: Federal and State Tax Rates on Sin Goods

Source: Tax Foundation 2018. `https://taxfoundation.org/facts-figures-2018/`.
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 the Tax Foundation

C. Appendix C: 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 C1, we calculate the average tax paid for different categories of sin goods by household income. We also report the ratio that the top income bin ($> \$100K$) pays relative to the bottom income bin ($< \$25K$). We see that beer taxes are pretty evenly distributed across income groups. The highest income taxes 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 $3\times$ as much as the richest. However, taxes on wine and distilled spirits appear to be strongly progressive with the richest households paying $1.7 - 2\times$ 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**	0.90	2.06	1.69	0.34	0.77	0.58
< $24,999	4.59	2.31	6.24	48.70	33.87	61.84
$25,000 - $44,999	4.26	2.70	7.53	34.44	33.82	48.93
$45,000-$69,999	4.67	3.22	9.20	25.97	32.57	43.06
$70,000-$99,999	4.78	3.75	10.22	20.93	29.32	39.68
> $100,000	4.15	4.77	10.53	16.49	26.04	35.94

Table C1: 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 C1 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 C1 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.[25] If we compare the average sin taxes paid for each income level, sin taxes look highly regressive and the correlation coeffiicent is $\rho = -0.81$. However, if we plot the log of sin taxes paid (plus one dollar) we find that the correlation is strongly positive $\rho = 0.81$.[26] The problem is the extreme heteroskedasticity where the standard deviation of sin taxes paid is more than $210 for lower income households and only $135 for the highest income households, this and the extreme skewness of

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

[26] 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.

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).

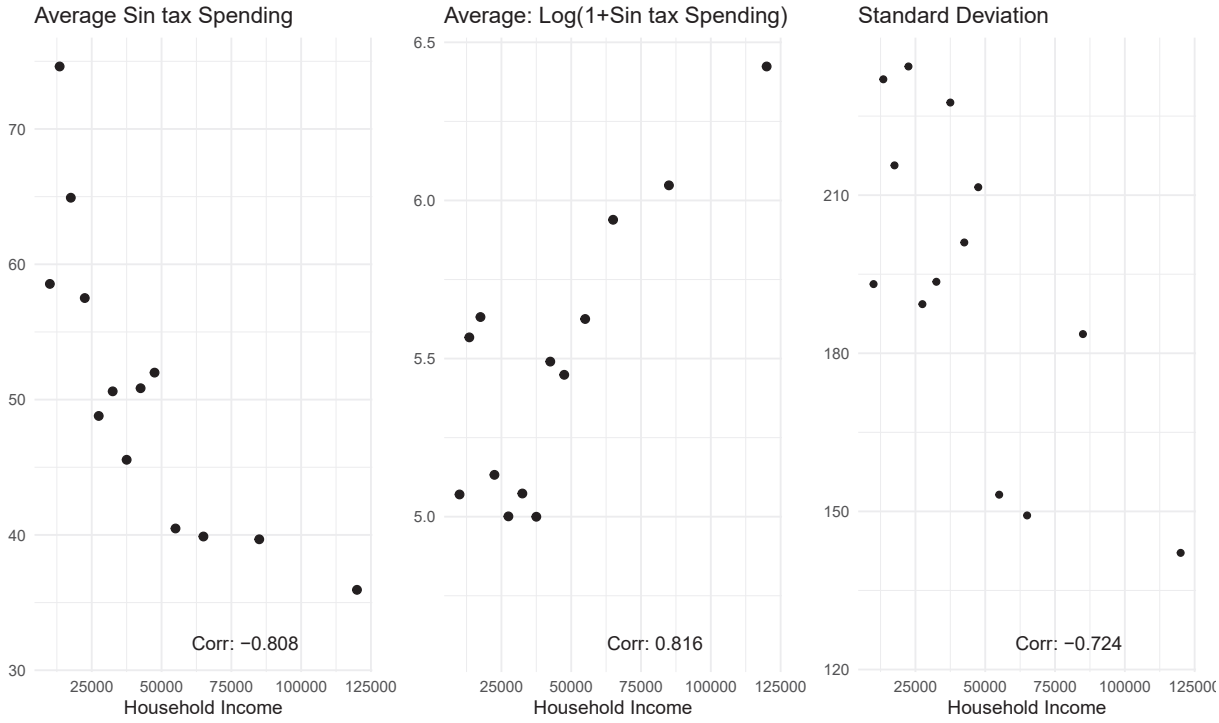

Figure C1: Sin Taxes Paid by Income (Levels vs. Logs)

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

Table C2 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 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 Table C3 we report the results of the regression described in (2). The differences are the cluster fixed effects (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

Panel A: Existing Sin Taxes on Alcohol and Tobacco						
	tax 25%	tax 50%	tax 75%	tax 90%	tax 95%	mean
White	0.0	2.87	21.29	95.46	231.40	50.44
Black	0.0	2.04	14.11	58.92	136.73	31.61
Asian	0.0	1.04	6.95	29.78	69.52	19.96
Other	0.0	2.82	15.61	58.57	123.91	34.34
Panel B: Existing Sin Taxes + SSB Taxes						
	tax 25%	tax 50%	tax 75%	tax 90%	tax 95%	mean
White	8.98	27.75	72.43	167.05	304.39	81.02
Black	12.79	32.35	70.88	140.59	222.22	67.00
Asian	4.10	12.32	35.31	81.14	130.54	38.32
Other	10.86	28.53	63.67	129.00	213.10	65.85

Panel C: Existing Sin Taxes on Alcohol and Tobacco						
	tax 25%	tax 50%	tax 75%	tax 90%	tax 95%	mean
Hispanic	0.0	3.01	14.60	54.40	107.33	27.14
Non-Hispanic	0.0	2.52	19.43	89.26	217.50	48.18
Panel D: Existing Sin Taxes + SSB Taxes						
	tax 25%	tax 50%	tax 75%	tax 90%	tax 95%	mean
Hispanic	11.54	28.55	63.43	124.79	198.49	58.32
Non-Hispanic	8.89	27.35	70.70	160.57	293.54	78.82

Table C2: Sin Tax Paid by Race and Ethnicity

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.

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 C1.

In Table C4 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

	Log(Sin Tax)		Log(Sin Tax+SSB Tax)	
	(1)	(2)	(3)	(4)
Income: < 24,999	-0.152	-0.051	-0.094	-0.021
	(0.041)	(0.032)	(0.025)	(0.038)
Income: 25,000-44,999	-0.137	-0.051	-0.059	-0.013
	(0.031)	(0.018)	(0.026)	(0.025)
Income: 70,000-99,999	0.089	0.017	0.025	-0.015
	(0.035)	(0.018)	(0.029)	(0.011)
Income: > 100,000	0.217	0.010	0.084	-0.026
	(0.028)	(0.032)	(0.019)	(0.022)
Race: Black	-0.206	-0.047	0.109	0.024
	(0.045)	(0.040)	(0.036)	(0.018)
Race: Asian	-0.534	-0.105	-0.514	-0.173
	(0.029)	(0.017)	(0.032)	(0.025)
Race: Other	-0.049	0.005	0.021	-0.011
	(0.034)	(0.010)	(0.040)	(0.009)
Hispanic: Yes	0.076	0.041	-0.005	0.017
	(0.040)	(0.032)	(0.043)	(0.017)
Children: Yes	-0.150	-0.055	0.315	0.134
	(0.031)	(0.037)	(0.022)	(0.042)
Edu: High School or Less	0.151	0.014	0.371	0.149
	(0.034)	(0.025)	(0.022)	(0.047)
Edu: Some College	0.174	0.037	0.250	0.078
	(0.036)	(0.010)	(0.021)	(0.014)
Edu: Post College Grad	-0.133	-0.012	-0.276	-0.107
	(0.032)	(0.032)	(0.020)	(0.025)
Age: Under 35	0.016	0.002	-0.021	-0.043
	(0.041)	(0.098)	(0.030)	(0.084)
Age: 35-44	0.071	-0.007	0.137	0.014
	(0.041)	(0.083)	(0.028)	(0.063)
Age: 45-54	0.199	0.018	0.289	0.081
	(0.038)	(0.059)	(0.029)	(0.052)
Age: 55-64	0.206	0.040	0.232	0.069
	(0.035)	(0.025)	(0.027)	(0.019)
Standard-Errors	State	State & Cluster Assignment	State	State & Cluster Assignment
Observations	61,365	61,365	61,365	61,365
R^2	0.0359	0.8042	0.0579	0.6884
Within R^2	0.0168	0.0044	0.0433	0.0191
State FE	✓	✓	✓	✓
Cluster Assignment FE		✓		✓

Table C3: 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.

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 C5 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 C5 and vice versa, so this is purely for convenience.

As an alternative to the odds ratios 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 C7 we provide the bootstrapped confidence intervals for the odds ratio 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 $N = 61{,}365$ households with replacement.
3. Compute $Pr(h \in Demog | h \in Cluster)$ as in Table C5
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.

Panel A: Existing Sin Taxes on Alcohol and Tobacco

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing
min	0.017	0.005	0.011	0.002	0.003	0.001	0.000	0.000
ratio 25%	0.252	0.125	0.043	0.012	0.013	0.007	0.000	0.000
ratio 50%	0.647	0.428	0.089	0.027	0.026	0.017	0.000	0.000
ratio 75%	1.787	1.510	0.205	0.062	0.063	0.044	0.003	0.003
ratio 90%	4.208	3.569	0.480	0.163	0.149	0.118	0.009	0.011
ratio 95%	7.288	5.711	0.829	0.290	0.268	0.242	0.018	0.024
mean	2.032	1.561	0.231	0.083	0.077	0.062	0.004	0.006
max	78.584	67.099	12.965	5.118	5.180	3.896	0.572	4.397

Panel B: Existing Sin Taxes Plus New Taxes on SSBs

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing
min	0.021	0.013	0.015	0.004	0.006	0.004	0.003	0.000
ratio 25%	0.334	0.232	0.066	0.021	0.033	0.025	0.020	0.001
ratio 50%	0.782	0.668	0.120	0.042	0.070	0.057	0.046	0.004
ratio 75%	2.086	1.848	0.267	0.091	0.169	0.141	0.120	0.010
ratio 90%	4.402	4.392	0.610	0.218	0.379	0.354	0.309	0.023
ratio 95%	7.671	6.542	1.014	0.367	0.665	0.698	0.531	0.039
mean	2.238	1.846	0.303	0.112	0.189	0.200	0.148	0.011
max	82.854	70.009	15.673	7.851	9.691	55.203	10.731	4.397

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

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing
min	0.021	0.009	0.011	0.002	0.003	0.001	0.000	0.000
ratio 25%	0.329	0.176	0.043	0.013	0.013	0.007	0.000	0.000
ratio 50%	0.848	0.618	0.089	0.027	0.026	0.017	0.000	0.000
ratio 75%	2.415	2.116	0.205	0.063	0.064	0.045	0.003	0.003
ratio 90%	5.573	4.974	0.482	0.163	0.152	0.123	0.010	0.011
ratio 95%	9.341	8.406	0.834	0.292	0.276	0.258	0.020	0.026
mean	2.693	2.171	0.231	0.084	0.079	0.063	0.005	0.007
max	113.551	83.743	12.965	5.118	5.340	3.896	0.692	5.637

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

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing
min	0.024	0.017	0.015	0.004	0.006	0.004	0.003	0.000
ratio 25%	0.413	0.299	0.066	0.021	0.034	0.025	0.020	0.001
ratio 50%	0.975	0.842	0.121	0.042	0.070	0.057	0.046	0.004
ratio 75%	2.637	2.412	0.267	0.093	0.169	0.143	0.120	0.010
ratio 90%	5.686	5.972	0.613	0.219	0.380	0.356	0.311	0.023
ratio 95%	9.862	8.838	1.019	0.367	0.665	0.698	0.533	0.040
mean	2.899	2.456	0.304	0.113	0.190	0.201	0.149	0.012
max	117.821	86.652	15.673	7.851	9.831	55.203	10.731	5.637

Table C4: Distribution of Sin Tax to Income Ratios (%)

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing
Race: White (74.8%)	0.78	0.82	0.80	0.81	0.70	0.76	0.71	0.78
Race: Black (12.5%)	0.12	0.11	0.09	0.09	0.17	0.10	0.16	0.08
Race: Asian (4.5%)	0.02	0.02	0.03	0.03	0.04	0.04	0.04	0.07
Race: Other (8.2%)	0.07	0.05	0.08	0.06	0.10	0.11	0.09	0.07
Hispanic: Yes (13.2%)	0.10	0.08	0.13	0.13	0.15	0.20	0.14	0.10
Hispanic: No (86.8%)	0.90	0.92	0.87	0.87	0.85	0.80	0.86	0.90
Children: Yes (31.3%)	0.22	0.29	0.27	0.24	0.36	0.36	0.38	0.19
Children: No (68.7%)	0.78	0.71	0.73	0.76	0.64	0.64	0.62	0.81
Age: $¡$ 35 (12.9%)	0.07	0.08	0.11	0.12	0.14	0.15	0.13	0.14
Age: 35 to 44 (18.0%)	0.14	0.16	0.15	0.15	0.20	0.20	0.20	0.14
Age: 45 to 54 (21.8%)	0.25	0.25	0.23	0.19	0.26	0.23	0.22	0.18
Age: 55 to 64 (22.7%)	0.31	0.28	0.26	0.22	0.21	0.22	0.21	0.23
Age: $¿$ 65 (24.6%)	0.22	0.22	0.26	0.32	0.19	0.20	0.23	0.31
Income: $¡$ 24,999 (20.4%)	0.28	0.38	0.08	0.11	0.15	0.15	0.22	0.24
Income: 25,000 - 44,999 (17.7%)	0.21	0.21	0.12	0.13	0.16	0.16	0.19	0.18
Income: 45,000-69,999 (18.2%)	0.18	0.18	0.17	0.17	0.18	0.19	0.19	0.17
Income: 70,000-99,999 (15.5%)	0.14	0.09	0.19	0.17	0.17	0.19	0.15	0.14
Income: $¿$ 100,000 (28.1%)	0.18	0.14	0.43	0.42	0.34	0.31	0.25	0.27
Edu: High School or less (27.4%)	0.36	0.44	0.21	0.18	0.23	0.27	0.29	0.25
Edu: Some College (31.4%)	0.41	0.37	0.31	0.26	0.34	0.32	0.32	0.27
Edu: Graduated College (26.3%)	0.19	0.15	0.31	0.32	0.28	0.28	0.25	0.28
Edu: Post College Grad (14.9%)	0.05	0.05	0.18	0.24	0.15	0.14	0.13	0.19

Table C5: Share of households with certain demographics by cluster: $Pr(h \in Demog | h \in Cluster)$

	cluster = Everything		**cluster = Smokers**		**cluster = Heavy Drinkers**		**cluster = Mostly Wine**	
	coef	std err	coef	std err	coef	std err	coef	std err
Intercept	-2.4394	0.085	-1.8462	0.064	-0.7625	0.050	-0.6548	0.050
Edu: High School or Less	0.7638	0.081	0.9361	0.060	0.0892	0.058	-0.2914	0.064
Edu: Post College Grad	-0.8766	0.107	-0.7793	0.081	-0.4370	0.048	-0.0802	0.046
Edu: Some College	0.6310	0.071	0.6025	0.055	0.1181	0.046	-0.1060	0.049
Race: Black	0.4823	0.091	0.2362	0.074	0.0461	0.066	-0.0151	0.069
Race: Asian	-0.9860	0.209	-0.8907	0.147	-1.2347	0.112	-0.9250	0.101
Race: Other	0.3257	0.132	-0.0719	0.111	-0.0337	0.092	-0.0183	0.094
Hispanic: Yes	-0.1064	0.131	-0.1226	0.10	0.1103	0.079	0.1916	0.080
Age: 35 - 44	0.3475	0.108	0.3439	0.079	-0.2503	0.067	-0.4466	0.068
Age: 45 - 54	0.6847	0.084	0.6304	0.063	-0.0866	0.055	-0.4232	0.057
Age: 55 - 64	0.5159	0.072	0.4612	0.054	-0.0236	0.044	-0.3473	0.046
Age: Under 35	-0.2821	0.158	-0.3031	0.111	-0.4420	0.084	-0.6247	0.084
Children: Yes	0.2122	0.086	0.6483	0.061	0.2689	0.054	0.1887	0.057
Income: $<$ 24,999	-0.3902	0.087	0.0218	0.061	-1.2901	0.077	-0.9169	0.074
Income: 25,000 - 44,999	-0.3732	0.079	-0.2322	0.059	-0.6267	0.057	-0.4343	0.057
Income: 70,000 - 99,999	-0.2951	0.085	-0.6053	0.070	0.1327	0.052	0.1228	0.054
Income: $>$ 100,000	-0.4380	0.097	-0.7002	0.079	0.4606	0.052	0.3310	0.054
	cluster = Moderate Spirits		**cluster = Moderate Beer**		**cluster = SSB only**			
	coef	std err	coef	std err	coef	std err		
Intercept	-1.0969	0.050	-0.8837	0.048	0.5317	0.033		
Edu: High School or Less	0.2738	0.054	0.4752	0.050	0.4297	0.035		
Edu: Post College Grad	-0.4659	0.047	-0.5286	0.046	-0.3451	0.031		
Edu: Some College	0.2361	0.043	0.2611	0.042	0.2132	0.030		
Race: Black	0.7362	0.054	0.2011	0.059	0.6733	0.040		
Race: Asian	-0.7100	0.089	-0.7432	0.086	-0.4352	0.052		
Race: Other	0.2056	0.082	0.0223	0.078	0.1748	0.057		
Hispanic: Yes	0.1487	0.072	0.5365	0.064	0.0974	0.051		
Age: 35 - 44	0.3596	0.061	0.2173	0.059	0.1219	0.041		
Age: 45 - 54	0.5136	0.052	0.2647	0.051	0.2409	0.035		
Age: 55 - 64	0.2230	0.046	0.1701	0.043	0.0974	0.029		
Age: Under 35	0.2604	0.072	0.1316	0.069	-0.1107	0.048		
Children: Yes	0.5921	0.047	0.6605	0.046	0.8906	0.034		
Income: $<$ 24,999	-0.8093	0.062	-0.9479	0.061	-0.4070	0.038		
Income: 25,000 - 44,999	-0.5254	0.052	-0.5785	0.050	-0.2757	0.034		
Income: 70,000 - 99,999	-0.0671	0.050	0.0417	0.047	-0.1521	0.035		
Income: $>$ 100,000	0.1129	0.051	0.0841	0.050	-0.2053	0.036		

Table C6: Multinomial Logit Regression

	Everything	Smokers	Heavy Drinkers	Mostly Wine	Moderate Spirits	Moderate Beer	SSB only	Nothing
Race: White (74.8 %)	1.05	1.10	1.07	1.08	0.93	1.01	0.95	1.04
	(1.01, 1.09)	(1.07, 1.13)	(1.04, 1.09)	(1.06, 1.11)	(0.91, 0.95)	(0.99, 1.03)	(0.95, 0.96)	(1.03, 1.05)
Race: Black (12.5 %)	0.98	0.85	0.72	0.71	1.33	0.83	1.24	0.66
	(0.79, 1.16)	(0.72, 0.98)	(0.63, 0.82)	(0.6, 0.82)	(1.21, 1.43)	(0.74, 0.91)	(1.2, 1.28)	(0.61, 0.71)
Race: Asian (4.5 %)	0.49	0.43	0.66	0.78	0.83	0.79	0.98	1.62
	(0.26, 0.76)	(0.27, 0.6)	(0.49, 0.84)	(0.6, 1.0)	(0.67, 1.0)	(0.64, 0.96)	(0.92, 1.05)	(1.49, 1.75)
Race: Other (8.2 %)	0.90	0.64	1.01	0.79	1.23	1.28	1.06	0.81
	(0.66, 1.15)	(0.5, 0.79)	(0.86, 1.17)	(0.64, 0.94)	(1.06, 1.4)	(1.14, 1.45)	(1.01, 1.11)	(0.73, 0.89)
Hispanic: Yes (13.2 %)	0.77	0.64	0.96	0.97	1.10	1.48	1.05	0.78
	(0.58, 0.97)	(0.52, 0.76)	(0.84, 1.07)	(0.82, 1.11)	(0.99, 1.22)	(1.36, 1.6)	(1.01, 1.09)	(0.72, 0.84)
Hispanic: No (86.8 %)	1.03	1.05	1.01	1.00	0.98	0.93	0.99	1.03
	(1.0, 1.06)	(1.04, 1.07)	(0.99, 1.02)	(0.98, 1.03)	(0.97, 1.0)	(0.91, 0.94)	(0.99, 1.0)	(1.02, 1.04)
Children: Yes (31.3 %)	0.72	0.94	0.85	0.77	1.17	1.15	1.20	0.61
	(0.61, 0.81)	(0.86, 1.01)	(0.79, 0.92)	(0.71, 0.84)	(1.11, 1.22)	(1.09, 1.2)	(1.18, 1.22)	(0.58, 0.64)
Children: No (68.7 %)	1.13	1.03	1.07	1.10	0.92	0.93	0.91	1.18
	(1.08, 1.17)	(1.0, 1.06)	(1.04, 1.09)	(1.07, 1.13)	(0.9, 0.95)	(0.91, 0.96)	(0.9, 0.92)	(1.16, 1.19)
Age: $<$ 35 (12.9 %)	0.58	0.65	0.82	0.89	1.10	1.14	1.03	1.08
	(0.42, 0.75)	(0.53, 0.78)	(0.7, 0.94)	(0.77, 1.02)	(0.99, 1.2)	(1.03, 1.25)	(0.99, 1.07)	(1.01, 1.15)
Age: 35 to 44 (18.0 %)	0.77	0.91	0.85	0.85	1.13	1.10	1.12	0.79
	(0.65, 0.92)	(0.81, 1.01)	(0.76, 0.93)	(0.75, 0.94)	(1.05, 1.23)	(1.03, 1.19)	(1.09, 1.15)	(0.74, 0.84)
Age: 45 to 54 (21.8 %)	1.12	1.14	1.06	0.86	1.19	1.05	1.03	0.80
	(0.98, 1.25)	(1.05, 1.23)	(0.98, 1.15)	(0.79, 0.94)	(1.12, 1.26)	(0.99, 1.11)	(1.0, 1.05)	(0.76, 0.84)
Age: 55 to 64 (22.7 %)	1.38	1.23	1.13	0.99	0.91	0.98	0.94	1.02
	(1.25, 1.51)	(1.14, 1.32)	(1.05, 1.21)	(0.91, 1.06)	(0.85, 0.97)	(0.92, 1.04)	(0.92, 0.96)	(0.98, 1.06)
Age: $>$ 65 (24.6 %)	0.92	0.91	1.04	1.30	0.77	0.83	0.93	1.27
	(0.8, 1.04)	(0.84, 0.99)	(0.96, 1.1)	(1.22, 1.37)	(0.71, 0.82)	(0.77, 0.88)	(0.91, 0.95)	(1.23, 1.31)
Income: $<$ 24,999 (20.4 %)	1.40	1.85	0.41	0.53	0.73	0.74	1.06	1.16
	(1.24, 1.56)	(1.74, 1.97)	(0.35, 0.47)	(0.46, 0.6)	(0.66, 0.8)	(0.67, 0.81)	(1.03, 1.09)	(1.12, 1.21)
Income: 25,000 - 44,999 (17.7 %)	1.20	1.19	0.70	0.74	0.90	0.90	1.09	1.01
	(1.06, 1.36)	(1.09, 1.29)	(0.63, 0.77)	(0.67, 0.82)	(0.83, 0.98)	(0.83, 0.98)	(1.06, 1.11)	(0.96, 1.05)
Income: 45,000-69,999 (18.2 %)	1.00	0.96	0.95	0.93	1.00	1.05	1.04	0.95
	(0.87, 1.15)	(0.87, 1.06)	(0.87, 1.03)	(0.84, 1.01)	(0.93, 1.07)	(0.98, 1.12)	(1.01, 1.06)	(0.91, 0.99)
Income: 70,000-99,999 (15.5 %)	0.92	0.60	1.20	1.13	1.09	1.24	0.98	0.91
	(0.77, 1.07)	(0.52, 0.68)	(1.11, 1.31)	(1.03, 1.23)	(1.01, 1.18)	(1.16, 1.32)	(0.95, 1.01)	(0.86, 0.96)
Income: $>$ 100,000 (28.1 %)	0.63	0.51	1.54	1.48	1.21	1.09	0.89	0.96
	(0.52, 0.72)	(0.44, 0.58)	(1.47, 1.62)	(1.4, 1.55)	(1.14, 1.27)	(1.03, 1.15)	(0.87, 0.91)	(0.92, 1.0)
Edu: High School or less (27.4 %)	1.30	1.61	0.75	0.67	0.84	0.97	1.07	0.91
	(1.18, 1.43)	(1.52, 1.69)	(0.69, 0.82)	(0.6, 0.73)	(0.78, 0.9)	(0.91, 1.03)	(1.05, 1.09)	(0.87, 0.95)
Edu: Some College (31.4 %)	1.29	1.18	0.99	0.83	1.09	1.01	1.02	0.88
	(1.19, 1.41)	(1.1, 1.25)	(0.93, 1.05)	(0.77, 0.89)	(1.03, 1.15)	(0.96, 1.06)	(1.0, 1.04)	(0.84, 0.91)
Edu: Graduated College (26.3 %)	0.72	0.55	1.17	1.20	1.06	1.07	0.96	1.07
	(0.63, 0.83)	(0.5, 0.61)	(1.1, 1.24)	(1.12, 1.28)	(1.0, 1.12)	(1.01, 1.13)	(0.94, 0.98)	(1.03, 1.1)
Edu: Post College Grad (14.9 %)	0.31	0.31	1.18	1.63	1.00	0.91	0.89	1.31
	(0.23, 0.4)	(0.24, 0.37)	(1.08, 1.28)	(1.5, 1.75)	(0.92, 1.08)	(0.83, 0.99)	(0.86, 0.92)	(1.26, 1.37)

Table C7: Odds Ratio - Bootstrap Result

D. Appendix D: Alternative Specifications

In Table D1, and Table D2 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	Smokers	Heavy Drinkers	Mostly Wine	Mostly Beer	Nothing	SSB only	Population
Beer 50%	41.88	0.00	26.24	2.10	21.00	0.00	0.00	0.00
Beer 75%	169.57	0.00	81.90	12.12	48.30	0.00	0.00	6.30
Beer 95%	546.00	11.34	350.24	59.50	227.73	10.28	4.20	83.54
Beer mean	140.97	1.95	81.17	15.92	57.16	1.77	0.58	18.78
Wine 50%	2.25	0.00	10.00	20.00	0.75	0.00	0.00	0.00
Wine 75%	12.20	0.00	27.75	42.00	3.00	0.75	0.75	3.00
Wine 95%	123.12	3.75	128.00	144.75	7.50	5.00	4.50	29.00
Wine mean	21.65	0.70	29.47	41.69	1.92	0.90	0.90	6.91
Spirits 50%	8.25	0.00	18.75	1.50	0.85	0.00	0.00	0.00
Spirits 75%	24.69	0.38	35.92	4.00	3.58	0.00	0.00	1.78
Spirits 95%	104.65	4.45	91.00	10.00	10.00	3.25	4.25	19.25
Spirits mean	23.38	0.78	29.99	2.69	2.50	0.52	0.67	3.99
Cigarettes 50%	74.00	60.00	0.00	0.00	0.00	0.00	0.00	0.00
Cigarettes 75%	211.00	178.00	0.00	0.00	0.00	0.00	0.00	0.00
Cigarettes 95%	524.00	460.00	1.00	1.00	2.00	0.00	0.38	39.00
Cigarettes mean	148.05	130.83	0.23	0.34	0.30	0.12	0.11	11.21
SSB 50%	90.35	96.03	56.60	23.70	65.52	3.89	71.99	42.88
SSB 75%	213.35	237.18	129.00	57.40	143.65	8.74	147.98	114.14
SSB 95%	572.48	551.53	332.59	166.13	362.53	14.40	372.20	340.01
SSB mean	158.04	169.70	97.37	47.71	111.31	5.06	119.06	90.84
Ethanol p.a.(L)	9.34	0.20	9.38	4.24	1.97	0.22	0.15	1.67
Cigarette packs p.a.	88.18	81.47	0.13	0.21	0.17	0.09	0.06	6.87
SSBs p.p.(L)	72.57	77.68	42.46	22.02	46.51	3.15	54.49	41.09
Total tax share	0.27	0.41	0.18	0.05	0.05	0.01	0.02	1.00
Alcohol tax share	0.15	0.01	0.49	0.14	0.14	0.02	0.05	1.00
Cigarette tax share	0.34	0.65	0.00	0.00	0.00	0.00	0.00	1.00
SSB tax share	0.04	0.10	0.09	0.04	0.14	0.01	0.57	1.00
Total tax p.a.	284.76	210.07	58.18	20.03	12.32	1.69	1.35	27.58
SSB tax p.a.	28.63	32.16	17.19	8.94	19.75	1.17	22.59	16.97
Effective Ethanol Tax Rate $/L	6.17	7.56	6.08	4.59	6.07	6.36	7.94	5.88
Number of households	1431	2841	5317	5203	6906	13524	26143	61365
Weighted share (%)	2.60	5.50	8.20	7.40	11.50	21.10	43.70	100.00

Table D1: 7 clusters

In Table D3 and Table D4 we replicate Table 1 and Table 2 from the main text. Instead of using the clustering, we instead segment households based on membership in the top decile of purchasers for each sin good category. This is no longer exhaustive and mutually exclusive, so some households will appear in multiple groups while others will appear in none. The demographic patterns in Table D4 and Table 2 are mostly similar, but the results in Table D3 are somewhat difficult to interpret, and the combined tax burden is less clear. The point is to illustrate exactly what clustering "does" here - it enables us to address the combined burden of sin taxes by segmenting households.

	Everything	Smokers	Heavy Drinkers	Heavy Beer	Mostly Wine	Moderate Spirits	Moderate Beer	Nothing	SSB only	Overall
Beer 50%	42.70	0.00	25.92	90.41	4.20	0.00	10.28	0.00	0.00	0.00
Beer 75%	165.90	0.00	69.36	195.64	18.20	6.08	20.40	0.00	0.00	6.30
Beer 95%	540.60	11.10	279.30	619.50	73.50	15.04	52.50	10.50	1.76	83.54
Beer mean	138.25	1.90	67.73	169.50	17.99	3.62	16.19	1.76	0.13	18.78
Wine 50%	2.25	0.00	25.00	0.75	22.17	2.25	0.00	0.00	0.00	0.00
Wine 75%	11.25	0.00	55.50	3.63	45.00	6.00	2.13	0.75	0.75	3.00
Wine 95%	106.50	3.75	186.50	9.00	147.50	13.25	6.00	5.25	4.50	29.00
Wine mean	20.68	0.69	51.91	2.41	43.63	3.92	1.28	0.92	0.83	6.91
Spirits 50%	7.97	0.00	19.50	4.26	0.75	7.00	0.00	0.00	0.00	0.00
Spirits 75%	25.27	0.38	38.50	10.27	2.13	11.77	0.75	0.00	0.00	1.78
Spirits 95%	105.08	4.30	94.75	36.75	5.38	35.96	2.84	2.56	1.75	19.25
Spirits mean	23.39	0.75	31.60	9.56	1.40	11.80	0.56	0.44	0.24	3.99
Cigarettes 50%	77.00	61.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cigarettes 75%	223.00	179.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cigarettes 95%	530.00	460.00	2.00	2.00	1.00	2.00	2.00	0.00	0.05	39.00
Cigarettes mean	151.26	132.32	0.39	0.26	0.34	0.29	0.25	0.14	0.10	11.21
SSB 50%	92.22	96.95	43.97	58.18	20.30	67.11	67.21	2.80	65.37	42.88
SSB 75%	214.08	237.84	110.77	141.31	50.37	144.51	141.88	6.68	138.90	114.14
SSB 95%	573.47	557.82	282.03	379.28	158.54	343.67	369.60	11.51	361.22	340.01
SSB mean	159.49	170.10	82.61	110.79	43.70	110.39	112.06	3.85	112.14	90.84
Ethanol p.a.(L)	9.19	0.19	11.00	6.36	4.30	2.02	0.57	0.20	0.09	1.67
Cigarette packs p.a.	90.31	82.30	0.21	0.16	0.21	0.17	0.15	0.10	0.06	6.87
SSBs p.p.(L)	72.83	77.77	36.06	48.69	20.00	46.61	47.21	2.45	52.15	41.09
Total tax share	0.27	0.41	0.13	0.07	0.04	0.05	0.02	0.01	0.01	1.00
Alcohol tax share	0.15	0.01	0.35	0.18	0.11	0.14	0.04	0.02	0.02	1.00
Cigarette tax share	0.34	0.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SSB tax share	0.04	0.10	0.05	0.06	0.03	0.10	0.13	0.01	0.48	1.00
Total tax p.a.	289.10	212.42	64.11	37.54	19.37	17.86	3.66	1.61	0.72	27.58
SSB tax p.a.	28.84	32.23	14.41	19.76	8.10	19.66	20.12	0.91	21.54	16.97
Effective Ethanol Tax Rate \$/L	6.21	7.53	5.70	5.96	4.39	8.64	5.75	6.14	6.17	5.88
Number of households	1400	2792	3506	2982	4235	4745	6161	11774	23770	61365
Weighted share (%)	2.50	5.40	5.20	4.80	6.00	7.90	10.50	18.30	39.30	100.00
Share of households (%)	2.28	4.55	5.71	4.86	6.90	7.73	10.04	19.19	38.74	100.00

Table D2: 9 clusters

We also report the probability that a household is in the top decile for more than one category in Table D5. For example only 3.3% of households are in the top decile for both Spirts and Beer purchases. We use top 5% instead of top 10% for cigarettes since only around 8% of households purchase cigarettes (and 91% of households are tied with zero cigarette purchases).

	Beer	Wine	Spirits	Cigarettes	SSB	Diet	Population
Beer 25%	49.52	0.00	0.65	0.00	0.00	0.00	0.00
Beer 50%	83.40	8.40	12.60	0.00	0.00	0.00	0.00
Beer 75%	176.88	37.98	50.40	16.72	8.40	8.40	6.30
Beer mean	160.91	44.06	56.25	57.27	26.50	22.83	18.78
Wine 25%	0.00	18.00	0.75	0.00	0.00	0.00	0.00
Wine 50%	3.00	28.96	5.25	0.00	0.00	0.00	0.00
Wine 75%	14.25	60.00	20.25	1.50	1.50	3.00	3.00
Wine mean	18.74	57.00	23.11	9.46	4.59	8.11	6.91
Spirits 25%	0.00	0.00	12.25	0.00	0.00	0.00	0.00
Spirits 50%	3.50	3.98	19.25	0.00	0.00	0.00	0.00
Spirits 75%	12.78	13.77	36.32	4.25	2.25	3.00	1.78
Spirits mean	12.86	12.83	32.22	9.19	6.11	6.60	3.99
Cigarettes 25%	0.00	0.00	0.00	75.00	0.00	0.00	0.00
Cigarettes 50%	0.00	0.00	0.00	145.00	0.00	0.00	0.00
Cigarettes 75%	0.00	0.00	0.00	274.00	1.00	0.00	0.00
Cigarettes mean	23.02	10.51	19.76	209.46	31.77	15.55	11.21
SSB 25%	19.20	11.11	20.08	25.91	276.03	13.10	12.39
SSB 50%	57.62	36.00	59.20	97.68	340.01	47.73	42.88
SSB 75%	143.24	91.66	141.56	237.69	456.94	125.36	114.14
SSB mean	110.60	73.58	106.74	171.19	407.40	100.34	90.84
Ethanol per adult (L)	7.57	7.64	8.54	3.75	1.70	2.18	1.67
Cigarette packs per adult	13.63	6.60	11.49	128.71	17.82	8.80	6.87
SSBs per capita (L)	48.23	32.74	47.06	82.24	166.75	42.67	41.09
Total tax share	0.30	0.21	0.32	0.63	0.21	0.14	1.00
Alcohol tax share	0.46	0.40	0.57	0.12	0.13	0.15	1.00
Cigarette tax share	0.20	0.09	0.17	0.93	0.26	0.14	1.00
SSB tax share	0.12	0.08	0.12	0.09	0.45	0.11	1.00
Total tax per adult	78.86	56.08	85.40	352.91	52.75	35.77	27.58
SSB tax per adult	19.66	13.31	18.79	32.57	71.12	16.95	16.97
Effective Ethanol Tax Rate $/L	5.84	5.17	6.49	6.18	6.46	5.97	5.88
Number of households	6331	6987	6312	2705	5430	6862	61365
Weighted share	10.00	10.00	10.00	5.00	10.00	10.00	100.00
Share of households	6.60	7.28	6.58	2.82	5.66	7.15	63.93

Table D3: Annual Household Purchases by Top Decile Buyers

	Beer	Wine	Spirits	Cigarettes	SSB	Diet
Race: White (74.8%)	1.07	1.07	1.04	1.11	1.02	1.18
Race: Black (12.5%)	0.67	0.76	0.96	0.81	1.09	0.31
Race: Asian (4.5%)	0.49	0.64	0.57	0.41	0.47	0.47
Race: Other (8.2%)	1.12	0.88	0.94	0.64	1.00	0.74
Hispanic: Yes (13.2%)	1.15	0.95	0.82	0.59	1.00	0.67
Hispanic: No (86.8%)	0.98	1.01	1.03	1.06	1.00	1.05
Children: Yes (31.3%)	0.92	0.80	0.83	0.72	1.39	0.79
Children: No (68.7%)	1.04	1.09	1.08	1.13	0.82	1.10
Age: $<$ 35 (12.9%)	0.81	0.75	0.74	0.49	0.87	0.45
Age: 35 to 44 (18.0%)	0.89	0.84	0.89	0.68	1.15	0.75
Age: 45 to 54 (21.8%	1.04	0.95	1.09	1.06	1.26	1.23
Age: 55 to 64 (22.7%)	1.21	1.03	1.14	1.43	1.03	1.24
Age: $>$ 65 (24.6%)	0.95	1.26	1.01	1.04	0.70	1.05
Income: $<$ 24,999 (20.4%)	0.79	0.53	0.67	1.76	1.23	0.82
Income: 25,000 - 44,999 (17.7%)	0.89	0.73	0.86	1.20	1.19	0.80
Income: 45,000-69,999 (18.2%)	1.02	0.92	0.98	0.98	1.08	0.93
Income: 70,000-99,999 (15.5%)	1.16	1.14	1.12	0.68	0.85	1.09
Income: $>$ 100,000 (28.1%)	1.12	1.49	1.27	0.52	0.75	1.25
Edu: High School or less (27.4%)	1.08	0.71	0.89	1.49	1.43	1.00
Edu: Some College (31.4%)	1.04	0.92	1.08	1.25	1.08	1.04
Edu: Graduated College (26.3%)	0.98	1.16	1.04	0.61	0.77	0.93
Edu: Post College Grad (14.9%)	0.80	1.43	0.94	0.27	0.45	1.03

Table D4: Odds Ratio by Extreme Consumers

	Beer	Wine	Spirits	Cigarettes	SSB	Diet
Beer	0.100	0.027	0.033	0.010	0.013	0.012
Wine	0.027	0.100	0.034	0.005	0.006	0.011
Spirits	0.033	0.034	0.100	0.009	0.012	0.014
Cigarettes	0.010	0.005	0.009	0.050	0.013	0.007
SSB	0.013	0.006	0.012	0.013	0.100	0.012
Diet	0.012	0.011	0.014	0.007	0.012	0.100

Table D5: Pr(Top Decile Row, Top Decile Column)

E. Appendix E: Sample Selection and Outliers

In our analysis we drop a total of 19 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 19 cases. These observations are excluded from the analysis. Summary stats for those households are provided in Table E1.

Panel A: Alcohol Outliers - 16

	Beer	Spirits	Wine	Ethanol	Cigarette	Household Size	Adult	SSB	Income
Mean	1406.8	100.0	234.4	128.5	81.8	1.2	1.2	78.8	33406
Std	2733.8	119.3	333.6	103.3	156.8	0.4	0.4	95.3	21989
Min	0.0	0.0	0.0	75.0	0.0	1.0	1.0	0.0	2500
25%	0.0	2.6	17.4	86.3	0.0	1.0	1.0	16.0	16500
50%	57.8	36.2	36.8	102.1	0.0	1.0	1.0	39.0	32500
75%	1885.2	190.8	502.6	114.0	58.1	1.0	1.0	106.9	38750
Max	10752.9	380.9	925.0	502.7	446.0	2.0	2.0	330.2	85000

Panel B: Tobacco Outliers - 3

	Beer	Spirits	Wine	Ethanol	Cigarette	Household Size	Adult	SSB	Income
Mean	0.7	1.7	1.0	0.4	1275.0	1.0	1.0	89.5	34500
Std	1.2	2.9	1.7	0.4	47.7	0.0	0.0	75.9	27672
Min	0.0	0.0	0.0	0.0	1245.0	1.0	1.0	3.6	11000
25%	0.0	0.0	0.0	0.2	1247.5	1.0	1.0	60.6	19250
50%	0.0	0.0	0.0	0.5	1250.0	1.0	1.0	117.6	27500
75%	1.0	2.5	1.4	0.6	1290.0	1.0	1.0	132.4	46250
Max	2.1	5.0	2.9	0.8	1330.0	1.0	1.0	147.3	65000

Table E1: 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 19 households of the 61,384 households in our sample exceed these purchase thresholds and are removed from the main sample.