Regressive Sin Taxes

Benjamin B. Lockwood and Dmitry Taubinsky*

March 6, 2017

Abstract

A common objection to "sin taxes"—corrective taxes on goods like cigarettes, alcohol, and sugary drinks, which are believed to be over-consumed—is that they fall disproportionately on low-income consumers. This paper studies the interaction between corrective and redistributive motives in a general optimal taxation framework. On the one hand, redistributive concerns amplify the corrective benefits of a sin tax when sin good consumption is concentrated on the poor, even when bias and demand elasticities are constant across incomes. On the other hand, a sin tax can generate regressivity costs, raising more revenue from the poor than from the rich. Sin tax regressivity can be offset by targeted transfers or income tax reforms if differences in sin good consumption are driven by income effects, but not if they are driven by preference heterogeneity, and not if the indirect incentives the sin tax generates for labor supply decisions are not salient. The price elasticity of demand determines the extent to which corrective benefits versus regressivity costs determine the size of the optimal tax. We implement our optimal tax formulas in a calibrated model of sugar-sweetened beverage consumption for a range of parameter values suggested by empirical work.

*Lockwood: The Wharton School, University of Pennsylvania (E-mail: ben.lockwood@wharton.upenn.edu). Taubinsky: Dartmouth and NBER (E-mail: dtaubinsky@gmail.com). We thank Hunt Allcott, Alan Auerbach, Raj Chetty, Stefano DellaVigna, Emmanuel Farhi, Xavier Gabaix, Nathaniel Hendren, Louis Kaplow, David Laibson, Erzo F.P. Luttmer, Matthew Rabin, Alex Rees-Jones, Emmanuel Saez, Jim Sallee, Florian Scheuer, Stefanie Stantcheva, Matthew Weinzierl, and participants at seminars and conferences for helpful comments and discussions.

"The only way to protect all of us, including the poor, from further harm is through a sugary drink tax...."

– *Forbes* article [April 2016]

"A tax on soda and juice drinks would disproportionately increase taxes on low-income families in Philadelphia."

– Bernie Sanders, U.S. Senator [April 2016]

"They've [big soda] made their money of the backs of poor people, but this money [soda tax revenue] will stay in poor neighborhoods."

– Jim Kenney, Philadelphia Mayor [in *The Nation*, May 2016]

1 Introduction

A large literature in behavioral economics shows that poor self control, inattention, and incorrect beliefs lead to overconsumption of goods such as cigarettes, alcohol, unhealthy junk foods, and energy inefficient appliances. Consequently, "sin taxes" that discourage the consumption of such goods have the potential to increase social welfare.[1] This corrective logic is frequently mentioned in policy debates, and has played a role in the passage of taxes on cigarettes and alcohol, as well as the recent rise of taxes on sugar-sweetened beverages (SSBs) in US cities that include Chicago, Philadelphia, and San Francisco.

One simple principle for the optimal tax that has been emphasized in behavioral economics is the basic logic of Pigou: the optimal bias-correcting tax should equal the average mistake (or "internality") of marginal consumers.[2] But this efficiency-maximizing principle depends on strong assumptions that ensure that consumers do not vary in their marginal utility from wealth.

Such strong assumptions are starkly out of sync with public debates about sin taxes. Indeed, a common objection to sin taxes is that they are regressive. Cigarettes and sugary drinks are consumed disproportionately by the poor, and energy efficiency subsidies are taken up disproportionately by the rich.[3] This has lead to forceful opposition to the taxes on the grounds of fairness and equality.[4]

Yet at the same time, others argue that the regressivity of the tax can be offset through progressive "revenue recycling": government revenue from the tax can be used to fund progressive initiatives and transfers to low-income consumers.[5] And some also point out that the biases that cause over-

[1] See, e.g., Chetty (2015), Mullainathan et al. (2012), and Bernheim and Rangel (2009), for an overview of the various biases and their implications for welfare calculations.

[2] See, e.g., O'Donoghue and Rabin (2006), Gruber and Kőszegi (2001), Allcott et al. (2014), Allcott and Taubinsky (2015), Heutel (2015), Tsvetanov and segerson (2013), Mullainathan et al. (2012).

[3] See, e.g., Gruber and Kőszegi (2004) and Goldin and Homonoff (2013) on cigarettes, and Allcott et al. (2015), Davis and Borenstein (2016), Davis and Knittel (2016) on energy efficiency subsidies and fuel economy standards.

[4] See, e.g., Lavin (2013) on cigarettes taxes or Sanders (2016) on SSB taxes.

[5] For example, this reasoning led Philadelphia to direct a share of SSB tax revenues toward the development of Pre-K education initiatives in low income neighborhoods. Philadelphia mayor James Kenney has proposed the soda tax under the promise that its revenues "will stay in poor neighborhoods." Consistent with this, Blumgart (2016), for example, writes that "By earmarking the revenues to programs that are at the top of the progressive agenda, Kenney has provided a strong counterargument to those concerned by the regressive nature of the tax: The money will go right back into those very neighborhoods that are hit hardest."

consumption are *themselves* regressive—e.g., cigarette consumption among the poor leads to higher insurance premiums for the poor—and that, therefore, a corrective tax might confer greater benefits on the poor than on the rich.[6]

What is the optimal tax in the presence of such diverse policy considerations? This paper is the first to thoroughly address this question in a context sufficiently comprehensive to account for each of these arguments. We allow for individuals who over- or under-consume some "sin good," the consumption of which varies across the income distribution. We allow for differences in consumption to be driven by income effects, preference heterogeneity, or some combination of the two. And we incorporate a motive for redistribution and a nonlinear income tax that can be used to recycle sin tax revenue progressively.

Our baseline model builds on Saez's (2002a) extension of Atkinson and Stiglitz (1976) by considering an economy of consumers with heterogenous earning abilities and tastes, who choose labor supply and a consumption bundle that exhausts their after-tax income. The policymaker chooses a set of linear commodity taxes and a non-linear income tax which can be used to provide transfers to low income consumers, to raise money for commodity subsidies, or to distribute commodity tax revenue (in a progressive way, if desired). But while the standard approach in optimal tax theory assumes that the planner is in agreement with consumers about what is best for them, we instead analyze the general case in which the policymaker disagrees with the consumers' assessment due to "mistakes" such as present bias, limited inattention, or incorrect beliefs. This approach nests a variety of behavioral economics models of biases as special cases, and generates robust economic insights across a number of domains of behavior. Moreover, although externalities are not the primary focus of the paper, optimal tax policy in the presence of externalities is obtained as a corollary of our more general results.

Our main theoretical results show that redistributive concerns have a nuanced effect on the optimal commodity tax, and can make the tax substantially larger or smaller than a benchmark Pigouvian tax. We show that the optimal commodity tax can be written as the sum of two terms: the "corrective benefits" (representing the welfare gains from reducing harmful internalities), and the "regressivity costs" (representing the welfare costs of shifting net resources from poorer to richer consumers). When a harmful good is consumed primarily by poorer consumers, redistributive motives tend to amplify the corrective benefits, since the beneficial reductions in harmful internalities accrue mostly to the poor.

On the other hand, redistributive motives also generate regressivity costs when the taxed good is consumed more heavily by the poor than by the rich. The extent to which the regressivity of the commodity tax can be offset (e.g., by making the income tax more progressive, or by targeting commodity tax revenues toward programs that benefit the poor) depends on how much of the sin good consumption profile is driven by income effects vs. preference heterogeneity. In the extreme

[6]See Gruber and Kőszegi (2004) for a formal argument along these lines for cigarette taxes. Less formally on the recent SSB debate, Huehnergarth (2016) writes "Here's why Sanders' position on sugary drink taxes plays right into the hands of Big Soda, and hurts America's most vulnerable citizens: Big Soda's targeted marketing to communities of color and low-income communities is regressive.... Type 2 diabetes, linked to excessive sugary drink consumption, is regressive...."

case where all differences are from income effects, regressivity can be fully offset, so that only the corrective benefits are relevant, and the optimal tax is unambiguously higher than the Pigouvian benchmark when elasticities and biases are non-decreasing with income. But when income effects play a smaller role, no such offset is feasible because a more progressive income tax will lead to labor supply distortions whose cost to public funds outweighs the redistributive benefits.

The price elasticity of demand determines the relative importance of corrective benefits versus regressivity costs. When the elasticity is low, the tax has little effect on behavior and thus little corrective benefit—and so its regressivity costs dominate. Conversely, corrective benefits dominate when the elasticity is high.

We provide a number of extensions to our baseline model and results. In theory, small changes in the tax (or price) of any particular good affect the returns from labor supply, and the standard model assumes that individuals adjust their labor supply perfectly in response to all possible changes. In our first extension, we relax this assumption and allow consumers to under-react to such indirect incentives for labor supply. We show formally that such underreaction to commodity taxes makes progressive revenue recycling of commodity tax revenue inefficient, and thus *decreases* the size of the optimal sin tax. In fact, we show that a non-salient (for labor supply choice) commodity tax operates on behavior almost as if the consumption profile were driven purely by preference heterogeneity, even when sin good consumption is driven entirely by income effects.

Second, we consider an extension to multiple goods (possibly complements or substitutes) that may be over- or under-consumed to different degrees. We present general formulas that shed light on questions like how the government should set taxes jointly on standard tobacco cigarettes and electronic cigarettes, or on regular soda and diet soda. Third, we show how results for externalities (or a combination of both externalities and internalities) can be obtained as corollaries of our results.

To demonstrate the quantitative importance of our theoretical results, we apply our optimal tax formulas to a calibrated model of sugar-sweetened beverage (SSB) consumption for a range of parameter values suggested by empirical work. Consumption of SSBs generates large health costs that are likely to be at under-appreciated by many consumers, and it also declines rapidly across the income distribution. We find that if SSBs are inferior goods, so that the profile of SSB consumption is driven by income effects, the optimal tax is modestly higher than a Pigouvian tax. When instead the SSB consumption profile is the consequence of preference heterogeneity, regressivity costs lower the optimal tax. At the lower end of empirical demand elasticity estimates, regressivity costs are large relative to corrective benefits, lowering the optimal tax by about 50%. For moderate levels of consumer bias, the welfare gains from the optimal tax range from \$0.14 billion to \$34.1 billion, depending on the elasticity and on how much of the consumption profile is due to preference heterogeneity. The simulations thus show that the comparative statics implied by our theoretical formulas are quantitatively very significant. At the same time, our results also suggest that for a broad swath of reasonable parameters, the optimal tax is moderately higher than the 1 to 2 cent per ounce taxes currently in place in US cities.

We derive our optimal tax formulas in terms of sufficient statistics that consist of elasticities and an empirically estimable, price-metric definition of bias. We generalize the seminal theorem of

Atkinson and Stiglitz (1976) and the qualitative results of Saez (2002a) by providing expressions for the commodity tax in the presence of arbitrary preference heterogeneity, arbitrary bias in consumption of the sin good, and potential underreaction to commodity taxes on the labor supply margin. As a technical contribution, our results provide a novel generalization of the Atkinson-Stiglitz Theorem even in the absence of any consumer misoptimization.[7]

Our results also generalize and qualify several recent advances in behavioral public economics. A number of papers abstract from redistributive concerns and focus purely on the corrective benefits of a corrective "internality tax."[8] Such frameworks can essentially be obtained as a special case of our model by setting social marginal welfare weights to 1.[9]

Amongst the small number of papers that do consider redistributive concerns, Gruber and Kőszegi (2004) study the incidence of cigarette taxes on low- and high-income consumers with self-control problems, and show that low-income consumers with sufficiently elastic consumption can benefit from such taxes, but they do not characterize the optimal tax implications. Bernheim and Rangel (2004) and Farhi and Gabaix (2015) study optimal commodity taxation with behavioral consumers who differ in their marginal utility from income. Bernheim and Rangel (2004) study the optimal tax on addictive substances in a dynamic framework in which environmental cues trigger mindless consumption, and differences in the marginal utility from wealth arise from differences in spending on the addictive good or on rehabilitation. Farhi and Gabaix (2015) focus on an application of their general framework that assumes two types of consumers—behavioral and rational—with potential differences in the marginal utility from income. Both Farhi and Gabaix (2015) and Bernheim and Rangel (2004) assume that commodity taxes are the sole source of redistribution: the commodity tax revenue has to be distributed lump-sum and cannot, for example, be spent on transfers to the poor or programs that benefit the poor. Our model complements and extends their approach by allowing for a nonlinear income tax in addition to commodity taxes, while also allowing for richer assumptions about biases and heterogeneity.[10]

The rest of the paper proceeds as follows. Section 2 presents our general model. Section 3 derives results for a simple, illustrative example. Section 4 characterizes the optimal commodity and income taxes in our general model with a continuum of consumers. Section 5 presents extensions to the baseline model. Section 6 presents numerical results for the optimal tax in a calibrated model of sugar-sweetened beverage consumption. Section 7 concludes.

[7] Moreover, our results also generalize Jacobs and Boadway (2014), who study optimal commodity taxation in the absence of any market failure under the assumption that consumption preferences are homogeneous while allowing for non-separability between consumption and labor in the utility function, and also Jacobs and de Mooij (2015) and Kaplow (2012), who study the taxation of environmental externalities under the assumptions of homogeneous preferences. By considering a flexible role of income effects vs. preference heterogeneity, our model accounts for the potential labor supply distortions discussed in Goulder and Williams (2003), and contributes to the "double dividend" debate in environmental economics (see Goulder (1995) for a review).

[8] See, e.g., the papers in footnote 2.

[9] See, in particular, Corollary 2 in Section 4.3.1.

[10] Also closely related is the work on health insurance by Baicker et al. (2015), Spinnewijn (forthcoming), and Handel et al. (2016). See also Mullainathan et al. (2012) for a model that links the insights from the health insurance literature to commodity tax results; they derive formulas similar to those in Farhi and Gabaix (2015) for an economy in which consumers make a binary choice.

2 Model

2.1 The environment

We consider individuals with multidimensional types denoted $\theta \in \Theta$, distributed with measure $\mu(\theta)$. Individuals choose a level of earnings z, which is taxed according to the nonlinear income tax $T(z)$. Consumers use their net income to consume a commodity c_2 (the "sin good") which is subject to a linear commodity tax t, and a composite of all other goods c_1 (the "numeraire") which does not face a commodity tax. We normalize the price of c_1 to 1, so that the individual's budget constraint is $c_1 + (p+t)c_2 \leq z - T(z)$, where p is the price of c_2. Individuals make their choices to maximize "decision utility" $U(c_1, c_2, z; \theta)$ subject to this constraint. U is assumed to be increasing and weakly concave in its first two arguments, and decreasing and strictly concave in the third.

The policymaker believes consumers should be maximizing $V(c_1, c_2, z; \theta)$, which may differ from from U because of the mistakes that the policymaker believes that consumers might be making. The policymaker selects $T(\cdot)$ and t to maximize aggregate experienced utility, subject to a budget constraint and individual optimization. That is, the policymaker's problem is to maximize aggregate experienced utility, weighted by type-specific Pareto weights $\alpha(\theta)$:

$$\max_{T,t} \int \alpha(\theta) V(c_1(\theta), c_2(\theta), z(\theta); \theta) \mu(\theta),$$

subject to the budget constraint

$$\int (tc_2(\theta) + T(z(\theta))) \mu(\theta) = 0$$

and to individual optimization

$$\{c_1(\theta), c_2(\theta), z(\theta)\} = \arg \max_{\{c_1, c_2, z\}} U(c_1, c_2, z; \theta) \quad \text{s.t.} \quad c_1 + (1+t)c_2 \leq z - T(z) \text{ for all } \theta.$$

The difference between U and V can capture a variety of different psychological biases, such as:

1. *Incorrect beliefs*, as documented for food choices and energy-efficiency choices by Allcott (2013), Attari et al. (2010), Bollinger et al. (2011), and Schofield (2015). Consumers may have incorrect beliefs about certain attributes of c_2, such as its calorie content, it's future health costs, or its energy efficiency.

2. *Limited attention* or *salience bias* with respect to certain attributes of c_2, as modeled in Gabaix and Laibson (2006), DellaVigna (2009), Gabaix (2014), and Bordalo et al. (2013), and as documented by Allcott and Taubinsky (2015) for energy-efficient appliances. For example, consumers may be attentive to the product features advertised by the manufactures, but less attentive to features not advertised: A sugary ice cream that is advertised as "fat free" may appear healthy to consumers who did not examine the less conspicuously displayed information on sugar content. Alternatively, suppose that c_2 represents investment in energy-efficiency, which generates

both hedonic utility and energy cost savings. Inattentive consumers may underweight the future costs.

3. *Present bias/time inconsistency.* As in O'Donoghue and Rabin (2006), suppose that consumers underweight the future health costs of c_2 (e.g, potato chips) because of $\beta - \delta$ time-inconsistent preferences. For example, if true utility from c_2 is given by $m(c_2) - h(c_2)$, where h is health costs and m is immediate consumption utility, consumers may instead perform optimization over $m(c_2) - \beta h(c_2)$. Our framework allows us to treat β as a bias. However, our framework also allows us to study other welfare criteria that may be applied to the model. For example, a policymaker maker's objective function may assign $m(c_2) - (\kappa + (1-\kappa)\beta)h(c_2)$ to be the utility from c_2, reflecting that the policymaker gives weight κ to the "future-oriented self" and weight $1 - \kappa$ to the "in the moment self."[11]

4. Any combination of the above biases, as well as any other biases that allow consumer choice to be represented by a differentiable utility function.

However, it is not always clear that all deviations of behavior from the standard model constitute mistakes, and it is not the purpose of this paper to contribute to the relevant debates. The analysis in this paper is perhaps best understood as a policy design calculation, given a well-specified objective function *from the policymaker*. For example, if a policymaker deems that V should correspond to the "future-oriented" self's preferences in cases of time inconsistency, then our framework and results deliver sharp policy design implications for such an objective function. Alternatively, a policymaker may instead deem that decisions made in the present are "truer," and that the future-oriented self suffers from an over-abstraction bias. This also constitutes a well-specified objective function, and in this case our framework and results deliver sharp policy design implications as well. Relatedly, in the framework of Bernheim and Rangel (2009), the utility function U can be thought of as representing choices in the normal market frame, whereas V represents choices in the welfare-relevant domain.[12]

2.2 Defining consumer bias

A key goal of our theoretical analysis is to derive optimal formulas that on the one hand can be applied to a variety of possible consumer biases, while on the other hand can be directly connected to empirical measurement through an intuitive and operationalizable definition of bias. We thus adopt a general, yet empirically grounded definition of consumer bias in the spirit of the sufficient statistics approach. We do this by constructing a price metric for consumer bias.

[11] See Bernheim and Rangel (2009) and Bernheim (2016) for a discussion of why treating $\beta < 1$ fully as a bias may be problematic.

[12] What our framework does not allow explicitly, however, is ambiguity about which choices are optimal for a consumer; that is, we assume that the "welfare-relevant choices" can be represented by a utility function and do not, e.g., exhibit cycles. This may be a strong assumption in some circumstances, as argued by Bernheim and Rangel (2009) and Bernheim (2016). Note, however, that all of our derivations are immediately relevant even if there is uncertainty about whether the normative criterion should be V^1 or V^2, or V^3 .... If the policymaker attaches probability π_i to each possible V^i, and maximizes $\sum \pi_i V^i$, then our results immediately carry over to this policy problem by setting $V := \sum \pi_i V^i$. More generally, because we can characterize optimal policy for any criterion $V := \sum \hat{\pi}_i V^i$ for any set of weights $\hat{\pi}_i$, our results can be used to characterize the set of all "justifiable policies."

Formally, let $c_2(\theta, y, p, t, T)$ be the consumption chosen by a type θ individual with after-tax earnings $y = z - T(z)$ at price p and tax t. Analogously, define $c_2^V(\theta, y, p, t, T)$ to be the amount of c_2 that would be chosen if the individual were maximizing V instead. We define the bias $\gamma(\theta, z, t, T)$ as the value for which $c_2(\theta, y, p, t, T) = c_2^V(\theta, y - c_2\gamma, p - \gamma, t, T)$. In words, γ is equal to the compensated price change that produces the same effects on demand as the bias does. In terms of primitives, $\gamma = \frac{U_2}{U_1} - \frac{V_2}{V_1}$.[13] Sometimes we suppress some of the arguments and simply write $\gamma(\theta)$.

A key feature of γ is that it is not tied to the policymaker's redistributive preferences, and that it is invariant to any monotonic transformation of the utility functions U and V. The statistic depends only on a consumer's observable choices in the "biased" and "debiased" states.[14]

Although in some cases γ can only be measured indirectly by, for example, using structural estimates of present bias and then constructing the price metric, in other cases it can be measured directly. Chetty et al. (2009) estimate the (average) value of γ directly when bias arises from lack of tax salience by computing the change in upfront prices that would alter demand as much as a debiasing intervention that displays tax-inclusive final prices. Similarly, Allcott and Taubinsky (2015) run an experiment that provides a direct estimate of γ for each consumer's biased valuation of energy efficient lightbulbs (CFLs). They compute willingness to pay (WTP) for more versus less energy efficient lightbulbs in a standard market frame, and then measure how the distribution of WTP changes when biases arising from inattention or incorrect beliefs are eliminated.[15]

2.3 Redistributive motives

To represent redistributive motives concisely, we adopt the notation, common in the optimal taxation literature, of "marginal social welfare weights"—the marginal social welfare generated by a marginal unit of composite consumption for a particular individual, in terms of public funds—formally defined here as $g(\theta) = \alpha(\theta)V_1/\lambda$, where V_1 represents the derivative of V with respect to its first argument, and λ is the marginal value of public funds (i.e., the multiplier on the government budget constraint at the optimum). These weights are endogenous to the tax system, but are useful for characterizing the necessary conditions which must hold at the optimum. We use $\bar{g} = \int_\Theta g(\theta)d\mu(\theta)$ to denote the average marginal social welfare weight. If there are no income effects on consumption and labor supply, then $\bar{g} = 1$ by construction. In this case, a social marginal welfare weight is simply the (normalized) value of giving one unit of (after-tax) income to this consumer.

Note that because the weights $\alpha(\theta)$ are exogenous, and because U and V produce identical

[13] To see this, note that by definition, $U_2/U_1 = p + t$ and $V_2/V_1 = p + t - \gamma$ when the arguments c_1, c_2, z are the same in the U_i and V_i terms . This also shows that the equivalence only holds under the assumption that the tax t is fully salient; otherwise, γ may deviate from $\frac{U_2}{U_1} - \frac{V_2}{V_1}$.

[14] Assuming that "debiased" choices are observable, as they are for most behavioral models; see Bernheim (2016) for the argument that the identification of almost all biases in behavioral economics can be done, and typically is, by directly measuring choices V in the "unbiased frame" and comparing those to the choices U in the potentially biased frame. Otherwise, γ can also be thought of as a function of observable choices and the policymaker's desired choice.

[15] Even in contexts in which the analyst has a particular psychological bias in mind, estimating γ may sometimes be easier than estimating the primitives of the particular psychological model. For example, if the analyst believes that individuals overconsume sugary drinks due to present bias, γ can be estimated directly by comparing an individual's period 0 demand for period 1 sugary drink consumption (reflecting choices according to V) to the individual's period 1 demand for period 1 sugary drink consumption (reflecting choices according to U).

behavior (and identical choice-based measures of bias) up to monotonic transformations, redistributive motives correspond to a policymaker's or society's tastes for reducing wealth inequality, and cannot be derived from consumers' choices alone. As in the rest of the optimal taxation literature, our formulas for optimal taxes will thus depend not just on observable behavior (and people's quantifiable mistakes), but also on the policymaker's (or society's) preferences for wealth equality, captured by these weights.

3 The optimal tax in a simple model with discrete types

We first derive the optimal sin tax in two contrasting contexts where the consumption profile of c_2 is driven by either preference heterogeneity or income effects. For simplicity, in this section we assume two types of consumers, $\theta \in \{L, H\}$, in equal proportion, with wages $w_L < w_H$. For concreteness, suppose the internality is harmful ($\gamma(\theta) > 0$) and that L consumes more c_2 than H. We normalize the units of c_2 so $p = 1$, so that each individual's budget constraint requires that $c_1^*(\theta) = z^*(\theta) - T_\theta - (1+t)c_2^*(\theta)$, where $c_i^*(\theta)$ denotes type θ's of c_i.

3.1 Example 1: Regressivity caused by heterogeneous preferences

We assume the following functional forms:

$$\begin{aligned} U(c_1, c_2, z; \theta) =& G(c_1 + v(c_2, \theta) - \Psi(z/w_\theta)) \\ V(c_1, c_2, z; \theta) =& G(c_1 + v(c_2, \theta) - \gamma(\theta)c_2 - \Psi(z/w_\theta)) \end{aligned}$$

where v is increasing and strictly concave in its first argument, and its dependence on θ represents preference heterogeneity, which generates differences in consumption of c_2 even if both types were given the same net income. Moreover we assume $v'_{c_2}(c_2, H) < v'_{c_2}(c_2, L)$ for all c_2, so that $c_2^*(H) < c_2^*(L)$.[16] Labor supply disutility Ψ is increasing and convex, implying a single crossing condition, and G is a concave function, implying diminishing marginal utility of consumption. Thus $\gamma(\theta)$ represents the money-metric bias wedge, in keeping with the definition from the previous section. One implication of this specification is that there are no income effects—shocks to a person's wealth do not change one's choice of c_2, nor one's chosen labor supply.

The policy maker chooses a commodity tax t and taxes T_L and T_H (possibly negative) on the the earnings of the two types to maximize social welfare

$$W := \sum_\theta V\left(c_1^*(\theta), c_2^*(\theta), z^*(\theta); \theta\right) \mu(\theta)$$

subject to the resource constraint $\frac{1}{2}\sum_\theta \left(T_\theta + tc_2^*(\theta)\right) \geq 0$ and subject to the individual optimization constraint that $(c_1^*(\theta), c_2^*(\theta), z^*(\theta))$ maximizes $U(c_1, c_2, z; \theta)$ subject to each individual's budget constraint. To solve for the optimal commodity tax t, we form the Lagrangian

[16] Notationally, we use $f'_x(x, y)$ to denote the partial derivative of f with respect to x, $f''_{xy}(x, y)$ the cross partial, etc., throughout the paper.

$$\mathcal{L} = W + \frac{\lambda}{2} \sum_{\theta} \left(T_\theta + t c_2^*(\theta) \right).$$

The multiplier λ represents the marginal value of public funds. The social marginal welfare weights are given by $g(\theta) = G'\left(c_1^*(\theta) + v(c_2^*(\theta), \theta) - \gamma(\theta) c_2^*(\theta) - \Psi(z^*(\theta)/w_\theta)\right)/\lambda$, and average to one because the utility function does not exhibit income effects. The derivative of the Lagrangian with respect to t must equal zero at the optimum, implying

$$0 = \sum_{\theta} g(\theta) \left(-\gamma(\theta) \frac{dc_2^*(\theta)}{dt} - c_2^*(\theta) \right) + \sum_{\theta} \left(c_2^*(\theta) + t \frac{dc_2^*(\theta)}{dt} \right),$$

where we have used the fact that individual optimization implies $v'_{c_2}(c_2, \theta) = 1 + t$. Rearranging, the optimal commodity tax t^* satisfies

$$t^* = \underbrace{\frac{\sum_\theta g(\theta)\gamma(\theta)\frac{dc_2^*(\theta)}{dt}}{\sum_\theta \frac{dc_2^*(\theta)}{dt}}}_{\text{corrective benefits}} - \underbrace{\frac{\sum_\theta c_2^*(\theta)(1 - g(\theta))}{\sum_\theta \frac{dc_2^*(\theta)}{dt}}}_{\text{regressivity costs}} \tag{1}$$

Written this way, the optimal tax can be transparently decomposed into two terms. The first term represents the corrective benefits from reducing c_2—correction is more valuable to the extent that the consumers who respond most to the tax have (1) greater bias $\gamma(\theta)$, or (2) higher welfare weight $g(\theta)$. The second term represents the regressivity costs of taxing a commodity consumed primarily by the poor. Since $c_2^*(H) < c_2^*(L)$ and the welfare weights average to one, the regressivity costs term reduces the optimal tax. This force captures the common intuition that corrective taxes on goods consumed by the poor should be reduced for redistributive reasons. As will be apparent from the next example, this intuition depends crucially on the the assumption that *preference heterogeneity* is the reason low earners consume more c_2.

3.2 Example 2: Regressivity caused by income effects

We now consider the case where income effects are responsible for differences in c_2 consumption. Here we assume the following functional forms:

$$\begin{aligned} U(c_1, c_2, z; \theta) =& G(c_1 + v(c_2, c_1) - \Psi(z/w_\theta)) \\ V(c_1, c_2, z; \theta) =& G(c_1 + v(c_2, c_1) - \gamma c_2 - \Psi(z/w_\theta)) \end{aligned}$$

This specification is similar to that in the previous example, with a key modification: the function v now has numeraire consumption c_1 as its second argument, rather than type θ.[17] The money-metric value of marginal bias $\gamma(\theta)$ for θ now differs slightly from the structural value of γ, with $\gamma(\theta) = \gamma / \left(1 + v'_{c_1}(c_2^*(\theta), c_1^*(\theta))\right)$; the difference is small provided that dependence of utility from c_2 on numeraire consumption only minimally changes the marginal utility of numeraire consumption.

[17] As a result, the expression for the marginal social welfare weight (defined as the social value of marginally more numeraire consumption for θ) is now $g(\theta) = G'\left(c_1^*(\theta) + v(c_2^*(\theta), c_1^*(\theta)) - \gamma c_2^*(\theta) - \psi(z^*(\theta)/w_\theta)\right)\left(1 + v'_{c_1}(c_2^*(\theta), c_1^*(\theta))\right)/\lambda$.

We let $\bar{g}$ denote the average social marginal welfare weights, which will be greater than one because income effects generate a fiscal externality from a lump-sum tax.

To characterize the optimal commodity tax in this case, consider raising the commodity tax by dt while adjusting the income tax to neutralize the effect on wealth, so that consumers do not change their choice of income.[18] The income tax reform must therefore reduce T_L by $dt \cdot c_2^*(L)$ and reduce T_H by $dt \cdot c_2^*(H)$. That reform exactly absorbs the funds generated mechanically by the commodity tax increase, so the only first-order effect on government revenues comes through the fiscal externality from the adjustment of c_2, equal to $\frac{t}{2}\left(\frac{dc_2^*(\theta)}{dt}\Big|_u + \frac{dc_2^*(\theta)}{dt}\Big|_u\right)$.[19] Note that these are compensated demand responses, since the income effect from the change in t has been offset by the income tax perturbation.

The only effect of this reform on consumer welfare is through changing the consumption of c_2, given by $\frac{1}{2}\sum_\theta \left(-g(\theta)\gamma(\theta)\ \frac{dc_2^*(\theta)}{dt}\Big|_u\right)$. Absent bias, this term would be equal to the zero due to the envelope theorem. Under the optimal tax system, the perturbation of the commodity and income taxes must have no first-order effect on welfare, requiring that the net effects on government revenues and consumer welfare sum to zero:

$$t\left(\sum_\theta \frac{dc_2^*(\theta)}{dt}\Big|_u\right) - \sum_\theta \left(g(\theta)\gamma\ \frac{dc_2^*(\theta)}{dt}\Big|_u\right) = 0.$$

Rearranging this expression, we find that the optimal commodity tax satisfies

$$t^* = \frac{\sum_\theta \left(g(\theta)\gamma\ \frac{dc_2^*(\theta)}{dt}\Big|_u\right)}{\sum_\theta \frac{dc_2^*(\theta)}{dt}\Big|_u}. \tag{2}$$

The optimal tax in this case is equivalent to the "corrective benefits" term in (1). However, the "regressivity costs" term does not carry through to this context. This is because when consumption differences are due to income effects, as in this example, an income-tax reform can (and, at the optimum, must) perfectly neutralize the effects of the commodity tax on income.

3.3 Discussion

These examples illustrate how redistributive concerns play a nuanced role in determining the optimal tax. Equation (1) illustrates how the optimal tax can be decomposed into two components representing corrective benefits and regressivity costs, only the former of which is relevant when consumption heterogeneity comes from income effects.

Although it is intuitive that redistributive concerns affect regressivity costs, it is perhaps more surprising that they also affect the size of corrective benefits. In the case of a constant elasticity of demand for c_2, redistributive concerns push the tax in the opposite direction from the standard

[18] In the previous example, since there were no income effects on labor supply, changes in the commodity tax did not affect labor supply, and thus no such neutralizing income tax reform was required.

[19] The notation $\frac{dc_j^*(\theta)}{dt}\Big|_u$ denotes a compensated demand response, with U held constant through a compensatory income adjustment.

regressivity concern: greater inequality aversion pushes toward a *larger* corrective tax. Intuitively, the corrective benefit is proportional to the change in c_2 consumption, and constant elasticity of demand implies that those who consume the most of the sin good (poor consumers) reap the largest benefits from correction. A more redistributive planner values this concentration of corrective benefits on the poor more heavily, and thus perceives a larger corrective benefit from taxation. This result stands in contrast to the common intuition that concentration of sin good consumption among the poor should dampen the optimal tax, and it highlights the importance of understanding the source of variation in the consumption of internality-producing goods across the income distribution.

When consumption differences are due to preference heterogeneity, redistributive concerns magnify both corrective benefits and regressivity costs and so the net effect is ambiguous. We generalize this idea, and discuss the intuition in more detail in Section 4.3.4. In the numerical simulations in Section 6 we illustrate that the optimal tax can indeed be either larger or smaller than the Pigouvian benchmark.

4 A general formula for the optimal commodity tax

4.1 Assumptions and elasticity concepts

We now turn to the general case with a continuum of types. For the remainder of the paper we impose the following assumptions, common in the optimal commodity taxation literature, in order to focus on the interesting features of sin taxes in a tractable context.

1. No labor supply misoptimization: $\frac{U_3}{U_1} = \frac{V_3}{V_1}$ for all types, at all allocations. That is, although there is a bias wedge between c_1 and c_2, there is no such wedge between c_1 and z. This assumption is not crucial, and we relax it in Section 5.5.

2. Constant marginal social welfare weights conditional on income: $g(\theta) = g(\theta')$ if $z(\theta) = z(\theta')$. This assumption is analogous to Assumption 1 in Saez (2002a). It holds immediately if types are homogeneous conditional on income. More generally, Saez (2002a) argues this is a reasonable normative requirement even under heterogeneity: "if we want to model a government that does not want to discriminate between different consumption patterns, it seems reasonable to assume that the primitive conditions on utilities and social weights have been specified so that Assumption 1 is true at the optimum."

3. U and V are smooth functions that are strictly concave in c_1, c_2, z, and μ is differentiable with full support.

4. The optimal income tax function $T(\cdot)$ is twice differentiable, and each consumer's choice of income z admits a unique global optimum, with the second-order condition holding strictly at the optimum.

Assumptions 3 and 4 ensure that the income distribution does not exhibit any atoms and individuals' labor supply and consumption decisions respond smoothly to perturbations of the tax system (see

Jacquet and Lehmann (2014) for the role of these assumptions). We represent the behavioral responses of $c_2(\theta)$ to taxes using the following parameters, which can be measured empirically:

- $\zeta(\theta, t, T)$: the price elasticity of demand for c_2 from type θ, formally equal to $-\left(\frac{dc_2(\theta,t,z)}{dt}\right)\frac{p+t}{c_2(\theta,t,T)}$.
- $\zeta^c(\theta, t, T)$: the compensated price elasticity of demand for c_2, equal to $-\left(\frac{dc_2(\theta,t,z)}{dt}\Big|_u\right)\frac{p+t}{c_2(\theta,t,T)}$.
- $\eta(\theta, t, T)$: the income effect on c_2, equal to $\zeta - \zeta^c$.

Additionally, we represent the labor supply response to tax reforms using the following parameters, which are defined formally in Appendix A:

- $\zeta_z^c(\theta, t, T)$: the compensated elasticity of taxable income with respect to the marginal income tax rate.
- $\eta_z(\theta, t, T)$: the income effect on labor supply.

These responses are defined to include the full sequence of adjustments due to any nonlinearities in the income tax. See Jacquet and Lehmann (2014) for an extensive discussion of such "circularities."

For each of the above parameters, we will sometimes suppress the arguments t and T, though endogeneity to the tax is acknowledged throughout. Moreover, it will prove convenient to average these statistics across all agents with a given income; such averages are denoted $\bar{c}_2(z)$, etc., with behavioral responses then denoted $\bar{\zeta}(z) = -\left(\frac{d\bar{c}_2(z)}{dt}\right)\frac{p+t}{\bar{c}_2(z)}$, and similarly for other parameters. Aggregate consumption of c_2 is denoted $C_2 = \int_\Theta c_2(\theta)d\mu(\theta)$, with aggregate elasticity of demand $\bar{\zeta} = -\left(\frac{dC_2}{dt}\right)\frac{p+t}{C_2}$, etc. The income distribution is denoted $H(z) = \int_\Theta 1\{z(\theta) \leq z\}\, d\mu(\theta)$, with density $h(z)$.

Finally, as has been suggested in Section 3, and as will be formalized in this section, it is important to distinguish between two sources of cross-sectional variation in $\bar{c}_2(z)$: income effects, and preference heterogeneity. We use $\bar{c}_2'(z)$ to denote the total change in $\bar{c}_2(z)$ with respect to z. This incorporates both that 1) a given θ-type may choose to consume a different amount of c_2 when his income goes up (the income effect) and that 2) the types of consumers earning z' rather than z have different tastes for c_2 (preference heterogeneity). The portion of this change that is explained by income effects is defined $\psi(z) := E\left[\frac{\partial}{\partial z}c_2(z,\theta)|z^*(\theta) = z\right]$, where $z^*(\theta)$ is the z-choice of a type θ. We set $\phi(z) := \int_{x=0}^{z}\psi(x)dx + c_2(0)$; intuitively, $\phi(z)$ is how much c_2 an average z-earner would consume if all variation in c_2 was explained solely by income effects. Indeed, $\bar{c}_2(z) = \phi(z)$ when preferences over c_1 and c_2 are homogeneous:

Lemma 1. *If for all z, $U_2(c_1, c_2, \theta, z)/U_1(c_1, c_2, \theta, z)$ does not depend on θ, then $\phi(z) = \bar{c}_2(z)$.*

We set $\tilde{\phi}(z) := \frac{\bar{c}_2(z)-\phi(z)}{C_2}$. This unitless index $\tilde{\phi}(z)$ measures how much of the difference between $\bar{c}_2(z)$ and $\bar{c}_2(0)$ is explained by preference heterogeneity, rather than income effects, normalized by average consumption of c_2.

4.2 An expression for the optimal commodity tax

To aggregate bias across individuals, we will follow Allcott et al. (2014) and Allcott and Taubinsky (2015) in defining the *average marginal bias* (i.e., bias weighted by compensated demand response):

$$\bar{\gamma}(t,T) = \frac{\int_\Theta \gamma(\theta,t,T)\left(\frac{dc_2(\theta,t,T)}{dt}\Big|_u\right) d\mu(\theta)}{\int_\Theta \left(\frac{dc_2(\theta,t,T)}{dt}\Big|_u\right) d\mu(\theta)}$$

as well as the average marginal bias among z-earners:

$$\bar{\gamma}(z,t,T) = \frac{\int_\Theta \gamma(\theta,t,T)\left(\frac{dc_2(\theta,t,T)}{dt}\Big|_u\right) 1\left\{z(\theta,t,T)=z\right\} d\mu(\theta)}{\int_\Theta \left(\frac{dc_2(\theta,t,T)}{dt}\Big|_u\right) 1\left\{z(\theta,t,T)=z\right\} d\mu(\theta)}$$

Intuitively, this aggregation represents the marginal bias weighted by individuals' marginal responses to a tax reform which raises the tax t while reforming T to offset the average effect on wealth at each income z. If dC_2 is the marginal change in total consumption of c_2 generated by such a tax reform, $\bar{\gamma}$ is the average amount by which consumers over- or under-estimate the change in utility from that change in consumption. We also define

$$\sigma := Cov_H\left[g(z), \frac{\bar{\gamma}(z)}{\bar{\gamma}}\frac{\bar{\zeta}^c(z)}{\bar{\zeta}^c}\frac{\bar{c}_2(z)}{C_2}\right]$$

as the covariance of welfare weight with consumption-weighted bias and elasticity. If bias and elasticity are invariant with income, then this is positive when low earners consume more c_2. In general, σ captures the extent to which the bias correction is concentrated on the low-end of the income distribution.

We proceed under assumption A below, which simplifies the ensuing equations but is not crucial for the derivation. We explore the effect of relaxing Assumption A in Appendix C.1, which provides guidance for the directional effect of non-zero income conditional covariances on the optimal tax. A particularly simple case in which assumption A holds is when heterogeneity is uni-dimensional; that is, when individuals earning the same income z have the same tastes for c_2.

Assumption A. *At each before-tax income z, the elasticities η, ζ^c, η_z, ζ^c_z, the bias γ, and consumption c_2 are all mutually independent at the optimal tax system.*

We begin by characterizing taxes not in terms of the social marginal welfare weights $g(z)$, but in terms of the social marginal utility of income $\hat{g}(z)$, defined (as in Farhi and Gabaix, 2015) as the average welfare effect of marginally increasing the incomes of consumers currently earning income z. The weights $\hat{g}(z)$ account for fiscal externalities resulting from income effects, and for the fact that some of this additional consumption will be mis-spent due to bias.[20] We then show that under empirically realistic assumptions, we can replace $\hat{g}(z)$ with $g(z)$ in the commodity tax formula. We conclude this section by providing formulas for the general case in terms of the weights $g(z)$.

[20] Formally, $\hat{g}(z) = g(z) - (1+\eta_z)\frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z) + \eta_z\frac{T'}{1-T'}$.

Proposition 1. *The optimal commodity tax t and income tax T satisfy the following at the optimum:*

$$t = \bar{\gamma}(\bar{g}+\sigma) - \frac{p+t}{\bar{\zeta}^c} Cov\left[\hat{g}(z), \tilde{\phi}(z)\right] \tag{3}$$

$$= \frac{\bar{\zeta}^c\bar{\gamma}(\bar{g}+\sigma) - pCov\left[\hat{g}(z), \tilde{\phi}(z)\right]}{\bar{\zeta}^c + Cov\left[\hat{g}(z), \tilde{\phi}(z)\right]} \tag{4}$$

$$\frac{T'(z)}{1-T'(z)} = \frac{g(z)\bar{\gamma}(z) - t}{p+t}\eta(z) + \frac{1}{\zeta_z^c z h(z)} \int_z^\infty (1-\hat{g}(x))dH(x)$$

Proposition 1 shows that the optimal tax is the combination of two terms. The term $\bar{\gamma}(\bar{g}+\sigma)$ corresponds to the corrective benefits of the tax. The corrective benefits are increasing in 1) the average marginal bias $\bar{\gamma}$, 2) the average social welfare weight $\bar{g}$ and 3) the extent to which bias correction is concentrated on the low-income consumers, σ.

The term $Cov\left[\hat{g}(z), \tilde{\phi}(z)\right]$ roughly corresponds to the regressivity costs of the tax that cannot be offset by progressive revenue recycling. Importantly, this term depends on the extent to which the variation of c_2 consumption by income is due to preference heterogeneity rather than income effects. When preference heterogeneity explains a constant portion κ of c_2 consumption, $\tilde{\phi}(z) = \kappa c_2(z)/C_2$, and regressivity costs are an increasing function of κ.

Equation (3), written in a form which echoes equation (2) from the binary types model in the previous section, illustrates that the optimal tax can be written as the corrective benefits per unit change in c_2 consumption—given by $\bar{\gamma}(\bar{g}+\sigma)$—less regressivity costs per unit of c_2 consumption—given by $\frac{p+t}{\bar{\zeta}^c} Cov\left[\hat{g}(z), \tilde{\phi}(z)\right]$. In (4) we rearrange so that t does not appear on the right side. The terms $\bar{\zeta}^c\bar{\gamma}(\bar{g}+\sigma)$ and $pCov\left[\hat{g}(z), \tilde{\phi}(z)\right]$ roughly correspond to the corrective benefits and regressivity costs per percent change in the price.

Equation (4) provides key guidance about the relative influence of the corrective benefits versus the regressivity costs in the optimal tax formula. The higher is the elasticity, the higher are the corrective benefits per unit change in the tax. Correspondingly, the relative importance of corrective benefits, as well as *how* sensitive the optimal tax t is to different levels of bias $\bar{\gamma}$ depends crucially on the price elasticity of demand—an issue we return to in subsection 4.3.6.

A complication in the formulas in Proposition 1 that lead to expressions involving $\hat{g}$ instead of g is the potential relevance of labor supply income effects η_z. However, Gruber and Saez (2002) find small and insignificant income effects and Saez et al. (2012) review the empirical literature on labor supply elasticities and argue that "in the absence of compelling evidence about significant income effects in the case of overall reported income, it seems reasonable to consider the case with no income effects." We now consider this case and also simplify further by assuming that consumers' expenditures on c_2 are sufficiently small to have a negligible impact on the marginal utility from consumption of c_1, rendering income effects on c_2 negligible as well.

Corollary 1. *Assume that $\eta_z \approx 0$. And assume, additionally, that terms $O\left(c_2 \frac{(p+t)U_{11}-U_{12}}{U_1}\right)$ are negligible; that is, consumers' expenditures on c_2, given by $(p+t)c_2$, are sufficiently small as to have a negligible impact on the marginal utility from the numeraire good c_1. Then*

$$t \approx \frac{\bar{\zeta}^c \bar{\gamma}(\bar{g}+\sigma) - pCov\left[g(z), \tilde{\phi}(z)\right]}{\bar{\zeta}^c + Cov\left[\hat{g}(z), \tilde{\phi}(z)\right]}$$

The corollary shows that that under the simplifying assumptions, the optimal commodity tax formula in Proposition 1 can be simplified by replacing $\hat{g}(z)$ with g when income effects are small relative to income inequality. This simplification is exactly accurate in the special cases of pure preference heterogeneity (since income effects are nonexistent) and pure income effects (since the regressivity costs term is then zero). Moreover, we show in Appendix F.3 that this approximation is very accurate numerically even outside these special cases. For brevity, we defer the intermediate case, in which $\eta_z \approx 0$ but expenditures on c_2 are large, to Appendix C.2.

Of course, the assumption of negligible income effects is not necessary to provide optimal tax formulas that are expressed in terms of the social marginal welfare weights. The key variable needed for this solution is the following: "When we raise the commodity tax by dt, how much money can we give back to each person to offset the tax so as to keep the average labor supply choices of each z-earner constant?" Roughly, we seek to characterize the extent to which we can offset the regressivity costs of the commodity tax by making the income tax more progressive. We begin with a lemma providing this characterization:

Lemma 2. *Let* $\chi(z) := \phi(z) - \int_0^z w(x,z)\frac{\eta_z}{\bar{\zeta}^c_z x}(c_2(x) - \phi(x))dx$, *where* $w(x,z) = e^{\int_{z'=x}^{z'=z} -\frac{\eta_z}{\zeta^c_z z}dx'}$. *Then increasing the commodity tax by dt and decreasing the income tax by* $\chi(z)dt$ *leaves the average labor supply of z-earners unchanged.*

Note that $\chi(z) = \phi(z)$ when $\eta_z = 0$. That is, in the absence of labor supply income effects, the term $\phi(z)$—which we defined as the portion of c_2 consumption explained by consumption income effects—is the extent to which the income tax should be optimally lowered for z-earners. When $\eta_z \neq 0$ and $\phi(z) < c_2(z)$, $\chi(z)$ will be higher than $\phi(z)$. Intuitively, this is because giving back $\phi(z)dt$ to consumers offsets the effective impact on the marginal keep-rate from before-tax earnings, but does not fully offset the income effect as it leaves consumers poorer by an amount $(c_2 - \phi)dt$. Analogous to $\tilde{\phi}(z)$, we define the index $\tilde{\chi}(z) := \frac{\bar{c}_2(z) - \chi(z)}{C_2}$.

Proposition 2. *The optimal commodity tax t satisfies*

$$t = \overbrace{\bar{\gamma}(\bar{g}+\sigma)}^{\textit{corrective benefits}} + \overbrace{\frac{p+t}{\bar{\zeta}^c}E[(1-g(z))\tilde{\chi}(z)]}^{\textit{regressivity costs}} - \overbrace{\frac{1}{\bar{\zeta}^c}\int \tilde{\chi}(z)\eta(z)(t - g(z)\bar{\gamma}(z))}^{\textit{additional impact from income effect}} \tag{5}$$

Proposition 2 provides a commodity tax formula similar to Corollary 1, with two differences. First, the covariance term $Cov[g(z), \tilde{\phi}(z)]$ corresponding to regressivity costs is written more generally as $E[(g(z)-1)\tilde{\chi}(z)]$. These two terms are equivalent when $E[g(z)] = 1$, but differ when $E[g(z)] > 1$ in the presence of labor supply income effects ($\eta_z \neq 0$). Second, the presence of income effects η means that a tax reform that decreases consumers' after-tax income by $c_2(z) - \chi(z)$ also impacts consumption of c_2, beyond the compensated elasticity with respect to prices. This generates fiscal externalities

proportional to $t\eta(c_2(z) - \chi(z))$ and additional corrective benefits proportional to $g\bar{\gamma}(c_2(z) - \chi(z))$. Combined, this generates the third term in equation (5) above. In the absence of income effects, the formula reduces to

$$t = \bar{\gamma}(\bar{g} + \sigma) - \frac{p+t}{\bar{\zeta}^c} Cov\left[g(z), \tilde{\phi}(z)\right],$$

which can be solved for $t$to get the formula in Corollary 1.

4.3 Interpretation and implications

4.3.1 Optimal taxes in the absence of redistributive concerns

We begin by relating the general formula in Proposition 1 to known sin tax results, derived under the assumptions of no redistributive concerns.

Corollary 2. *Suppose that either 1) $z(\theta)$ is constant in θ or 2) $g(\theta) = 1$ for all θ. Then $t^* = \bar{\gamma}$.*

Corollary 2 considers two special cases that correspond to existing work on sin taxes. In the first case, we assume that there is no wealth inequality: $z(\theta)$ is constant in θ. In the second case, we allow for differences in wealth, but assume that utility is quasilinear and that the policymaker has no aversion to this wealth inequality. In both cases, the optimal commodity must exactly offset the average marginal bias. The first case is a slight generalization of known sin tax results (e.g., O'Donoghue and Rabin 2006; Mullainathan et al. 2012; Allcott et al. 2014; Allcott and Taubinsky 2015) to the case of income effects, while the second case is a slight generalization of known sin tax results to situations in which the policymaker also has a nonlinear income tax at his disposal.

4.3.2 Optimal taxes in the absence of corrective concerns

Another special case of our framework is the case in which there are no corrective concerns. Proposition 1 implies that for this case,

$$t = -\frac{pCov\left[\hat{g}(z), \tilde{\phi}(z)\right]}{\bar{\zeta}^c + Cov\left[\hat{g}(z), \tilde{\phi}(z)\right]}, \tag{6}$$

and Proposition 2 can similarly be used to obtain a formula in terms of $g(z)$ rather than $\hat{g}(z)$. Equation (6) generalizes the seminal Atkinson-Stiglitz theorem to the case of arbitrary preference heterogeneity. The Atkinson-Stiglitz theorem itself obtains as a special case of (6) when all variation in c_2 consumption is driven by income effects, which then implies that $t = 0$. Additionally, formula (6) generalizes the qualitative analyses about the usefulness of commodity taxes by Saez (2002b) and Edwards et al. (1994), the results for two-type models by Edwards et al. (1994) and Nava et al. (1996), and results for homogeneous preferences by Jacobs and Boadway (2014).

4.3.3 Optimal taxes when income effects do not affect c_2 consumption

Consider next the case in which all differences in c_2 consumption are due to differences in tastes; that is, there are no income effects.

Corollary 3. *Suppose that there are no income effects: $\eta \equiv 0$ and $\eta_z \equiv 0$. Then*

$$t^* = \underbrace{\bar{\gamma}\left(\bar{g} + \sigma\right)}_{\textit{Corrective benefits}} - \underbrace{\frac{p + t^*}{\bar{\zeta}^c} Cov[g(z), \bar{c}_2(z)]}_{\textit{Regressivity costs}} \tag{7}$$

The formula, which generalizes our results from Example 1 in Section 3, consists of two terms. The first term—corrective benefits—reflects the extent to which changing consumer demand for c_2 increases welfare because it counteracts consumer biases. Note that the magnitude of the first term is now dependent on σ, the extent to which the corrective benefits are concentrated on the low-income consumers.

The second term—regressivity costs—reflects the extent to which the tax (or subsidy) is regressive and contributes to wealth inequality. This term persists in the formula despite the policymakers' ability to offset the commodity regressivity using a more progressive income tax. This is because making the income tax more progressive in this case will generate fiscal externalities that outweigh its redistributive benefits.

To provide some intuition for the second term, consider the case in which low-wage consumers also have a higher preference for c_2. In this case, a positive tax falls more heavily on the low-income consumers, making the total tax burden less progressive. Despite the fact that the total tax burden has become less progressive, however, consumers are not more motivated to increase their labor supply—precisely because all differences in c_2 are due to heterogeneous preferences, rather than differences in wealth. Thus, any tax reform that increases a regressive sin tax and attempts to offset that by making the income tax more progressive will generate fiscal externalities by lowering overall labor supply in the economy.

A stark implication of (7) is that the optimal sin tax will be negative (a sin subsidy) when the consumers who over-purchase the good are inelastic to the tax. This captures the spirit of a key result of Bernheim and Rangel (2004) about the optimality of subsidizing addictive goods when the marginal utility of income is increasing with the consumption of the addictive good.[21] Although the Bernheim and Rangel (2004) result that the sin good should be subsidized is seemingly in stark contrast to the sin tax results in, e.g., O'Donoghue and Rabin (2006), our general tax formula clarifies the economic forces that lead each result.

At the same time, our results also extend this earlier work by allowing the policymaker to cary out redistribution through the income tax. Although Corollary 3 would continue to hold exactly if the policymaker did not have access to an income tax, or if the income tax was constrained to be fixed at particular levels,[22] we now discuss in more detail how the optimal commodity tax deviates from this formula when preferences heterogeneity does not explain all variation in c_2 consumption.

[21] Because in Bernheim and Rangel (2004) over-consumption of the good is a consequence of cue-triggered neural processes that render the individual inelastic to prices, the average bias of consumers who are elastic to the tax is zero. Thus, their subsidy result is a direct consequence of the positive covariance between consumption of the addictive good and the marginal utility from income—analogous to our "regressivity costs term.

[22] Note that more generally, we do not need to assume that the commodity tax authority have full control over the income tax per se—they could instead employ other progressive revenue recycling policies to offset regressivity, such as Philadelphia's promise to direct a share of SSB tax revenues toward targeted pre-K programs.

4.3.4 Optimal taxes when all differences in c_2 consumption are due to income effects

When differences in consumption are due purely to differences in wealth, any regressive consequences of a tax or subsidy can be perfectly offset by changes to the income tax, in such a way that the there are no fiscal externalities resulting from the full policy reform.

Corollary 4. *Suppose that $U_2(c_1, c_2, \theta, z)/U_1(c_1, c_2, \theta, z)$ is constant in θ for each z. Then*

$$t^* = \bar{\gamma}\,(\bar{g} + \sigma) \tag{8}$$

Formula (8) generalizes the results in our Example 2 in Section 3, and shows that when preferences for c_2 are homogeneous, the optimal commodity tax is proportional to the average marginal bias, irrespective of how consumption of c_2 varies across the income distribution.

Note, however, that redistributive concerns still play a role in the size of the optimal commodity tax, and are reflected in the σ term. The more elastic the low-income consumers are, and the more biased they are relative to the high income consumers, the higher is σ and thus the higher is the benefit of bias correction. The intuition for this is a consequence of two principles. First, as we illustrated in Example 2, the mechanical wealth effect of the commodity tax can be directly offset by making the income tax more progressive. The welfare effects of this tax reform can thus be decomposed into only two effects: 1) bias correction and 2) the fiscal externality from lowering consumption of c_2. The social welfare benefits of bias correction are shaped by the following principle: a policymaker with redistributive motives should be willing to spend more than \$1 to eliminate a \$1 mistake made by the poor consumers. Thus, corrective motives are amplified when "bias correction" is concentrated on the low-income consumers—either because they are more biased or because they are more elastic to the income tax. Conversely, when bias correction is concentrated on high-income concerns, a policymakers' motives for correcting that bias will be diminished.

4.3.5 Redistributive motives can amplify corrective benefits

The above discussion underscores a broader point: although in some cases redistributive motives can lower the size of the optimal tax because of regressivity concerns, in other cases redistributive motives can actually amplify the size of the optimal tax if bias correction is concentrated on the low income consumers. As we discuss next, redistributive concerns are most likely to amplify the size of the optimal tax when the elasticity is high.

4.3.6 The key role of the price elasticity of demand in determining the importance of corrective benefits

Proposition 1 shows that as that elasticity grows large, the optimal tax t approaches $\bar{\gamma}(\bar{g}+\sigma)$, which is the formula for the optimal tax when there are no regressivity costs due to heterogeneous preferences. In this case, if $\sigma > 0$ then the size of the optimal tax will be higher than what it would be in the absence of redistributive concerns.

Intuitively, as the elasticity grows large, the corrective benefits per unit change in the tax, as well as the fiscal externalities from substitution away from c_2, grow very large. On the other hand, regressivity costs per unit change in the tax—which correspond to how much of the mechanical revenue effect from the unit increase in the tax falls on the low income consumers—do not change with the size of the tax. Thus, when the elasticity is large, the regressivity costs become negligible, and the optimal tax is determined by the tradeoff between corrective benefits and fiscal externalities from substitution.

At the opposite extreme, as the elasticity grows small, corrective benefits become negligible compared to the regressivity costs, and the optimal tax approaches the formula in equation (6). If preference heterogeneity accounts for *any* share of the decrease in c_2 consumption across incomes, then a sufficiently low elasticity implies the optimal tax becomes has to be negative (a subsidy) whenever it is regressive. Intuitively, if individuals do not respond to commodity taxes, then such taxes become a powerful instrument to enact redistribution through targeted subsidies.

More generally, the results clarify that the role of consumer bias in shaping both the sign and magnitude of the optimal commodity tax is modulated by the price elasticity of demand. But perhaps most importantly, the elasticity also provides practical guidance on how *sensitive* the optimal tax is to different values of the bias $\bar{\gamma}$. Consider, for example, two different estimates of bias in sugary drink consumption: 5 cents per ounce versus 3 cents per ounce. For a high elasticity, the high estimate implies an optimal tax that is almost 2 cents per ounce higher than under the low estimate. For a low elasticity, the difference in the estimates could translate to a difference in taxes that is, e.g., only 1 cent.

4.3.7 The importance of decomposing cross-sectional variation in c_2 into income effects and preference heterogeneity

As the two special cases we have examined so far show, whether the cross-sectional variation of c_2 is explained by preference heterogeneity or income effects plays a crucial role in whether concerns about regressivity are reflected in commodity taxes. In practice, of course, both preference heterogeneity and income effects like play a role in explaining the variation in consumption of the sin good. In the case of energy efficiency subsidies, both of these likely contribute to the fact that higher-income consumers purchase more energy efficient products. In the case of sin goods such as cigarettes or unhealthy foods, both likely play a role as well, though it is entirely possible that the sin goods are normal goods, meaning that income effects actually decrease the cross-sectional variation that we observe.

Our general results in Propositions 1 and 2 provide a quantitative, elasticity-based formula that covers all of these cases. Broadly speaking, the more that c_2 variation across incomes is driven by income effects, the smaller the magnitude of regressivity costs.

But at the same time, our results also imply that the regressivity costs may be higher than what many would conjecture. If c_2 is a normal good and yet consumption of c_2 decreases with income, our

results show that regressivity costs are actually larger than $Cov[\hat{g}(z), \bar{c}_2(z)]$,[23] since the commodity both falls more heavily on the poor than the rich *and* decreases labor supply.

5 Extensions

5.1 Tax Salience on the labor supply Margin

Most sin taxes take the form of excise taxes that are incorporated into the final posted price of the good, and are thus likely to be fully salient at the time of purchase. However, the assumption that people can correctly compute how various commodity taxes—or differences in prices, more generally—determine the implicit marginal tax rate on their labor income may be too demanding. A tax on a particular luxury good, for example, indirectly decreases the returns from labor supply; but in practice, its effect may be smaller than that of an equivalent, but direct, increase in marginal income tax rates on wealthy consumers.

A reasonable hypothesis is that the effects of commodity taxes on labor supply may be minimal, even when the variation consumption is due to income effects rather than preference heterogeneity.[24] If that is the case, then offsetting the effects of regressive commodity taxes by making the income tax more progressive will be undesirable even when there is little preference heterogeneity.

Formally, we consider consumers whose perceived price of c_2 is given by $(1-\vartheta)\hat{p} + \vartheta(p+t)$ when they are making a labor supply decision, where $\hat{p}$ is a "mental default" and $p+t$ is the actual after-tax price. The attention parameter ϑ captures the extent to which consumers' labor supply is sensitive to variations in prices generated by the commodity tax. However, we assume that prices are understood when consumers choose how to spend their after-tax earnings $z - T(z)$ on c_1 and c_2.

We provide a fully general characterization of the optimal commodity tax, as well as a derivation of the optimal income tax, in Appendix B.5. In this section, we discuss the results for the commodity tax in the illustrative case in which ϑ does not vary by z and $\eta_z = 0$.

Proposition 3. *Suppose that ϑ is constant and $\eta_z = 0$. Then the optimal commodity tax satisfies*

$$\begin{aligned} t &= \bar{\gamma}(\bar{g}+\sigma) - \frac{p+t^*}{\bar{\zeta}^c} Cov\left[\hat{g}(z), (1-\vartheta)c_2(z) + \vartheta\tilde{\phi}(z)\right] \\ &= \frac{\bar{\zeta}^c\bar{\gamma}(\bar{g}+\sigma) - pCov\left[\hat{g}(z), (1-\vartheta)c_2(z) + \vartheta\tilde{\phi}(z)\right]}{\bar{\zeta}^c + Cov\left[\hat{g}(z), (1-\vartheta)c_2(z) + \vartheta\tilde{\phi}(z)\right]} \end{aligned}$$

Proposition 3 show that the less people consider commodity taxes on the labor supply margin, the closer the formula is to the case of pure preference heterogeneity.

The result is particularly stark in the case of $\vartheta = 0$, leading to the following corollary of the derivations in Appendix B.5:

[23] This is because the term $\phi(z)$, corresponding to how much variation is explained by income effects, is actually negative in this case.

[24] A possible micro-foundation for this is along the lines of the sparsity-based bounded rationality model of Gabaix (2014). Because expenditures on any one of many dimensions of consumption are relatively small, while the total tax burden paid through the income tax is relatively high, the framework of Gabaix (2014) would predict that people pay less attention to changes in taxes or prices on particular commodities than they do to changes in the income tax.

Corollary 5. *Suppose that $\vartheta = 0$ for all individuals. Define $\bar{\gamma}^u$ and σ^u analogously to $\bar{\gamma}$ and σ, but with respect to uncompensated demand responses. Then the optimal commodity tax satisfies*

$$\begin{aligned} t &= \bar{\gamma}^u(\bar{g} + \sigma^u) - \frac{p+t}{\bar{\zeta}^c} Cov\left[g(z), c_2(z)\right] \\ &= \frac{\bar{\zeta}^c \bar{\gamma}^u(\bar{g} + \sigma^u) - pCov\left[g(z), c_2(z)\right]}{\bar{\zeta}^c + Cov\left[g(z), c_2(z)\right]}. \end{aligned}$$

In this case, even under the standard assumptions of the Atkinson-Stiglitz theorem—that preferences are homogeneous, so that income effects explain all variation in c_2 consumption and $\tilde{\phi} = 0$—the regressivity costs will still be given by $Cov[\hat{g}(z), c_2(z)]$. Intuitively, if labor supply does not respond to differences in the commodity tax, then the welfare effects of increasing the commodity tax depend only on 1) its corrective benefits and 2) the extent to which it is regressive. Because the commodity tax has no effect on the labor supply margin, any attempts to offset its regressivity through the income tax would generate additional fiscal externalities.[25]

Beyond the implications for sin taxes, these results also underscore that the Atkinson-Stiglitz theorem depends crucially on the assumption that consumers can perfectly calculate how changes in commodity taxes affect their returns to labor supply.

5.2 $N > 2$ Dimensions of Consumption

Our results easily generalize to $N > 2$ dimensions of consumption. As before, we treat c_1 as the numeraire, and assume that $U_1/U_{N+1} = V_1/V_{N+1}$. Let $\mathbf{S}$ denote the Slutsky matrix of compensated demand responses in which the j, i entry is the compensated demand response of c_j with respect to t_i.

Let $\gamma_i(\theta)$, for $i \geq 2$, denote the bias in consumption dimension i, and let $\bar{\gamma}_i(z)$ denote the statistic for z-earners. Let $\bar{\gamma}_{ij}$ denote the average marginal bias from consumption of c_j with respect to tax i:

$$\bar{\gamma}_{ij} = \frac{\int_\Theta \gamma_j(\theta, t, T) \left(\frac{dh_j(\theta,t,T)}{dt_i}\right) d\mu(\theta)}{\int_\Theta \left(\frac{dh_j(\theta,t,T)}{dt_i}\right) d\mu(\theta)}$$

where h_j is the compensated (Hicksian) demand for good j. We also define $\sigma_{ij} := Cov_H \left[g(z), \frac{\bar{\gamma}(z)}{\bar{\gamma}} \frac{\frac{dh_j(\theta,t,T)}{dt_i}}{\frac{dh_j(\theta,t,T)}{dt_i}} \right]$. We define $\mathbf{R}$ to be a column vector in which the jth entry is $R_i = -\sum \frac{dh_j}{dt_i} \bar{\gamma}_{ij}(\bar{g} + \sigma_{ij})$. That is, R_i is the total corrective benefit from increasing tax t_i, keeping income constant.

As before, we define $\phi_j(z)$ to be the portion of c_j consumption explained by income effects, and we define $\tilde{\phi}_j(z) := (c_j - \phi_j)/C_j$ to be the variation due to preference heterogeneity. We define $\boldsymbol{\rho}$ to be the column vector in which the jth entry is $Cov[\hat{g}(z), \tilde{\phi}_j(z)]$.

Proposition 4. *The optimal commodity taxes $\mathbf{t} = (t_1, \ldots, t_N)$ satisfy*

$$\mathbf{tS} = \boldsymbol{\rho} - \mathbf{R} \tag{9}$$

[25] Moreover, the formula for the commodity tax in Corollary 5 would continue to hold even when

and the optimal income tax satisfies

$$\frac{T'(z)}{1-T'(z)} = \sum_i \frac{g(z)\bar{\gamma}_i(z) - t_i}{p_i + t_i}\eta_i(z) + \frac{1}{\zeta_z^c z h(z)} \int_z^\infty (1-\hat{g}(x))dH(x) \tag{10}$$

The system of linear equations in (9) can be easily solved for the optimal commodity taxes using Cramer's rule. As an example, consider the case of three commodities, where c_1 is the numeraire as before. Solving the system of linear equations in (9) using Cramer's rule yields

$$t_2 = \frac{-(R_2-\rho_2)\frac{dh_3}{dt_3} + (R_3-\rho_3)\frac{dh_3}{dt_2}}{D} \tag{11}$$

$$t_3 = \frac{-(R_3-\rho_3)\frac{dh_2}{dt_2} + (R_2-\rho_2)\frac{dh_2}{dt_3}}{D} \tag{12}$$

where $D := \frac{dh_2}{dt_2}\frac{dh_3}{dt_3} - \frac{dh_2}{dt_3} \cdot \frac{dh_3}{dt_2} > 0$.

Equations (11) and (12) generate several economic insights. Assume for simplicity that $\bar{g} = 1$ and that γ_2 and γ_3 are homogeneous across incomes. Then the expressions above simplify to

$$t_2 = \frac{\gamma_2 D + \rho_2\frac{dh_3}{dt_3} - \rho_3\frac{dh_3}{dt_2}}{D} \tag{13}$$

$$t_3 = \frac{\gamma_3 D + \rho_3\frac{dh_2}{dt_2} - \rho_2\frac{dh_2}{dt_3}}{D} \tag{14}$$

We now consider two special cases with policy relevance. First, consider a stylized example in which c_2 represents regular soda, consumed primarily by poorer consumers ($\rho_2 > 0$) and c_3 represents diet soda, consumed more heavily by richer consumers ($\rho_3 < 0$). Suppose also that $R_2 - \rho_2 > 0$, meaning that corrective benefits outweigh regressivity costs when taxing good 2, while $-\rho_3 + R_3 > 0$, meaning that the progressivity benefits of taxing good 2 outweigh the indirect corrective costs from consumers substituting from good 3 to good 2. In this case it is straightforward to verify that, beginning from the case where $\frac{dh_2}{dt_3} = \frac{dh_3}{dt_2} = 0$ (independence) for which $t_2 > 0$ and $t_3 > 0$, the first-order effect of introducing a small degree of substitutability is to reduce t_3 while raising t_2.[26] Thus although a positive tax on diet soda may be optimal for redistributive reasons, the first order effect of substitutability between diet and regular soda reduces the magnitude of the tax on diet soda and magnifies the optimal tax on regular soda.[27] (For larger degrees of substitutability, this first-order effect need not dominate, so the ultimate effect of substitutability could be ambiguous.)

As a second special case, consider a stylized example in which c_2 represents regular tobacco cigarettes, and c_3 represents electronic cigarettes which, though overconsumed, are less harmful than regular cigarettes ($\gamma_2 > \gamma_3$). Both are consumed more heavily by poorer consumers ($\rho_2, \rho_3 > 0$). In this case, beginning again from the case of independence, the effect of substitutability is to lower the optimal tax on both goods. Thus, the emergence of e-cigarettes suggests, perhaps somewhat

[26] To see this, let $x = \frac{dh_3}{dt_2} = \frac{dh_2}{dt_3}$ (the latter equality is implied by symmetry of the Slutsky matrix), and compute the derivatives of each tax with respect to x, evaluated at $x = 0$: $\frac{dt_2}{dx}\big|_{x=0} = \frac{-\rho_3}{D}$ and $\frac{dt_2}{dx}\big|_{x=0} = \frac{-\rho_2}{D}$.

[27] This result is relevant for Philadelphia's SSB tax, which applies equally both to sugar-sweetened beverages and "diet beverages" with artificial sweeteners.

counterintuitively, that taxes on tobacco cigarettes should be lower.

5.3 Externalities

Of course, many sin goods also generate externalities. Driving gas guzzlers, for example, contributes to carbon emissions, while smoking cigarettes may decrease the welfare of those in the vicinity of the smoker. And consumption of unhealthy goods generates fiscal externalities arising from medical expenses that are not fully internalized by individuals with insurance.

To see how how externalities are a special case of our framework, suppose that marginal social damages from consumption of c_2 are given by $\mathcal{E}(C_2)$. Individuals maximize U, while the policymaker wants them to maximize $V = U - \mathcal{E}'(C_2)c_2/\alpha(\theta)$. A simple corollary of Proposition 1 is that the optimal commodity tax is then given by

$$t = \frac{\bar{\zeta}^c \mathcal{E}'(C_2) - pCov\left[\hat{g}(z), \tilde{\phi}(z)\right]}{\bar{\zeta}^c + Cov\left[\hat{g}(z), \tilde{\phi}(z)\right]} \tag{15}$$

Formula (15) above generalizes Jacobs and de Mooij (2015) and Kaplow (2012), both of which derive the optimal externality-correcting tax under the assumption that preferences are homogeneous. More generally, in the presence of both externalities and internalities, the optimal tax is given by

$$t = \frac{\bar{\zeta}^c(\bar{\gamma}(\bar{g} + \sigma) + \mathcal{E}'(C_2)) - pCov\left[\hat{g}(z), \tilde{\phi}(z)\right]}{\bar{\zeta}^c + Cov\left[\hat{g}(z), \tilde{\phi}(z)\right]}.$$

In addition to generalizing our main results, the formula in (15) also clarifies a key difference between externalities and internalities. In the case of consumer bias, it is necessary to know whether the bias is bigger for the rich or for the poor, and whether the rich or the poor are more elastic. With externalities—even if they are heterogeneously produced in the population—analogous questions are irrelevant. It is not necessary to know 1) whether marginal damages generated by the poor are higher than marginal damages generated by the rich or 2) how the price elasticity of c_2 consumption varies for the rich versus the poor.

5.4 Optimal tax formulas without the first-order approach

Throughout, we have assumed that the optimal tax system can be characterized by first order conditions. This assumption is made almost universally in the optimal taxation literature, including work on commodity taxation such as Saez (2002) or Jacobs and Boadway (2014). In Appendix D, however, we show that this assumption can be relaxed.

There are three special cases in which our results hold verbatim even without the differentiability assumption. The first case is when there are no income effects: $\eta_z = \eta = 0$. In this case, a commodity has no impact on labor supply, and the formula in Corollary 7 continues to hold. Similarly, if the commodity tax is completely not salient on the labor supply margin, as in Section 5.1 with $\vartheta = 0$,

then the commodity tax will again have no impact on labor supply, and the formula in Corollary 5 will continue to hold.

Third, when all variation of c_2 is explained by income effects, labor supply will always remain unchanged if an increase dt of the commodity tax is followed by a decrease $dtc_2(z)$ in the income tax $T(z)$. In this case, the formula in Corollary 8 continues to hold.

Finally, in Appendix D, we provide a general solution for the case of uni-dimensional θ. The case with no labor supply income effects is particularly simple. For $z \in \lim_{d\theta \to 0}[z(\theta), z(\theta + |d\theta|)]$, define

$$\omega(z) = \lim_{d\theta \to 0} \frac{c_2(\theta + |d\theta|, z(\theta + |d\theta|)) - c_2(\theta + |d\theta|, z(\theta))}{z(\theta + |d\theta|) - z(\theta)}$$

At all values θ at which z is differentiable, $\omega(z) = \phi(z)$. But when earnings z are discontinuous in θ, $\omega(z)$ is simply the slope of the line connecting $(z(\theta), c_2(z(\theta)))$ to $(z(\theta + d\theta), c_2(z(\theta + d\theta)))$. Now define $\chi(z) = c_2(0) + \int_{x=0}^{z} \omega(z)dz$. The solution is then exactly as in Proposition 2, using this more general definition of $\chi(z)$.

5.5 Labor supply misoptimization

It is also possible to relax the assumption that $U_1/U_3 = V_1/V_3$. Labor supply misoptimization modifies the optimal income tax formula to allow to correct for labor supply internalities, as shown in Farhi and Gabaix (2015), Lockwood (2016), and Gerritsen (forthcoming). Labor supply misoptimization does not alter the optimal commodity tax in two special cases: (1) c_2 consumption differences come from preference heterogeneity, or (2) the commodity tax t is not salient when labor supply is chosen (in this case, the formula from Corollary 5, employing uncompensated demand elasticities, should be used). More generally, the same formulas could be applied under labor supply misoptimization in the presence of income effects, with the exception that $\hat{g}(z)$—the welfare impact of giving z-earners one more unit of post-tax income—must be modified to incorporate the corrective benefit (or cost) from any change in labor supply behavior from this income shock, in addition to the usual fiscal externalities.

6 Numerical application: A tax on sugar-sweetened beverages

We now turn to the quantitative implications of our theoretical results in the context of a prominent current policy debate: a tax on soda and other sugar-sweetened beverages (SSBs). As we will discuss below, consumption of SSBs declines across the income distribution and there is a substantial body of research in the public health literature documenting the large health costs associated with SSBs (and sugar consumption generally). Although SSB taxes have thus far been levied only at the city-level, here we simulate the welfare impacts of a hypothetical federal SSB tax, analogous to the federal cigarette tax that is currently in place in the US.

Because little is known about consumer biases in SSB consumption—and because there is no consensus on what the elasticity of SSB consumption is or on how much of the cross-sectional variation in consumption is due to preference heterogeneity versus income effects—our simulations embrace

this uncertainty and calculate the shape of the optimal tax as a function of these key parameters. This allows us to quantify the key comparative statics from our theory—e.g., the central role of the elasticity and preference heterogeneity—and underscores the importance of empirical measurement of these statistics.

6.1 Empirical model

We employ the following functional forms for decision and experienced utility, extending the simple examples in Section 3:

$$\begin{aligned} U(c,s,z;\theta) =& c + v(s,c,\theta) - \Psi\left(\frac{z}{w_\theta}\right) \\ V(c,s,z;\theta) =& c + v(s,c,\theta) - \tilde{\gamma}s - \Psi\left(\frac{z}{w_\theta}\right) \end{aligned}$$

Consumption of SSBs (in ounces) is denoted s, and for simplicity we drop the subscript on numeraire consumption. Bias $\tilde{\gamma}$ denotes the unanticipated internality cost of from each ounce of SSB consumption. Disutility of labor effort is given by

$$\Psi\left(\frac{z}{w}\right) = \frac{1}{1+1/\epsilon}\left(\frac{z}{w_\theta}\right)^{1+1/\epsilon},$$

and subutility of SSB consumption is given by

$$v(s,c,\theta) = a(c,\theta)\left(\frac{s^{1-k}}{1-k}\right).$$

This functional form flexibly allows for heterogeneous levels of SSB consumption that can be driven by either pure preference heterogeneity (in which case a varies only with θ), pure income effects (in which case a varies only with c), or by some combination of the two. The parameter k controls the elasticity of demand for soda, which is $1/k$ in the case of pure preference heterogeneity.[28] To parameterize the relative influence of preference heterogeneity and income effects on soda consumption, we assume $a(c,\theta)$ is the product of two univariate functions $a_c(c)$ and $a_\theta(\theta)$—one varying solely with consumption, the other solely with type—with the parameter φ governing the weight on income effects:

$$a(c,\theta) = a_c(c)^{\varphi} a_\theta(\theta)^{(1-\varphi)}$$

Thus $\varphi = 0$ and $\varphi = 1$ correspond to the special cases where all SSB variation is driven by either preference heterogeneity or income effects, respectively. Intermediate cases with $\varphi \in (0,1)$ attribute a portion of SSB variation to each channel. We also consider $\varphi < 0$, which corresponds to positive income effects, so that SSBs are a normal good; in this case, the taste for SSBs must decline sufficiently steeply across incomes to more than compensate for the positive income effects. To relate

[28] One can allow the elasticity to vary across the income distribution by allowing k to vary with c and θ. We omit that possibility here for simplicity.

the structural parameter φ to the more economically interpretable parameters used in Section 4, we define the aggregate statistic Φ to represent the average amount of cross-sectional SSB variation which is locally driven by income effects:

$$\Phi = \mathbb{E}\left[\frac{\psi(z)}{c_2'(z)}\right] = \int \left(\frac{\psi(z)}{c_2'(z)}\right) dH(z),$$

so that $\Phi = 0$ and $\Phi = 1$ correspond $\varphi = 0$ and $\varphi = 1$, respectively.

For the purposes of crisp exposition of the comparative statics, we assume that the pareto weights $\alpha(\theta)$ are not endogenous to the tax system itself, and are given by $\alpha(\theta) = (c_{US}(\theta))^{-1}$, where c_{US} represents total consumption (aggregating numeraire and soda consumption) under the US status quo in the calibration. Appendix F.2 explores the sensitivity of our results to the strength of redistributive preference. We derive analytic expressions for elasticities and income effects using the specification above in Appendix E.

6.2 Calibration and data

The income distribution is drawn from CPS, equally weighting the 2001–2010 March survey waves, restricted to households with positive total income and two children. We assume an elasticity of taxable income of 0.33, the preferred value from Chetty (2012).[29] Following Saez (2001) and Saez (2002b), we calibrate the skill distribution so that it generates the observed income distribution in the presence of the status quo tax code, which we approximate for simplicity using a linear tax function with a marginal tax rate of 40% and a lump sum grant of $6000. Results are not sensitive to these values.

We calibrate $a(c, \theta)$ to match the pattern of soda consumption across the income distribution. Figure 1 plots the average number of non-diet soda drinks and other sugar sweetened beverages—juices with added sugar, energy, and sports drinks—consumed per day across the income distribution according to the 2009–2010 National Health and Nutrition Examination Survey.[30] Data is reported separately for soda and other SSBs. The horizontal axis reports the income quantile (in our income distribution from CPS) which corresponds to the central value of each income bin in the survey. To generate a smooth and continuous schedule of consumption across the income distribution, we regress soda and other SSB consumption on those income quantiles. The best fit lines are displayed in Figure 1, and we use the sum of these best fit lines as our calibration of total SSB consumption across incomes.[31] We assume 8 ounces of soda per drink, and we follow Long et al. (2015) in assuming an

[29] Specifically, we set $\epsilon = 0.33$; ϵ is the compensated elasticity of taxable income at points where the income tax is linear; it diverges if the income tax is very curved, see Appendix A, but in practice the difference is extremely small in our simulations.

[30] The question wording for the first category was: "During the past month, how often did you drink regular soda or pop that contains sugar? Do not include diet soda." The wording for the second category was: "During the past month, how often did drink sweetened fruit drinks, sports or energy drinks, such as Koolaid, lemonade, HiC, cranberry drink, Gatorade, Red Bull or Vitamin Water? Include fruit juices you made at home and added sugar to. Do not include diet drinks or artificially sweetened drinks."

[31] In our baseline specification, we exclude the data point corresponding to the lowest income bracket in the National Health and Nutrition Examination Survey. That point is an outlier, reporting a lower level of SSB consumption than

average price of \$0.06 per ounce of SSB.

We present results for a wide range of different price elasticities of demand for SSBs. A growing literature in public health and economics attempts to measure this elasticity, but estimates so far have been wide-ranging and inconclusive.[32] Due to a lack of credible exogenous variation, the estimates generally rely on cross-sectional variation. The estimates range from a low of 0.15 (Zheng and Kaiser, 2008) to a high of 2.25 (Dharmasena and Capps, 2011).[33] We present our results for five elasticity values spanning this range: 0.15, 0.3, 0.6, 1.2 , and 2.4.

We also present our results for a range of different levels bias, ranging from 0 to 8, measured in cents per ounce of SSB consumption.[34] To compute the level of bias, two statistics are needed: the marginal health costs of drinking soda, and the extent to which these health costs are under-appreciated. Long et al. (2015) estimate the impact of SSB consumption on quality-adjusted life years (QALY) by tracing its impact on the likelihood of suffering a stroke, ischemic heart disease, hypertensive heart disease, diabetes mellitus, osteoarthritis, postmenopausal breast cancer, colon cancer, endometrial cancer, and kidney cancer. Long et al. (2015) estimate that a permanent 20% reduction in SSB consumption generates an average increase of 0.00278 QALYs per person in the first ten years. A "back of the envelope" calculation suggests this equates to a total increase of 0.0021 QALYs per person from a one-year SSB reduction of 20%.[35] This translates to \$105 per person using the commonly used conservative estimate of \$50,000 for the value of a QALY (Hirth et al., 2000), or to \$210 per person under the less conservative estimate of \$100,000 (Hirth et al., 2000). These values correspond to marginal health costs of either \$0.10 or \$0.20 per ounce of SSB.[36]

To our knowledge, there is no empirical work directly measuring bias in SSB consumption, but to get a sense of how this may translate to $\tilde{\gamma}$, note for a present bias parameter β, we have bias equal to $\tilde{\gamma} = (1-\beta) * (\text{marginal health costs})$. Thus values of $\beta = 0.7$ or $\beta = 0.8$ lead to $\tilde{\gamma} = 3$ cents/oz or $\tilde{\gamma} = 2$ cents/oz for conservative valuations of a QALY, or to values of $\tilde{\gamma} = 6$ cents/oz or $\tilde{\gamma} = 4$ cents/oz for less conservative valuations of a QALY.

at slightly higher incomes, and we are concerned that it reflects misreporting by survey respondents who do not wish to reveal their incomes. We report results which include the lowest income point in Appendix F.1, and which show that the magnitudes of optimal taxes are not sensitive to this assumption.

[32]While significant variation in excise taxes enables precise estimates of elasticities for cigarettes and alcohol, such variation does not yet exist for sugary drinks. Excise taxes for SSBs currently exist in only a few cities, and many have not yet come into effect.

[33]See Andreyeva et al. (2010) for a review of the literature.

[34]Specifically, since the structural parameter $\tilde{\gamma}$ is slightly different from the money-metric bias $\gamma(\theta)$ in the presence of income effects, we select the structural $\tilde{\gamma}$ which gives rise to the reported money-metric bias.

[35]The simulated experiment in Long et al. (2015) is a 20% permanent reduction of SSB consumption, whereas we are interested in the QALY cost of each SSB consumed. To reach this, we first compute the approximate total QALYs gained from reducing SSB consumption for a single year. If the 10-year total of 0.00278 QALYs were composed of 10 years of exposure to the effects of a 20% reduction during year 1, 9 years of exposure from the reduction during year 2, etc., then the first year's reduction would account for $10/55 = 0.18$ of the 10 year total, or 0.00051 QALYs. This accounts only for the QALY difference in the first 10 years. If the per-year effect were the same in the ensuing years (a conservative assumption, given that health costs typically compound later in life), then this ten-year effect should be multiplied by four, given a median US age of 38 and life expectancy of 78. This yields a total approximate effect of 0.0021 QALYs for a one-year 20% SSB reduction.

[36]Long et al.'s (2015) estimates imply that on average, people consume 5475 oz of SSBs per year which implies that a 20% reduction would reduce consumption by 1095 oz.

6.3 Results

Figure 2 plots the optimal SSB tax as a function of the degree of bias $\tilde{\gamma}$ for several specifications. We first plot the optimal tax for a corrective tax with no redistributive motives (the "Pigouvian benchmark"), which lies on the 45-degree, exactly offsetting the negative internality of SSBs—this case is given by the dotted black line. The blue dashed line plots the optimal SSB tax when $\Phi = 1$; that is, when all SSB consumption variation is due to income effects. By the logic in Section 4.2, only the corrective force of the SSB tax is relevant in this case, since any regressivity costs can be offset by a more progressive income tax. Thus the optimal tax is larger than the Pigouvian benchmark in this case, illustrating how redistributive concerns magnify the optimal tax in the case of pure income effects.

We also plot the optimal tax across five elasticities of demand in the case where all variation is due to preference heterogeneity, $\Phi = 0$.[37] For low elasticity values, the optimal tax is substantially lower than in the case of pure income effects, since the change in behavior, and thus the corrective benefits, will be small. For a demand elasticity of 0.15, for example, regressivity costs reduce the optimal SSB tax substantially—to less than half of the optimal Pigouvian corrective tax. Under higher elasticity values, however, corrective benefits become relatively large compared to the regressivity costs. At an elasticity of 2.4, for example, regressivity costs are sufficiently small that the optimal SSB tax is higher than the Pigouvian benchmark, and is almost as high as the tax under pure income effects.[38] This highlights the importance of assessing both the source of SSB variation (income effects vs. preference heterogeneity) and the elasticity of demand when determining the optimal tax.[39]

Figure 2 also highlights the theoretical insight that the *sensitivity* of the tax to the bias is increasing in the elasticity when there are regressivity costs. As the elasticity decreases, the curve plotting the optimal tax against the bias not only shifts downward, but also becomes flatter.

Figure 3 explores the role of income effects vs. preference heterogeneity in greater detail, by plotting the optimal tax across Φ. The dotted line represents the Pigouvian benchmark without redistributive motives, while the solid lines plot the optimal SSB tax in the presence of redistributive concerns for a range of demand elasticities. The solid lines all converge at $\Phi = 1$ (pure income effects) to a value above the Pigouvian benchmark, but fan out due to the presence of regressivity costs when $\Phi < 1$. If SSBs are a normal good ($\Phi < 0$) then the regressivity costs are even larger than in the case of pure preference heterogeneity, and the optimal tax is reduced further still. The importance of the regressivity costs is modest at high demand elasticities, but becomes very large at low values.

Figure 3 also illustrates a helpful insight for computing a simple, yet accurate, approximation of the optimal SSB tax. As is evident from the figure, the optimal SSB tax varies nearly linearly with Φ. As a result, one can obtain an accurate approximation of the optimal tax at values other than $\Phi = 0$

[37] Note that the demand elasticity is not relevant for the optimal tax in the previous case of pure income effects.

[38] It is also helpful to note that for elasticities greater than one, a higher tax actually raises numeraire consumption, since consumers reduce their soda spending by more than enough to offset the higher tax. This implication echoes the point in (Gruber and Kőszegi, 2004) that a tax increase on harmful addictive goods can have negative incidence on highly elastic consumers.

[39] Appendix F.2 reproduces Figure 2 under stronger and weaker inequality aversion, which magnify and dampen the regressivity costs of the tax, respectively.

or $\Phi = 1$ by computing the "corrective benefit" and "regressivity costs" terms, and adding them, with weights equal to 1 and $1 - \Phi$, respectively. This approximation greatly simplifies the computation of the optimal tax for two reasons. First, simulations with pure income effects and pure preference heterogeneity are much simpler than those for intermediate cases, as is apparent from the analytic derivations of elasticities in Appendix E. Second, the simplifying technique discussed in Corollary 1 of using the weights g rather than $\hat{g}$ (which include income effects) is exactly correct for both $\Phi = 0$ (since income effects are nonexistant) and $\Phi = 1$ (since regressivity costs are zero). We confirm this insight numerically by reproducing Figure 3 in Appendix F.3 employing this simplification. The approximated optimal tax never differs from the true optimum by more than 0.01 cents per ounce.

Figure 4 considers the case in which all SSB variation is due to income effects, but the SSB tax is not salient at the time labor supply is chosen. Section 5.1 derived a striking theoretical result in this context: the expression for the optimal commodity tax in this case is remarkably similar to the case of pure preference heterogeneity (i.e., with full weight given to regressivity costs). We explore this insight numerically in Figure 4, which plots the optimal commodity tax across various degrees of SSB tax salience when labor supply choices are made.[40] The horizontal axis ranges from 0 (complete non-salience) to 1 (fully salient SSB tax). Figures 4 shows that the (non) salience of SSB taxes is a force very similar to role of preference heterogeneity. When SSB taxes are non-salient, the optimal tax is considerably reduced, particularly at low elasticities of demand.[41]

Figure 5 further explores the implications of income effects vs. preference heterogeneity by displaying the effect of the optimal SSB tax on the overall progressivity of the tax system. For this exercise we restrict to $\tilde{\gamma} = 4$ and elasticity $\epsilon = 0.6$. Results for alternative elasticities are displayed in Appendix F.4. We compute the optimal SSB tax, as well as the difference between the optimal income tax and the income tax that would be optimal if the commodity tax was constrained to $t = 0$. Panel (a) displays the difference in taxes paid (income taxes, SSB taxes, and combined taxes) under the full optimum relative to the constrained optimum, when SSB consumption is driven by income effects. The substantial positive SSB tax results in more SSB taxes paid at every level of income, particularly at the bottom of the income distribution where SSB consumption is concentrated. At the same time, the income tax becomes more progressive, largely offsetting the regressivity of the SSB tax, so that the total difference in taxes paid is mostly flat.[42]

In contrast, panel (b) shows that when SSB consumption differences are driven by preference heterogeneity, the income tax cannot offset the optimal SSB tax's regressivity. The revenue is recycled lump-sum and the total tax regime becomes more regressive. This shows that the smaller SSB tax which obtains under pure preference heterogeneity (relative to pure income effects) in Figures 2 and 3 should not be interpreted to imply a less regressive tax system—on the contrary, the tax is smaller precisely because it generates regressivity costs that cannot be efficiently offset.

[40] In all cases we assume the SSB tax is fully salient when choosing SSB consumption.

[41] This result also emphasizes the nuanced role of salience in determining optimal commodity taxes more generally—unlike the context of Chetty et al. (2007), where non-salience tends to dampen efficiency costs and thus magnify the optimal commodity tax, here nonsalience tends to reduce the optimal tax due to regressivity costs.

[42] The slight change in total taxes paid is due to the change in the income density when soda taxes change, as the labor leisure choice is slightly affected by the soda tax.

Finally, we explore the theoretical insight that the corrective component of the optimal tax varies with σ, which quantifies the covariation of marginal social welfare weight with the level of SSB consumption, the elasticity of demand, and bias. Figure 6 plots the optimal SSB tax as a function of *average* bias $\bar{\gamma}$, when bias declines with income. Specifically, bias is assumed to decline linearly across income percentiles, either from $2\bar{\gamma}$ to 0 or from $1.5\bar{\gamma}$ to $0.5\bar{\gamma}$. For simplicity, we restrict consideration to the midrange elasticity value of 0.6. The quantitative effect of this correlated heterogeneity is striking. Even the more modest decline from $1.5\bar{\gamma}$ to $0.5\bar{\gamma}$ is sufficient to reverse the directional implication of redistributive concerns in the pure preference heterogeneity case: with declining bias, the optimal SSB tax becomes slightly larger than the Pigouvian benchmark at high $\bar{\gamma}$. This result emphasizes the importance of empirically assessing not just the magnitude of consumption biases, but also how they covary with income.

6.4 Impact on social welfare

We conclude our simulations by quantifying the welfare improvements from implementing the optimal SSB tax, relative to the counterfactual in which the income tax is set optimally but the SSB tax is constrained to equal zero. We compute these benefits in terms of a policymaker's hypothetical "willingness to pay" for the optimal SSB tax policy. These tables report the per-capita gains multiplied by 240 million, the approximate number of US adults.[43]

Table 1 reports the welfare gains across the range of SSB demand elasticities explored above. Each row reports the willingness to pay for the optimal tax under 1) pure income effects, 2) pure preference heterogeneity, and 3) pure preference heterogeneity without regressivity costs. The last column is computed by calculating the willingness to pay for a "nudge" policy that achieves the same change in SSB demand as the optimal tax under pure preference heterogeneity, but without raising any revenue from consumers.

The third column, which reports the welfare gains from the pure preference heterogeneity policy without regressivity costs, contains values that lie between the values in columns 1 and 2. The values are smaller than those in the first column because the optimal tax under pure preference heterogeneity is smaller than the optimal tax under pure income effects, and thus achieves lower corrective benefits. The difference between columns 2 and 3 quantifies the welfare costs of regressivity from the optimal tax when that regressivity cannot be offset by income tax reforms.

Table 2 reports results similarly to Table 1, except that it measures the welfare benefit from implementing a 1 cent per oz SSB tax, rather than the optimal tax.

Several key quantitative lessons emerge from these tables. First, the gains are substantial, considering that SSB consumption constitutes a small share of total consumption. Indeed, at high values of bias and high elasticities, the gains are notably larger than total spending on SSBs. Relatedly, Tables 1 and 2 also demonstrate that the welfare benefits are much larger under higher demand elasticities—roughly an order of magnitude larger for an elasticity of 2.4 than for an elasticity of 0.15—because higher elasticities imply a larger response to the tax, and thus larger welfare gains

[43] In this respect they constitute a conservative estimate, as they ignore any benefits from corrective policy for SSB consumption among children.

from the corrective effect of the tax. Second, although the welfare benefit from existing one-cent-per-ounce SSB taxes is large, these tables show that for many plausible parameter values, welfare would rise substantially from increasing such taxes further.

7 Conclusion

Recent research and policy reforms suggest a role for corrective taxes to correct uninternalized harms (or benefits) from consumption of goods like sugar-sweetened beverages, cigarettes, or energy inefficient appliances. This paper addresses the common concern that such taxes may be highly regressive. We derive conditions for the optimal tax policy in a general setting with a nonlinear income tax and redistributive motives, and we show that distributional concerns affect the optimal tax in a nuanced way.

Our theoretical and numerical results clarify that the optimal policy does not just depend on the magnitude of consumer bias, but that it also depends critically on a number of other statistics: the extent to which variation in consumption of the sin good is explained by preference heterogeneity versus income effects, the extent to which the rich or the poor are more biased, the elasticity of demand and how that varies across the income distribution, and the salience of commodity taxes on the labor supply margin. Our numerical results show that the optimal SSB tax varies dramatically with all of these statistics. Yet there is little consensus about the SSB elasticity, and even less is known about the other key empirical parameters. For other goods such as cigarettes, alcohol, or energy-using durables, significant uncertainty about some or all of these key parameters remains as well. Our nuanced, but quantitatively significant theoretical insights thus underscore an enormous need for more empirical work to help guide public policy.

References

Allcott, Hunt, "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market," *American Economic Journal: Economic Policy*, 2013, *5* (3), 30–66.

__ **and Dmitry Taubinsky**, "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market," *American Economic Review*, 2015, *105* (8), 2501–2538.

__ **, Christopher Knittel, and Dmitry Taubinsky**, "Tagging and Targeting of Energy Efficiency Subsidies," *American Economic Review Papers and Proceedings*, 2015, *May.*

__ **, Sendhil Mullainathan, and Dmitry Taubinsky**, "Energy policy with externalities and internalities," *Journal of Public Economics*, 2014.

Andreyeva, Tatiana, Michael W. Long, , and Kelly D. Brownell, "The Impact of Food Prices on Consumption: A Systematic Review of Research on the Price Elasticity of Demand for Food," *American Journal of Public Health*, 2010, *100* (2), 216–222.

Atkinson, Anthony B and J.E. Stiglitz, "Design of Tax Structure - Direct Versus Indirect Taxation," *Journal of Public Economics*, 1976, *6*, 55–75.

Attari, Shahzeen, Michael DeKay, Cliff Davidson, and Wandi Bruine de Bruin, "Public Perceptions of Energy Consumption and Savings," *Proceedings of the National Academy of Sciences*, 2010, *37*, 16054–16059.

Baicker, Katherine, Sendhil Mullainathan, and Josh Schwartzstein, "Behavioral Hazard in Health Insurance," *working paper*, 2015.

Bernheim, B. Douglas, "The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics," *Journal of Cost Benefit Analysis*, 2016, pp. 1–57.

Bernheim, B Douglas and Antonio Rangel, "Addiction and cue-triggered decision processes," *American Economic Review*, 2004, pp. 1558–1590.

__ and __ , "Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics," *The Quarterly Journal of Economics*, 2009, *124* (1), 51–104.

Blumgart, Jake, "A Soda Tax is Fundamentally Regressive. What if It's Philadelphia's Best Hope for Progressive Change?," 2016.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen, "Calorie Posting in Chain Restaurants," *American Economic Journal: Economic Policy*, 2011, *3* (1), 91–128.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and Consumer Choice," *Journal of Political Economy*, 2013, *121*, 803–843.

Chetty, Raj, "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 2012, *80* (3), 969–1018.

__ , "Behavioral Economics and Public Policy: A Pragmatic Perspective," *American Economic Review Papers and Proceedings*, 2015, *105* (5), 1–33.

__ , Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," September 2007.

__ , __ , and __ , "Salience and Taxation: Theory and Evidence," *American Economic Review*, 2009, *99* (4), 1145–1177.

Davis, Lucas and Christopher Knittel, "Are Fuel Economy Standards Regressive?," *working paper*, 2016.

__ and Severin Borenstein, "The Distributional Effects of U.S. Clean Energy Tax Credits," *NBER Tax Policy and the Economy*, 2016, *30* (1), 191–234.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 2009, *47* (2), 315–372.

Dharmasena, Senarath and Oral Capps, "Intended and unintended consequences of a proposed national tax on sugar-sweetened beverages to combat the U.S. obesity problem.," *Health Economics*, 2011, *21*, 669–694.

Edwards, Jeremy, Michael Keen, and Matti Tuomala, "Income tax, commodity taxes and public good provision: A brief guide," *FinanzArchiv*, 1994, *51* (4), 83–95.

Farhi, Emmanuel and Xavier Gabaix, "Optimal Taxation with Behavioral Agents," 2015.

Gabaix, Xavier, "A Sparsity Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2014, *129*, 1661–1710.

__ and David Laibson, "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *Quarterly Journal of Economics*, 2006, *121* (2).

Gerritsen, Aart, "Optimal taxation when people do not maximize well-being," *Journal of Public Economics*, forthcoming.

Goldin, Jacob and Tatiana Homonoff, "Smoke gets in your eyes: cigarette tax salience and regressivity," *American Economic Journal: Economic Policy*, 2013, *5* (1), 302–336.

Goulder, Lawrence H., "Environmental taxation and the double dividend: A reader's guide," *International Tax and Public Finance*, 1995, *2* (2), 157–183.

__ and Roberton C. Williams, "The Substantial Bias from Ignoring General Equilibrium Effects in Estimating Excess Burden, and a Practical Solution," *Journal of Political Economy*, 2003, *111* (4), 898–927.

Gruber, Jon and Emmanuel Saez, "The elasticity of taxable income: evidence and implications," *Journal of public Economics*, 2002, *84* (1), 1–32.

Gruber, Jonathan and Botond Kőszegi, "Is Addiction "Rational"? Theory and Evidence," *The Quarterly Journal of Economics*, 2001, *116* (4), 1261–1303.

__ **and** __ , "Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes," *Journal of Public Economics*, 2004, *88* (9), 1959–1987.

Handel, Benjamin R., Jonathan T. Kolstand, and Johannes Spinnewijn, "Information Frictions and Adverse Selection: Policy Interventions in Health Insurance Markets," *working paper*, 2016.

Heutel, Garth, "Optimal Policy Instruments for Externality-Producing Durable Goods under Present Bias," *Journal of Environmental Economics and Management*, 2015, *72*, 54–70.

Hirth, Richard A., Michael E. Chernew, Edward Miller, Mark Fendrick, and William G. Weissert, "Willingness to Pay for a Quality-adjusted Life Year: In Search of a Standard," *Medical Decision Making*, 2000, *20*, 332–342.

Huehnergarth, Nancy Fink, "Senator Sanders Is Wrong: A Soda Tax Is Progressive, Not Regressive," 2016.

Jacobs, Bas and Robin Boadway, "Optimal Linear Commodity Taxation under Optimal Non-Linear Income Taxation," *Journal of Public Economics*, 2014, *117* (1), 201–210.

__ **and Ruud A. de Mooij**, "Pigou Meets Mirrlees: On the Irrelevance of Tax Distortions for the Second-Best Pigouvian Tax," *Journal of Environmental Economics and Management*, 2015, *71*, 90–108.

Jacquet, Laurence and Etienne Lehmann, "Optimal Nonlinear Income Taxation with Multidimensional Types: The Case with Heterogeneous Behavioral Responses," 2014.

Kaplow, Louis, "Optimal Control of Externalities in the Presence of Income Taxation," *International Economic Review*, 2012, *53* (2), 487–509.

Lavin, Timothy, "President Obama's Unserious Tax Plans," 2013.

Lockwood, Benjamin B, "Optimal Income Taxation with Present Bias," 2016.

Long, Michael W., Steven L. Gortmaker, Zachary J. Ward, Stephen C. Resch, Marj L. Moodie, Gary Sacks, Boyd A. Swinburn, Rob C. Carter, and Y. Claire Wang, "Cost Effectiveness of a Sugar-Sweetened Beverage Excise Tax in the U.S.," *American Journal of Preventive Medicine*, 2015, *49* (1), 112–123.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J Congdon, "A reduced-form approach to behavioral public finance," *Annual Review of Economics*, 2012, *4*.

Nava, Mario, Fred Shroyen, and Maurice Marchand, "Optimal fiscal and public expenditure policy in a two-class economy," *Journal of Public Economics*, 1996, *61* (1), 119–137.

O'Donoghue, Ted and Matthew Rabin, "Optimal sin taxes," *Journal of Public Economics*, 2006, *90* (10), 1825–1849.

Saez, Emmanuel, "Using elasticities to derive optimal income tax rates," *The Review of Economic Studies*, 2001, *68* (1), 205–229.

__ , "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics*, 2002, *83* (2), 217–230.

__ , "Optimal income transfer programs: intensive versus extensive labor supply responses," *The Quarterly Journal of Economics*, 2002, *117* (3), 1039–1073.

__ **, Joel Slemrod, and Seth H Giertz**, "The elasticity of taxable income with respect to marginal tax rates: A critical review," *Journal of Economic Literature*, 2012, *50* (1), 3–50.

Sanders, Bernie, "A Soda Tax Would Hurt Philly's Low-Income Families," 2016.

Schofield, Heather, "The economic costs of low caloric intake: Evidence from India," *working paper*, 2015.

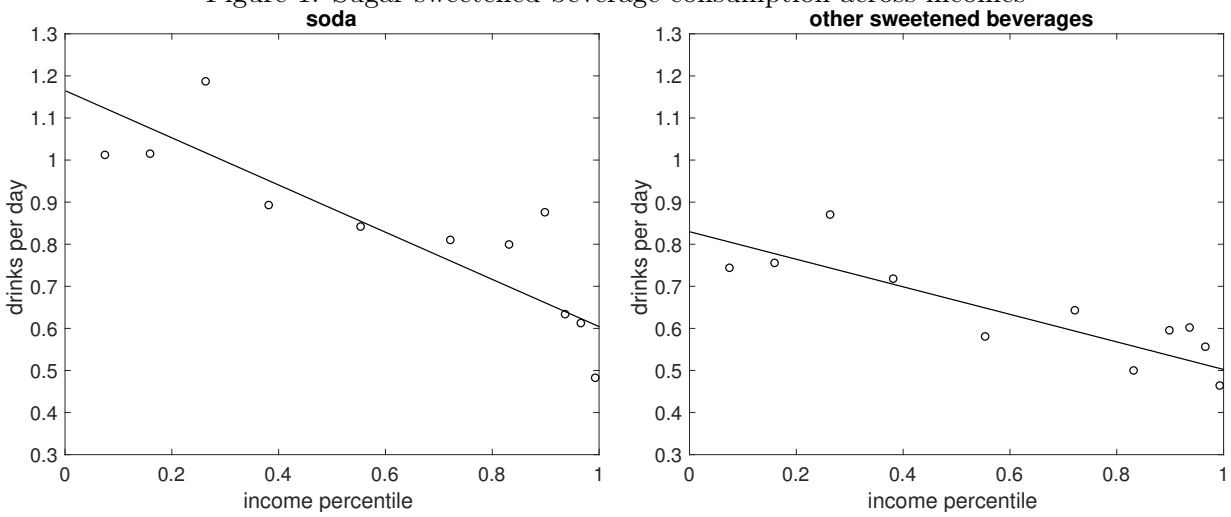
Spinnewijn, Johannes, "Heterogeneity, Demand for Insurance and Adverse Selection," *American Economic Journal: Economic Policy*, forthcoming.

Tsvetanov, Tsvetan and Kathleen segerson, "Re-evaluating the role of energy efficiency standards: A behavioral economics approach," *Journal of Environmental Economics and Management*, 2013.

Zheng, Yugin and Harry M. Kaiser, "Advertising and U.S. Nonalcoholic Beverage Demand," *Agricultural and Resource Economics Review*, 2008, *37* (2), 147–159.

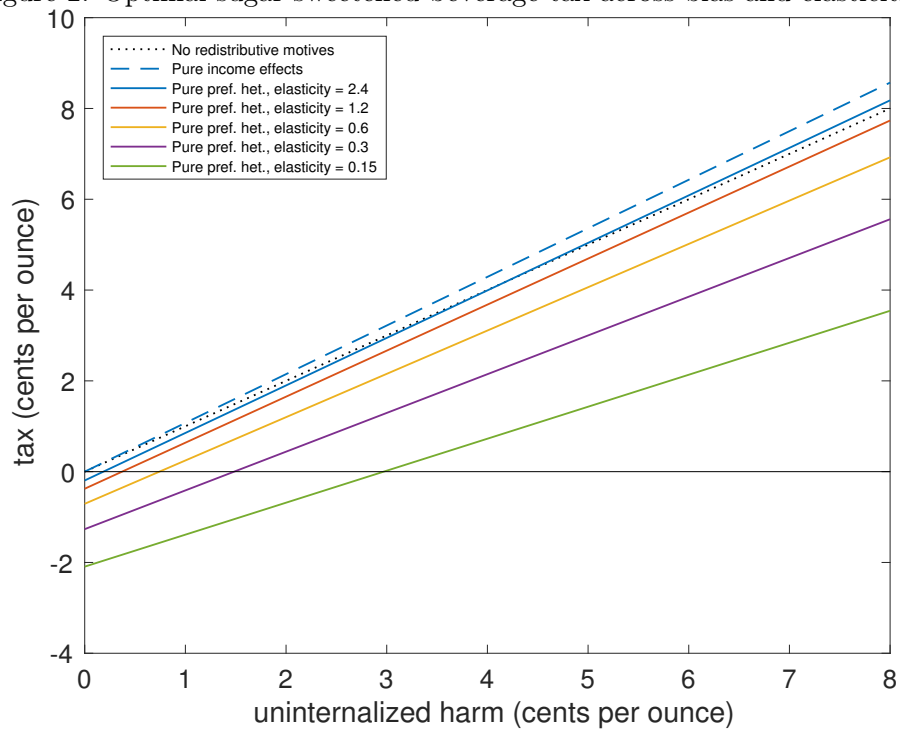
Figures

Figure 1: Sugar-sweetened beverage consumption across incomes



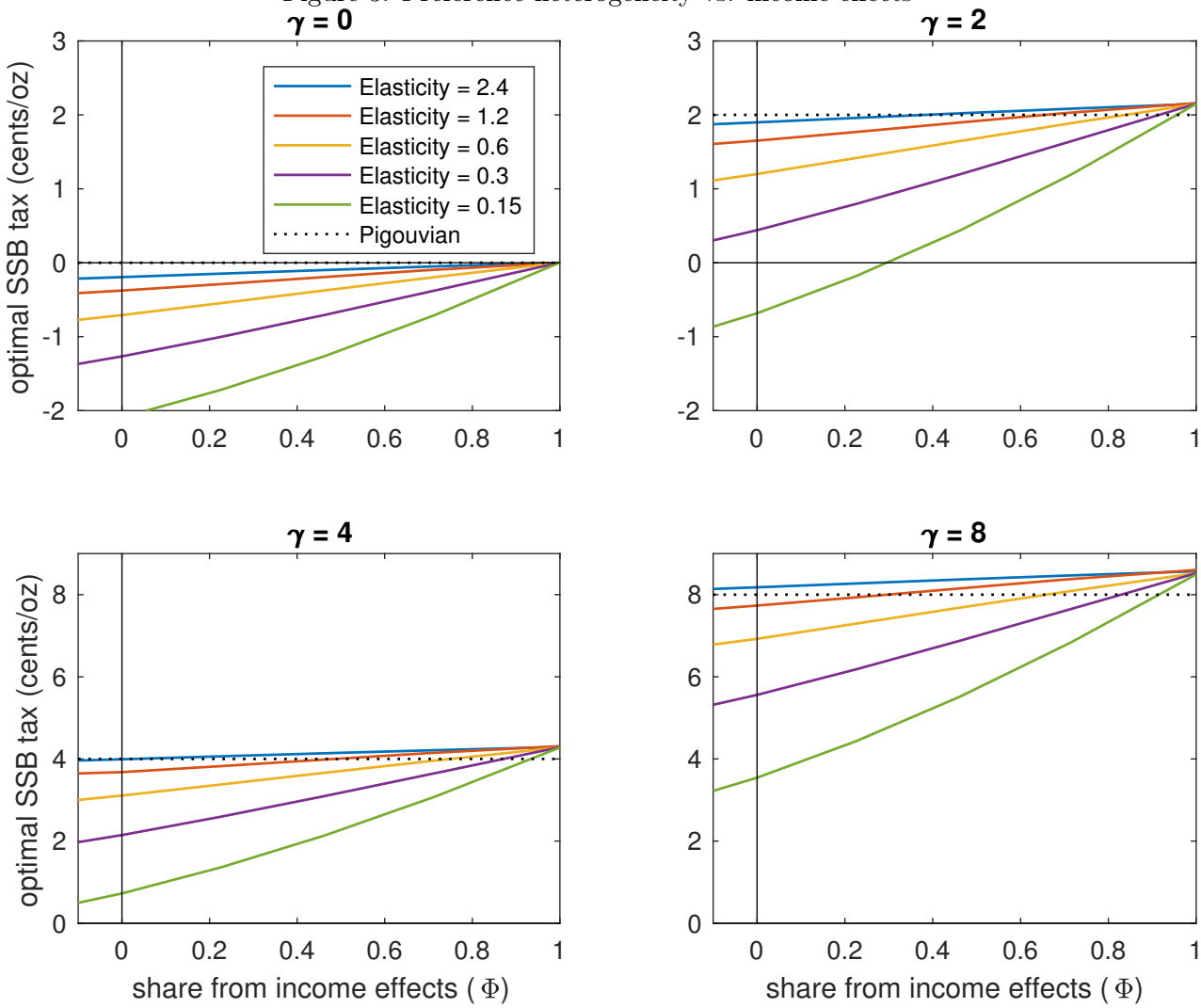
This figure plots the number of soda drinks and other sugar-sweetened beverages (SSBs) consumed across the income distribution, according to survey data drawn from the National Health and Nutrition Examination Survey in 2009–2010. We also plot the linear best fit lines, which we use to generate a continuous schedule of SSB consumption across incomes in our simulations. (See text for details.)

Figure 2: Optimal sugar-sweetened beverage tax across bias and elasticities



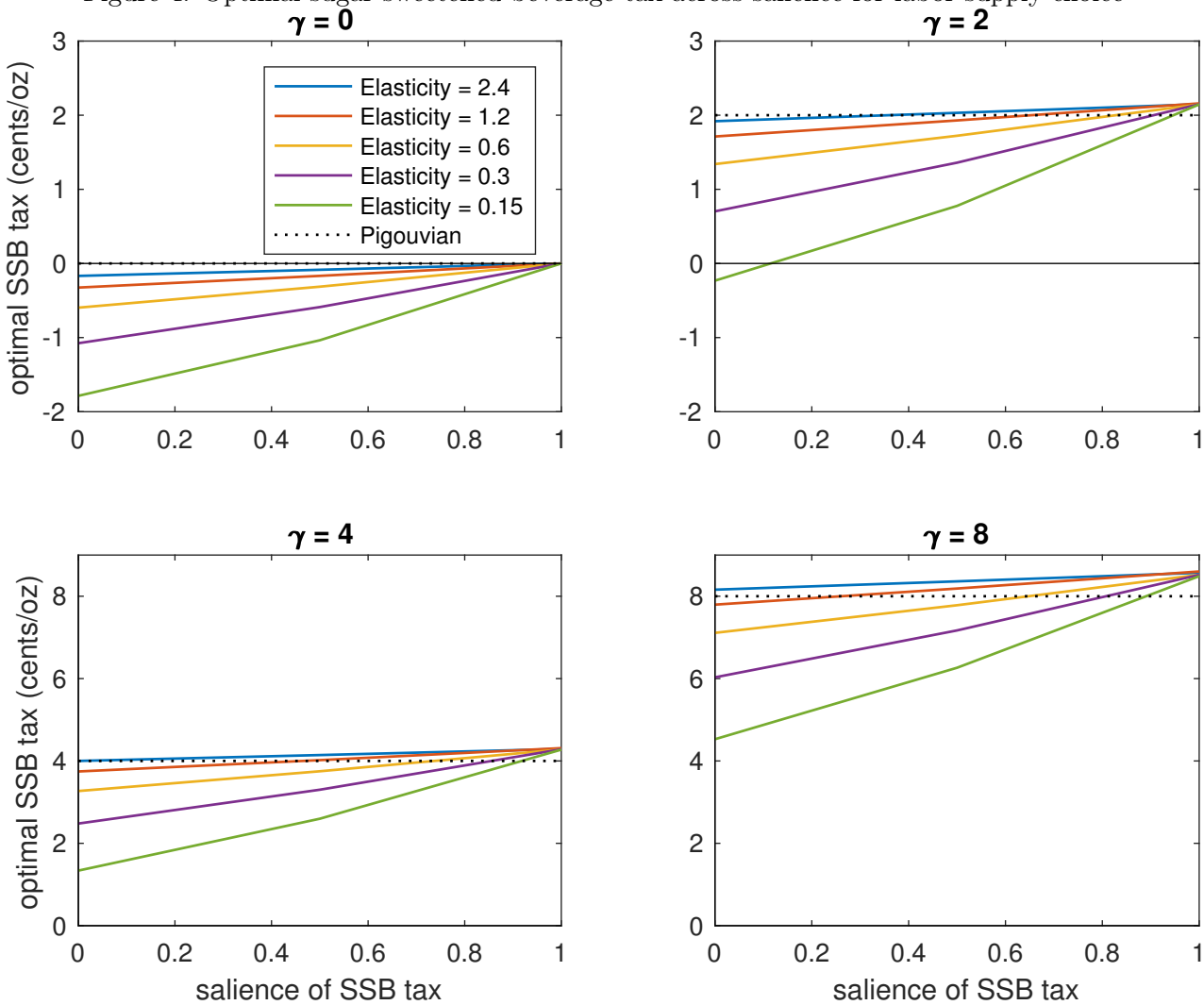
This figure plots the optimal sugar-sweetened beverage (SSB) tax as a function of the money-metric uninternalized harm from $1 of SSB consumption. The 45 degree line (dotted black) represents the optimal tax in the absence of redistributive concerns. The dashed blue line plots the optimal tax when SSB consumption is driven by income effects, while the solid lines plot the optimal tax when SSB consumption is driven by preference heterogeneity, across four values of the elasticity of demand for SSBs.

Figure 3: Preference heterogeneity vs. income effects



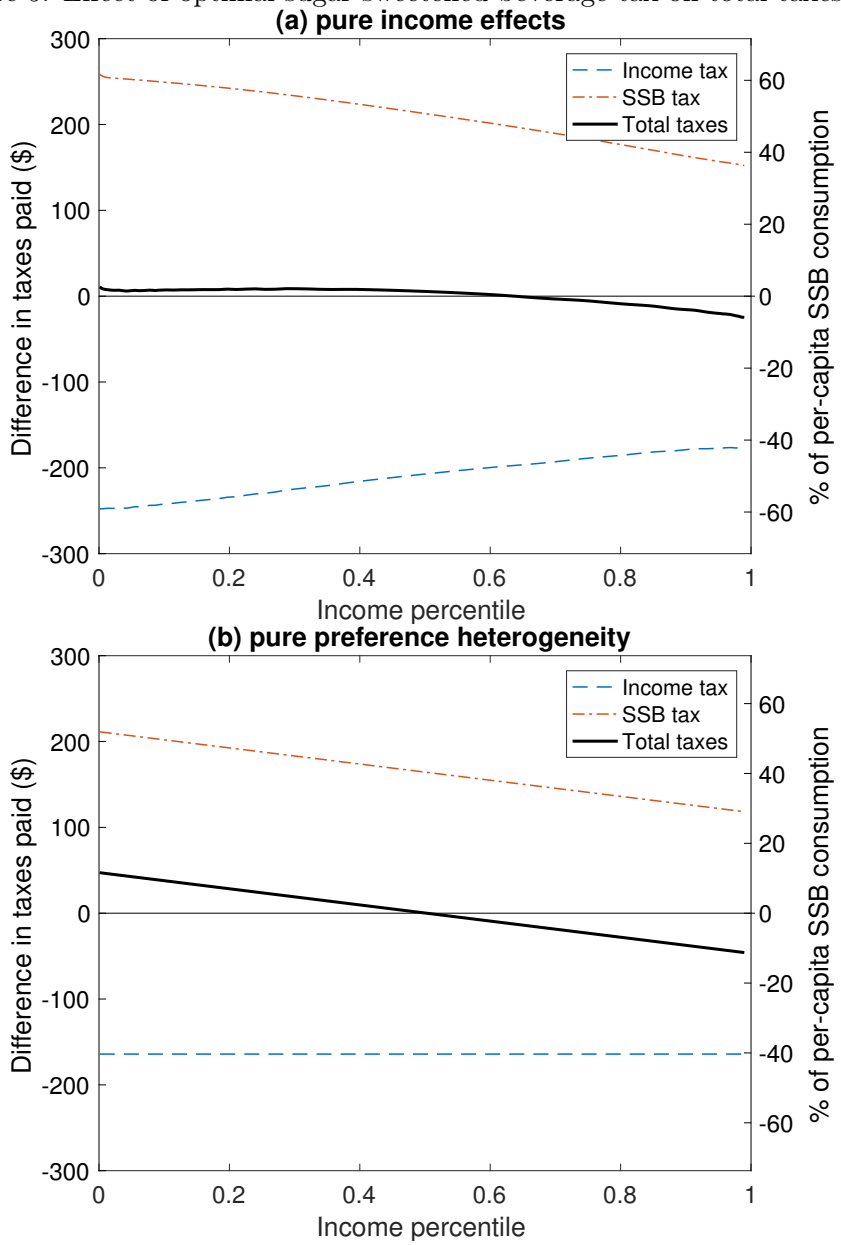
This figure plots the optimal sugar-sweetened beverage (SSB) tax against the strength of income effects in determining consumption of sugar sweetened beverages. When $\Phi = 1$, all differences in SSB consumption are due to income effects (the "Atkinson Stiglitz benchmark"), implying soda is a strongly inferior good. When $\Phi = 0$, all differences are due to preference heterogeneity. And when $\Phi < 0$, SSBs are a normal good. Each panel represents a different degree magnitude of the uninternalized harm from \$1 of SSB consumption.

Figure 4: Optimal sugar-sweetened beverage tax across salience for labor supply choice



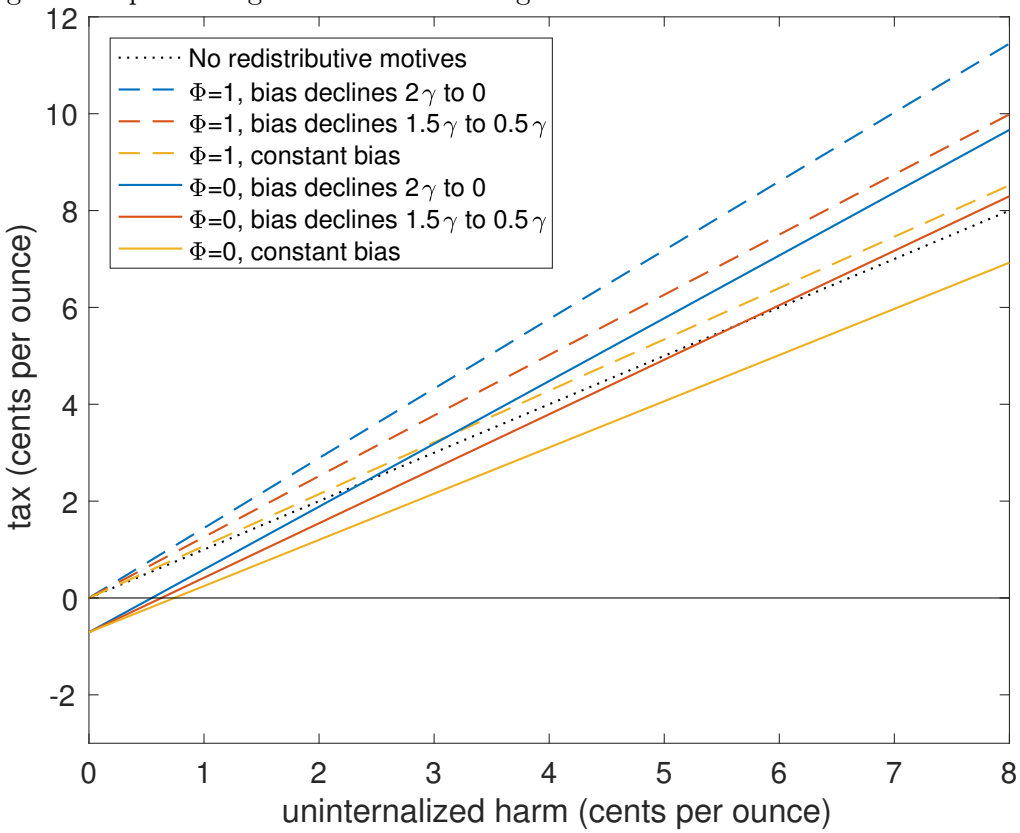
This figure plots the optimal sugar-sweetened beverage (SSB) tax when variation in SSB consumption is driven by income effects, but the SSB tax is not salient when consumers select their labor supply. The horizontal axis spans from 0 (completely non-salient tax) to 1 (fully salient tax).

Figure 5: Effect of optimal sugar-sweetened beverage tax on total taxes paid



These figures compare the total tax burden as a function of income under the optimal soda tax (assuming $\tilde{\gamma} = 1$ and a sugar-sweetened beverage (SSB) demand elasticity of 0.6), relative to the case in which the soda tax is constrained to equal zero. Net income taxes and soda taxes are displayed for the specification in which differences in soda consumption are driven by income effects (panel a) and preference heterogeneity (panel b).

Figure 6: Optimal sugar-sweetened beverage tax when bias declines across incomes



This figure plots the optimal tax across values of the uninternalized harm ("bias") from consuming sugar-sweetened beverages (SSBs), assuming that bias declines across incomes. Average bias $\bar{\gamma}$ is displayed on the horizontal axis, and simulations plot the optimal tax when bias declines from $2\bar{\gamma}$ (at the bottom of the income distribution) to 0 (at the top) linearly interpolated, and also for a decline from $1.5\bar{\gamma}$ to $0.5\bar{\gamma}$. This figure assumes a demand elasticity for SSBs of 0.6.

Tables

Table 1: Welfare gains from optimal sugar-sweetened beverage tax

SSB elasticity	Pure income effects	Pure pref. het.	Pref. het. without regressivity costs
$\tilde{\gamma} = 2$			
0.15	\$0.78 Bn	\$0.17 Bn	—
0.3	\$1.6 Bn	\$0.09 Bn	\$0.66 Bn
0.6	\$3.0 Bn	\$1.2 Bn	\$2.6 Bn
1.2	\$6.9 Bn	\$3.9 Bn	\$5.6 Bn
2.4	\$11.1 Bn	\$8.9 Bn	\$10.2 Bn
$\tilde{\gamma} = 4$			
0.15	\$2.9 Bn	\$0.14 Bn	\$1.1 Bn
0.3	\$5.7 Bn	\$2.0 Bn	\$4.6 Bn
0.6	\$10.6 Bn	\$7.0 Bn	\$10.2 Bn
1.2	\$22.7 Bn	\$16.3 Bn	\$19.1 Bn
2.4	\$34.1 Bn	\$30.3 Bn	\$31.9 Bn
$\tilde{\gamma} = 8$			
0.15	\$9.9 Bn	\$3.2 Bn	\$7.6 Bn
0.3	\$19.1 Bn	\$11.6 Bn	\$17.7 Bn
0.6	\$34.3 Bn	\$28.2 Bn	\$34.0 Bn
1.2	\$68.0 Bn	\$54.7 Bn	\$58.5 Bn
2.4	\$92.7 Bn	\$87.1 Bn	\$88.4 Bn

This table reports the welfare gains from the optimal sugar-sweetened beverage (SSB) tax in the US by quantifying a hypothetical policymaker's willingness to pay to implement the optimal policy. This is computed by calculating the welfare change from the economy where the SSB tax is constrained to equal zero (but the income tax is optimal) to the economy where both the SSB tax and the income tax are optimal—the columns 2 and 3 report the resulting change in welfare, normalized by the marginal value of public funds. Column 4 reports willingness to pay for a "nudge" which has the same effect on SSB consumption as the tax in Column 3, but which does not raise revenue and therefore has no regressivity costs. (Note: WTP is not reported for $\gamma = 2$, and an elasticity of 0.15, as the optimal tax is negative in that case, and thus there is not a welfare benefit from a pure nudge policy.)

Table 2: Welfare gains from a \$0.01 per ounce sugar-sweetened beverage tax

SSB elasticity	Pure income effects	Pure pref. het.	Pref. het. without regressivity costs
$\tilde{\gamma} = 2$			
0.15	\$0.60 Bn	–\$0.69 Bn	—
0.3	\$1.2 Bn	–\$0.06 Bn	—
0.6	\$2.3 Bn	\$1.2 Bn	\$2.4 Bn
1.2	\$5.2 Bn	\$3.4 Bn	\$4.5 Bn
2.4	\$8.9 Bn	\$7.5 Bn	\$8.4 Bn
$\tilde{\gamma} = 4$			
0.15	\$1.4 Bn	\$0.12 Bn	\$1.4 Bn
0.3	\$2.8 Bn	\$1.5 Bn	\$2.8 Bn
0.6	\$5.4 Bn	\$4.3 Bn	\$5.5 Bn
1.2	\$11.8 Bn	\$9.4 Bn	\$10.5 Bn
2.4	\$20.1 Bn	\$18.4 Bn	\$19.3 Bn
$\tilde{\gamma} = 8$			
0.15	\$3.0 Bn	\$1.7 Bn	\$3.0 Bn
0.3	\$5.9 Bn	\$4.7 Bn	\$6.0 Bn
0.6	\$11.4 Bn	\$10.5 Bn	\$11.7 Bn
1.2	\$24.9 Bn	\$21.3 Bn	\$22.4 Bn
2.4	\$42.4 Bn	\$40.2 Bn	\$41.1 Bn

This table reports the welfare gains computed as in Table 1, except it compares the zero tax equilibrium to a one cent per ounce tax on SSBs.

Appendices (not for publication)

A Income Elasticity Definitions

We define labor supply responses to include any "circularities" due to the curvature of the income tax function, which is assumed to be differentiable. Thus, following Jacquet and Lehmann (2014), we define a tax function $\hat{T}$ which has been locally perturbed around the income level z_0 by raising the marginal tax rate by τ and reducing the marginal tax level by ν:

$$\hat{T}(z; z_0, \tau, \nu) = T(z) + \tau(z - z_0) - \nu.$$

Let $z^*(\theta)$ denote a type θ's choice of earnings under the status quo income tax T, and let $\hat{z}(\theta; \tau, \nu)$ denote θ's choice of earnings under the perturbed income tax $\hat{T}(z; z^*(\theta), \tau, \nu)$. Then the compensated elasticity of taxable income is defined in terms of the response of $\hat{z}$ to τ, evaluated at $\tau = \nu = 0$:

$$\zeta_z^c(\theta) := \left(- \left. \frac{\partial \hat{z}(\theta; \tau, 0)}{\partial \tau} \right|_{\tau=0} \right) \frac{1 - T'(z^*(\theta))}{z^*(\theta)}$$

The income tax is similarly defined in terms of the response of $\hat{z}$ to a tax credit ν (this statistic will be nonpositive if leisure is a non-inferior good):

$$\eta_z(\theta) := \left(\left. \frac{\partial \hat{z}(\theta; 0, \nu)}{\partial \nu} \right|_{\nu=0} \right) \left(1 - T'(z^*(\theta))\right).$$

These definitions are comparable to those in Saez (2001), except that they include circularities and thus permit a representation of the optimal income tax in terms of the actual earnings density, rather than the "virtual density" employed in that paper.

B Proofs of Propositions in the Paper

B.1 Proof of Proposition 1, with additional derivations without assumption A

Income tax:

Following Saez (2001), we have the following effect form increasing the marginal tax rate between z^* and $z^* + dz$:

1. Direct effect (fiscal and welfare): $\int_{z=z^*}^{\infty} (1 - g(z)) dH(z)$
2. Compensated elasticity effect (fiscal and welfare): $-\zeta_z^c z \frac{T'}{1-T'} h(z^*) - (\zeta_z^c z/(1-T'))h(z^*)(g(z^*)\bar{\gamma}(z^*) - t)\frac{dc_2}{dz}$ or $-\zeta_z^c z^* \frac{T'}{1-T'} h(z^*) + \zeta_z^c z^* \frac{g(z^*)\bar{\gamma}(z^*) - t}{p+t} \eta(z^*) h(z^*)$
   (a) Note: for simplicity I am assuming here an independence between ζ_z^c, η, and γ.

3. Income effect (fiscal and welfare): $\int_{z\geq z^*} \eta_z \frac{T'}{1-T'} h(z) - \int_{z\geq z^*} \eta_z \frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z)h(z) + \int_{z\geq z^*} \frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z)h(z)$ (note that the second term is the income effect on c_2 from the change earnings z, while the second term is the income effect on c_2 from the change in after-tax income $y = z(1-T)$.)

These three terms must sum to zero, as the net welfare effect of this reform should be zero. Using our definition of $\hat{g}(z) = \hat{g}(z) = g(z) - (1+\eta_z)\frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z) + \eta_z \frac{T'}{1-T'}$, this yields

$$\frac{T'(z)}{1-T'(z)} = \frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z) + \frac{1}{\zeta_z^c z h(z)} \int_z^\infty (1-\hat{g}(x))dH(x)$$

Commodity tax:

First, consider the total effect on welfare of a marginal increase in the commodity tax dt. The total welfare effect, written in terms of the marginal value of public funds, can be decomposed into the following components:

- *Mechanical revenue effect*: the reform mechanically raises revenue from each consumer by $dtc_2(\theta)$, for a total of dtC_2.
- *Mechanical welfare effect*: the reform mechanically reduces each consumer's net income by $dtc_2(\theta)$. To isolate the mechanical effect, we compute the loss in welfare as if this reduction all comes from composite consumption c_1 for a welfare loss of $dtc_2(\theta)g(\theta)$, and we account for adjustments in c_2 and z in the behavioral effects below. Thus the total mechanical welfare effect is $-dt \int_\Theta c_2(\theta)g(\theta)d\mu(\theta)$.
- *Direct effect on c_2 consumption:* the reform causes each consumer to decrease their c_2 consumption by $dt\zeta(\theta)\frac{c_2(\theta)}{p+t}$. This generates a fiscal externality equal to $dt \int_\Theta t\zeta(\theta)\frac{c_2(\theta)}{p+t}d\mu(\theta)$. It also generates a behavioral welfare effect equal to $\alpha(\theta)\left(V_2(\theta) - V_1(\theta)(1+t)\right)\left(dt\zeta(\theta)\frac{c_2(\theta)}{p+t}\right)$, or $dt \int_\Theta g(\theta)\gamma(\theta)\zeta(\theta)\frac{c_2(\theta)}{p+t}d\mu(\theta)$.
- *Effect on earnings:* The reform causes a change in labor supply equal to $\frac{dz(\theta,t,T)}{dt}T'(z(\theta,t,T))$, creating a fiscal externality through the income tax equal to $\int_\Theta \frac{T'(z(\theta))}{1-T'}\left(\zeta_z^c(\theta)z(\theta)\frac{\partial c_2(z,\theta)}{\partial z} + \eta_z c_2(\theta)\right)d\mu(\theta)$. To compute $\frac{dz(\theta,t,T)}{dt}$, we use Lemma 1 from Saez (2002a), which carries through in this context: the change in earnings for type θ due to the increase in the commodity tax dt is equal to the change in earnings which would be induced by imposing a type-specific income tax reform, raising the income tax by $dT(z) = dtc_2(z|\theta)$, where $c_2(z|\theta)$ is a function of z denoting the amount of c_2 that θ would consume at each possible income level. Intuitively, an increase in the commodity tax alters the utility that type θ would gain from selecting each possible level of earnings, in proportion to the amount of c_2 that θ would consume at each income. This alternative income tax reform reduces the utility θ would realize at each possible income by instead altering the income tax. The resulting adjustment in earnings can be decomposed into a compensated elasticity effect and an income effect. The former depends on the change in the marginal income

tax rate, equal to $dt\frac{dc_2(z|\theta)}{dz} = dt\eta(\theta)$; the latter depends on the change in the tax level, equal to $dtc_2(\theta)$. Combining these effects, we have $\frac{dz(\theta,t,T)}{dt} = \zeta_z^c(\theta)\left(\frac{z(\theta)}{1-T'(z(\theta))}\right)\eta(\theta) + \eta_z(\theta)c_2(\theta)$.

- *Indirect effects on c_2 consumption:* The change in earnings affects consumption indirectly. This generates both additional fiscal externalities and additional welfare effects. These total to $\int z\zeta_z^c\frac{\partial c_2(z,\theta)}{\partial z}\frac{\eta(g\gamma-t)}{p+t}d\mu(\theta) + \int \eta_z c_2\frac{\eta(g\gamma-t)}{p+t}d\mu(\theta)$

Combining these components, the total welfare effect of the tax reform dt is equal to

$$\begin{aligned}
\frac{dW}{dt} &= \int_\Theta \left(c_2(\theta)(1-g(\theta)) + c_2(\theta)\left(\frac{\zeta^c(\theta)}{p+t} + \frac{\eta(\theta)}{p+t}\right)(t-g(\theta)\gamma(\theta))\right)d\mu(\theta) \\
&- \int_\Theta \frac{T'(z(\theta))}{1-T'}\left(\zeta_z^c(\theta)z(\theta)\frac{\partial c_2(z,\theta)}{\partial z} + \eta_z c_2(\theta)\right)d\mu(\theta) \\
&+ \int z\zeta_z^c\frac{\partial c_2(z,\theta)}{\partial z}\frac{\eta(g\gamma-t)}{p+t}d\mu(\theta) + \int \eta_z c_2\frac{\eta(g\gamma-t)}{p+t}d\mu(\theta) \\
&= \int_\Theta \left(c_2(\theta)(1-g(\theta)) + c_2(\theta)\left(\frac{\zeta^c(\theta)}{p+t} + \frac{\eta(\theta)}{p+t}\right)(t-g(\theta)\gamma(\theta))\right)d\mu(\theta) \\
&- \int_\theta \int_{s\geq z(\theta)}(1-\hat{g}(s))ds\left(\frac{\partial c_2(z,\theta)}{\partial z} + \frac{\eta_z c_2(\theta)}{\zeta_z^c(\theta)z(\theta)}\right)d\mu(\theta)
\end{aligned}$$

If the income tax is optimal, any small perturbation to it generates no first-order change in welfare. In particular, this statement must hold for a reform which raises the income tax at each z in proportion to average c_2 consumption at that z, i.e., for the perturbation $dT(z) = dt\bar{c}_2(z)$. Letting $\frac{dW}{dT}$ denote the total welfare effect of this marginal reform, we have

$$\begin{aligned}
\frac{dW}{dt} &= \int_\Theta \left(\bar{c}_2(z(\theta))(1-g(\theta)) - \bar{c}_2(z(\theta))\frac{\eta(\theta)}{1-T'(z(\theta))}(t-g(\theta)\gamma(\theta))\right)d\mu(\theta) \\
&- \int_\Theta \frac{T'(z(\theta))}{1-T'}\left(\zeta_z^c(\theta)z(\theta)\bar{c}_2'(z(\theta)) + \eta_z\bar{c}_2(\theta)\right)d\mu(\theta) \\
&+ \int z\zeta_z^c\bar{c}_2'(z(\theta))\frac{\eta(g\gamma-t)}{p+t}d\mu(\theta) + \int \eta_z\bar{c}_2\frac{\eta(g\gamma-t)}{p+t}d\mu(\theta) \\
&= \int_\Theta \left(c_2(\theta)(1-g(\theta)) + c_2(\theta)\left(\frac{\zeta^c(\theta)}{p+t} + \frac{\eta(\theta)}{p+t}\right)(t-g(\theta)\gamma(\theta))\right)d\mu(\theta) \\
&- \int_\theta \int_{s\geq z(\theta)}(1-\hat{g}(s))ds\left(\bar{c}_2'(z(\theta)) + \frac{\eta_z c_2(\theta)}{\zeta_z^c(\theta)z(\theta)}\right)d\mu(\theta)
\end{aligned}$$

Since the income tax is optimal by assumption, we can simplify $\frac{dW}{dt}$ by subtracting $\frac{dW}{dT} = 0$. The resulting expression can be simplified by writing the integral over incomes and over types conditional on income. Then employing the assumption that $g(\theta)$ is constant conditional on income, the mechanical effects cancel out, leaving

$$\frac{dW}{dt} = \int_0^\infty \left(\frac{\bar{c}_2(z)\bar{\zeta}^c(z)}{p+t}\left(t - g(z)\bar{\gamma}(z)\right) + \frac{1}{1-T'(z)}\left(g(z)\Sigma^{(z)}_{\eta\gamma,c_2} - t\Sigma^{(z)}_{\eta,c_2}\right)\right) dH(z)$$
$$+ \int_0^\infty \left(\int_{s\geq z}(1-\hat{g}(s))ds \left(\left(\psi(z) - \bar{c}_2'(z)\right) + \frac{\Sigma^{(z)}_{\zeta^c_z,\eta}}{z\bar{\zeta}^c_z(z)}\right) + \frac{T'(z)\Sigma^{(z)}_{\eta_z,c_2}}{z\bar{\zeta}^c_z(z)}\right) dH(z) \quad (16)$$

where $\Sigma^{(z^*)}_{x,y} = Cov_{\mu|z^*}\left[x(\theta), y(\theta)|\, z(\theta) = z^*\right]$ denotes the income-conditional covariance between two variables $x(\theta)$ and $y(\theta)$ at the optimum. Under the optimal commodity tax, (16) is equal to zero. This expression is similar to the derivation in Saez (2002a) of the effect of introducing a small commodity tax (expression (5) in that paper) with three important differences. First, (16) depends on the bias γ, which is not present in that paper. Second, since Saez (2002a) considers the *introduction* of a small commodity tax (rather than the perturbation of a possibly non-zero tax) there is no first-order fiscal externality from the change in c_2 consumption. Since we are interested in characterizing the optimal tax in situations where the optimal t^* is non-zero, (16) includes that fiscal externality—it therefore represents a necessary condition at the optimum, rather than a test of whether the optimal t is zero as in Saez (2002a). Third, the income responses ζ^c_z and η_z in this expression represent the full income response, accounting for nonlinearities in $T(z)$, which simplifies the expression.

Employing assumption A, the Σ terms are equal to zero. This is the extension of Assumption 2 in Saez (2002a). (More generally, the same formula obtains if the level of c_2 consumption and the income effect η are orthogonal to other behavioral responses conditional on income.) In this case, the necessary condition for optimality reduces to

$$\frac{dW}{dt} = \int_0^\infty \left(\frac{\bar{c}_2(z)\bar{\zeta}^c(z)}{p+t}\left(t - g(z)\bar{\gamma}(z)\right) + \int_{s\geq z}(1-\hat{g}(s))ds\left(\bar{c}_2'(z(\theta)) - \psi(z)\right)\right) dH(z). \quad (17)$$

The second term can be integrated by parts and rewritten as follows:

$$\int_0^\infty \int_z^\infty (1-\hat{g}(s))dH(s)\left(\bar{\eta}(z) - \bar{c}_2'(z)\right)dz = \int_0^\infty (1-\hat{g}(z))h(z)\left(\int_0^z \psi(s)ds - \Delta\bar{c}_2(z)\right)dz$$
$$= \int_0^\infty (1-\hat{g}(z))\left(\phi(z) - \bar{c}_2(z)\right)dH(z) \quad (18)$$

Substituting this into 17 and rearranging yields Proposition 1.

B.2 Proof of Corollary 1

We show that η/ζ is negligible under the assumptions of the corollary. The result will follows directly from this, as Appendix C.2 shows that when $\eta_z = 0$,

$$t = \bar{\gamma}(\bar{g} + \sigma) - \frac{p+t}{\bar{\zeta}^c} Cov\left[g(z) - \frac{g(z)\bar{\gamma}(z) - t}{p+t}\eta(z), \tilde{\phi}(z)\right].$$

And when η/ζ is negligible, this reduces to

$$t = \bar{\gamma}(\bar{g} + \sigma) - \frac{p+t}{\bar{\zeta}^c} Cov\left[g(z), \tilde{\phi}(z)\right].$$

To derive the uncompensated elasticity, differentiate the FOC

$$-(p+t)U_1(y-(p+t)c_2, c_2, \dots) + U_2(y-(p+t)c_2, c_2, \dots) = 0,$$

where y denotes after-tax income, to get

$$-U_1 + (p+t)c_2U_{11} - c_2U_{12} + D\frac{dc_2}{dt} = 0,$$

where D is the partial derivative of the FOC with respect to c_2, the sign of which is negative since we assume a strict second-order condition. Thus

$$\frac{dc_2}{dt} = \frac{-U_1 + (p+t)c_2U_{11} - c_2U_{12}}{-D}.$$

The income effect is similarly given by

$$\frac{dc_2}{dy} = \frac{-(p+t)U_{11} + U_{12}}{-D}$$

Thus

$$\begin{aligned}
\eta &= (p+t)\left(\frac{(p+t)U_{11} - U_{12}}{D}\right) \\
\zeta &= (p+t)\left(\frac{-U_1}{c_2D} + \frac{(p+t)U_{11} - U_{12}}{D}\right) \\
&= (p+t)\left(\frac{-U_1}{c_2D}\right) + \eta
\end{aligned}$$

and so

$$\frac{\eta}{\zeta^c} = c_2\frac{(p+t)U_{11} - U_{12}}{-U_1}.$$

B.3 Proof of Lemma 2

We can instead ask the following intuitive question: "When we raise the commodity tax by dt, how much money can we give back to each person to offset the tax so as to keep the average labor supply choices of each z-earner constant?" Call this quantity $\chi(z)$. Under Assumption A, this term satisfies the following differential equation:

$$\bar{\zeta}^c_z z\chi'(z) + \eta_z\chi(z) = \bar{\zeta}^c_z z\psi(z) + \eta_z\bar{c}_2(z) \tag{19}$$

where $\psi(z) = E\left[\frac{dc_2(z,\theta)}{dz}|z^*(\theta) = z\right]$. The right-hand side is the impact of the commodity-tax, as shown in Appendix B.1. The left-hand side follows straightforwardly. The first term is just compensated effect of increasing the marginal tax rate, while the second term is the income effect. Rearranging

yields

$$\chi'(z) + \frac{\eta_z}{\bar{\zeta}^c_z z}\chi(z) = \psi(z) + \frac{\eta_z}{\bar{\zeta}^c_z z}c_2(z)$$

The solution to this first-order differential equation is

$$\chi(z) = \frac{\int_{x=0}^{z} e^{\int^x \frac{\eta_z}{\bar{\zeta}^c_z x'}dx'}\left(\psi(x) + \frac{\eta_z}{\bar{\zeta}^c_z x}c_2(x)\right)dx + K}{e^{\int^z \frac{\eta_z}{\bar{\zeta}^c_z z}dx'}}.$$

where K is some integration constant.

Now set $\phi(z) = \int_{x=0}^{z}\psi(x)dx + c_2(0)$. Then

$$\frac{d}{dx}\phi(x)e^{\int^x \frac{\eta_z}{\bar{\zeta}^c_z x'}dx'} = e^{\int^x \frac{\eta_z}{\bar{\zeta}^c_z x'}dx'}\left(\psi(x) + \frac{\eta_z}{\bar{\zeta}^c_z x}\phi(x)\right)$$

and thus

$$\chi(z) \quad = \quad \phi(z) + \frac{K - \phi(0)}{e^{\int^z \frac{\eta_z}{\bar{\zeta}^c_z z}dx'}} + \int_0^z w(x,z)\frac{\eta_z}{\bar{\zeta}^c_z x}(c_2(x))dx$$

where $w(x,z) = e^{\int_{x'=x}^{x'=z} -\frac{\eta_z}{\bar{\zeta}^c_z z}dx'}$. And to get the initial conditions right $\chi(0) = \phi(0)$, we must have $K = \phi(0)$, so that

$$\chi(z) = \phi(z) + \underbrace{\int_0^z w(x,z)\frac{\eta_z}{\bar{\zeta}^c_z x}(c_2(x) - \phi(x))dx}_{\text{labor supply income effect adjustment}}. \tag{20}$$

The intuition for the "income effect adjustment" term is that the higher the income effect, the more the commodity tax increases labor supply through the income effect, and thus the higher ϕ can be.

B.4 Proof of Proposition 2

Consider now the welfare impacts of a reform that increase the commodity tax by dt and the income tax $\chi(z)$ at each point z. Because by Lemma has the following impact on welfare, under Assumption A:

$$dW \quad = \quad \int_z \left(\overbrace{(1-g(z))(c_2(z)-\chi(z))}^{\text{Mechanical revnue and welfare}} - \overbrace{\bar{c}_2(z)(t-\gamma)\frac{\bar{\zeta}^c(z)}{p+t}}^{\text{Corrective benefit I}} - \overbrace{\frac{\eta(z)}{p+t}(t - g(z)\bar{\gamma}(z))(c_2(z)-\chi(z))}^{\text{Corrective benefit II}} \right) dH(z) \tag{21}$$

Equation (21) follows from the following effects. The first effect is the direct revenue and welfare effect of decreasing each individual's income by $\bar{c}_2(z) - \chi(z)$. The second effect corresponds to the compensated demand response, which generates both a welfare effect and a fiscal externality from substitution. The third effect comes from the effect that a wealth decrease of $c_2(z) - \chi(z)$ has on c_2

consumption; again, this generates both a fiscal externality and a welfare effect.

$$\begin{aligned} t &= \bar{\gamma}(\bar{g}+\sigma) - \frac{1}{\bar{\zeta}^c}\int \tilde{\chi}(z)\eta(z)(t-g(z)\bar{\gamma}(z)) + \frac{p+t}{\bar{\zeta}^c}E[(1-g(z))\tilde{\chi}(z)]. \\ &= \bar{\gamma}(\bar{g}+\sigma) + \frac{1}{\bar{\zeta}^c}E\left[\tilde{\chi}(z)\eta(z)g(z)\bar{\gamma}(z)\right] - \frac{t}{\bar{\zeta}^c}E[\tilde{\chi}(z)\eta(z)] + \frac{p+t}{\bar{\zeta}^c}E[(1-g(z))\tilde{\chi}(z)] \end{aligned}$$

Thus

$$\begin{aligned} t &= \left(\bar{\zeta}^c - E\left[\tilde{\chi}(z)(1-g(z)-\eta(z))\right]\right) \\ &= \bar{\zeta}^c\bar{\gamma}(\bar{g}+\sigma) + pE[(1-g(z))\tilde{\chi}(z)] + E\left[\tilde{\chi}(z)\eta(z)g(z)\bar{\gamma}(z)\right] \end{aligned}$$

from which the result follows.

B.5 General Analysis of Tax Salience, Including a Proof of Proposition 3

B.5.1 Income tax

Agent chooses labor supply to maximize $U(z-(\hat{p}(1-\vartheta)+(p+t)\vartheta)\hat{c}_2(z)-T(z),\hat{c}_2(z),z,\theta)$, where $\hat{p}$ is some anchor for agents' perceptions of the final post-tax price. The first order condition for the consumer is $(1-T')U_1(\hat{c}_1,\hat{c}_2,z,\theta)+U_3(\cdot)=0$. The perceived consumption bundle satisfies $-(p+\vartheta t)U_1(\hat{c}_1,\hat{c}_2,...)+U_2(\hat{c}_1,\hat{c}_2,...)=0$, while the actual consumption bundle satisfies $-(p+t)U_1(c_1,c_2,...)+U_2(c_1,c_2,...)=0$. The impact of changing z for this consumer is $(1-T'-(p+t)c_2')V_1+c_2'V_2+V_3$ which equals

$$(1-T')V_1(c_1,c_2,...)+V_3(c_1,c_2,...)-V_1[(U_2/U_1)-V_2/V_1]c_2'.$$

This equals

$$\begin{aligned} &V_1[1-T'+U_3(c_1,c_2,...)/U_1(c_1,c_2,...)]-V_1[(U_2/U_1)-V_2/V_1]\frac{\partial c_2}{\partial z} \\ =&V_1[U_3(c_1,c_2,...)/U_1(c_1,c_2,...)-U_3(\hat{c}_1,\hat{c}_2,...)/U_1(\hat{c}_1,\hat{c}_2,...)]-V_1[(U_2/U_1)-V_2/V_1]\frac{\partial c_2}{\partial z} \end{aligned}$$

We define the labor supply bias induced by salience $\gamma_\vartheta := U_3(c_1,c_2,...)/U_1(c_1,c_2,...)-U_3(\hat{c}_1,\hat{c}_2,...)/U_1(\hat{c}_1,\hat{c}_2,...)$. Note that the lower is ϑ, the higher is consumption of c_2 and thus U_1 (if uncompensated elasticity is positive). Thus $\gamma(\theta)>0$. Now the above equation can be rewritten as

$$g(z)\gamma_\vartheta - g(z)\gamma\frac{\partial c_2}{\partial z}$$

We derive the formula in the Theorem, and then we solve the differential equation. Following Saez (2001), we have the following effect form increasing the marginal tax rate between z^* and z^*+dz:

1. Direct effect (fiscal and welfare): $\int_{z=z^*}^{\infty}(1-g(z))dH(z)$
2. Compensated elasticity effect (fiscal and welfare): $-\zeta_z^c z\frac{T'}{1-T'}h(z^*)-(\zeta_z^c z/(1-T'))h(z^*)\left[(g(z^*)\bar{\gamma}(z^*)-t)\frac{\partial c_2}{\partial z}-\bar{\gamma}\right.$ or $-\zeta_z^c z^*\frac{T'}{1-T'}h(z^*)+\zeta_z^c z^*\left[\frac{g(z^*)\bar{\gamma}(z^*)-t}{p+t}\eta(z^*)+\frac{\bar{\gamma}_\vartheta}{1-T'}\right]h(z^*)$
   (a) Note: for simplicity I am assuming here an independence between ζ_z^c, η, and γ.
3. Income effect (fiscal and welfare): $\int_{z\geq z^*}\eta_z\frac{T'}{1-T'}h(z)-\int_{z\geq z^*}\eta_z\left[\frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z)+\frac{\bar{\gamma}_\vartheta}{1-T'}\right]h(z)-\int_{z\geq z^*}\frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z)h(z)$ (note that the second term is the income effect on c_2 from the change earnings z, while the third term is the income effect on c_2 from the change in after-tax income $y=z(1-T)$.)

Thus

$$\frac{T'}{1-T'}=\frac{g(z^*)\bar{\gamma}(z^*)-t}{p+t}\eta(z^*)+\frac{\bar{\gamma}_\vartheta}{1-T'}+\frac{1}{\zeta_z^c z^* h(z^*)}\int_{z\geq z^*}(1-\hat{g}(z))h(z)$$

where

$$\hat{g}(z)=g(z)-(1+\eta_z)\frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z)+\eta_z\frac{T'-\bar{\gamma}_\vartheta}{1-T'}$$

B.5.2 Commodity tax

$$\begin{aligned}
\frac{dW}{dt}&=\int_\Theta\left(c_2(\theta)(1-g(\theta))+c_2(\theta)\left(\frac{\zeta^c(\theta)}{p+t}-\frac{\eta(\theta)}{p+t}\right)(t-g(\theta)\gamma(\theta))\right)d\mu(\theta)\\
&-\int_\Theta\vartheta\frac{T'(z(\theta))}{1-T'}\left(\zeta_z^c(\theta)z(\theta)\frac{\partial c_2(z,\theta)}{\partial z}+\eta_z c_2(\theta)\right)d\mu(\theta)\\
&+\int\vartheta z\zeta_z^c\left[\frac{\partial c_2(z,\theta)}{\partial z}\eta(g\gamma-t)+\gamma_\vartheta\right]+\int\vartheta\eta_z c_2\left[\frac{\eta(g\gamma-t)}{p+t}+\bar{\gamma}_\vartheta\right]\\
&=\int_z\left(c_2(1-g(z))+c_2(z)\left(\frac{\zeta^c(z)}{p+t}+\frac{\eta(z)}{p+t}\right)(t-g(z)\gamma(z))\right)h(z) \qquad (22)\\
&-\int_z\vartheta\left(\frac{\partial c_2(z)}{\partial z}+\frac{\eta_z c_2(z)}{z\zeta_z^c}\right)\int(1-\hat{g}(x))dxh(z)dz \qquad (23)
\end{aligned}$$

The welfare impact of an income tax reform that gives back z-earners $c_2(z)dt$ is given by

$$\begin{aligned}
0=\frac{dW}{dT}&=\int_\Theta\left(\bar{c}_2(z(\theta))(1-g(\theta))-\bar{c}_2(z(\theta))\frac{\eta(\theta)}{p+t}(t-g(\theta)\gamma(\theta))\right)d\mu(\theta)\\
&-\int_\Theta\frac{T'(z(\theta))}{1-T'}\left(\zeta_z^c(\theta)z(\theta)\bar{c}_2'(z(\theta))+\eta_z(\theta)\bar{c}_2(z(\theta))\right)d\mu(\theta)\\
&+\int_\Theta z\zeta_z^c c_2'\left[\frac{\eta(g\gamma-t)}{p+t}+\frac{\bar{\gamma}_\vartheta}{1-T'}\right]d\mu(\theta)+\int\eta_z c_2\left[\frac{\eta(g\gamma-t)}{p+t}+\frac{\bar{\gamma}_\vartheta}{1-T'}\right]d\mu(\theta)\\
&=\int\left(c_2(z)(1-g(z))-\bar{c}_2(z)\frac{\eta(z)}{p+t}(t-g(z)\gamma(z))\right)h(z)dz\\
&-\int\left(\bar{c}_2'(z)+\frac{\eta_z c_2(z)}{z\zeta_z^c}\right)\int_{x\geq z}(1-\hat{g}(x))h(x)dz
\end{aligned}$$

Now under assumption A,

$$\begin{aligned}\frac{dW}{dt}-\frac{dW}{dT} &= \int \frac{\bar{c}_2(z)\bar{\zeta}^c(z)}{p+t}\left(t-g(z)\bar{\gamma}(z)\right)h(z)\\ &+\int\left(c_2'(z)-\vartheta\frac{\partial\bar{c}_2(z)}{\partial z}\right)\int_{x\geq z}(1-\hat{g}(x))dxh(z)\\ &+\int(1-\vartheta)\frac{\eta_z\bar{c}_2(z)}{z\zeta_z^c}\int_{x\geq z}(1-\hat{g}(x))dxh(z)dz\end{aligned}$$

When $\eta_z = 0$ and when ϑ is constant, we perform calculations analogous to the proof of Proposition 1 to get the modified formula

$$\begin{aligned}t &= \bar{\gamma}(\bar{g}+\sigma)-\frac{p+t}{\bar{\zeta}^c}Cov\left[\hat{g}(z),(1-\vartheta)c_2+\vartheta\tilde{\phi}(z)\right]\\ &= \frac{\bar{\zeta}^c\bar{\gamma}(\bar{g}+\sigma)-pCov\left[\hat{g}(z),(1-\vartheta)c_2+\vartheta\tilde{\phi}(z)\right]}{\bar{\zeta}^c+Cov\left[\hat{g}(z),(1-\vartheta)c_2+\vartheta\tilde{\phi}(z)\right]}\end{aligned}$$

Note that for the simple formula, we need ϑ to be constant. More generally, when the salience parameter might differ by incomes, we can define a new term $\tilde{\tilde{\phi}}$ equal to $\frac{1}{C_2}\int_{x=0}^{x=z}\left(c_2'(x)-\vartheta(x,t)\frac{\partial c_2(z,\theta)}{\partial z}\right)dx$, and obtain

$$\begin{aligned}t &= \bar{\gamma}(\bar{g}+\sigma)-\frac{p+t}{\bar{\zeta}^c}Cov\left[\hat{g}(z),\tilde{\tilde{\phi}}(z)\right]\\ &= \frac{\bar{\zeta}^c\bar{\gamma}(\bar{g}+\sigma)-pCov\left[\hat{g}(z),\tilde{\tilde{\phi}}(z)\right]}{\bar{\zeta}^c+Cov\left[\hat{g}(z),\tilde{\tilde{\phi}}(z)\right]}\end{aligned}$$

When $\eta_z \neq 0$, we also have to add an additional term given by $n(z) = \frac{1}{C_2}\int_{x=0}^{z}(1-\vartheta)\frac{\eta_z c_2(x)}{z\zeta_z^c}dx$. This corresponds to the additional revenue that can be raised from the commodity tax through the fact that people underestimate the effect it has on the total budget. Incorporating this term leads to the more general formula

$$\begin{aligned}t &= \bar{\gamma}(\bar{g}+\sigma)-\frac{p+t}{\bar{\zeta}^c}Cov\left[\hat{g}(z),\tilde{\tilde{\phi}}(z)+n(z)\right]\\ &= \frac{\bar{\zeta}^c\bar{\gamma}(\bar{g}+\sigma)-pCov\left[\hat{g}(z),\tilde{\tilde{\phi}}(z)+n(z)\right]}{\bar{\zeta}^c+Cov\left[\hat{g}(z),\tilde{\tilde{\phi}}(z)+n(z)\right]}\end{aligned}$$

When $\vartheta = 0$, the formula is particularly simple, as the expression in (23) drops out. Define $\bar{\gamma}^u$ and σ^u analogously to $\bar{\gamma}$ and σ, but with respect to uncompensated demand responses. Then expression

(22) implies that that the optimal tax satisfies

$$\begin{aligned} t &= \bar{\gamma}^u(\bar{g}+\sigma^u) + \frac{p+t}{\bar{\zeta}^c} E\left[(1-g(z))c_2(z)\right] \\ &= \frac{\bar{\zeta}^c\bar{\gamma}^u(\bar{g}+\sigma^u) + pE\left[(1-g(z))c_2(z)\right]}{\bar{\zeta}^c + E\left[(1-g(z))c_2(z)\right]}. \end{aligned}$$

B.6 Many Dimensions of Consumption: Proof of Proposition 4

Analogous to our calculations for the income tax in the proof of Proposition 1, we can show that the income tax must satisfy

$$\frac{T'(z)}{1-T'(z)} = \sum_i \frac{g(z)\bar{\gamma}_i(z) - t_i}{p+t_i}\eta_i(z) + \frac{1}{\zeta_z^c z h(z)} \int_z^\infty (1-\hat{g}(x))dH(x)$$

Under assumption A, the welfare impact of increasing commodity tax t_i is given by

$$\begin{aligned} \frac{dW}{dt_i} &= \overbrace{\sum_j \int \left(\bar{c}_i(z)(1-g(z)) + \frac{\partial \bar{c}_j(z)}{\partial t_i}(t_j - g(z)\gamma_j)\right) dH(z)}^{\text{Direct effect on consumption}} \\ &- \overbrace{\int \frac{T'(z)}{1-T'}\left(\zeta_z^c z \frac{\partial \bar{c}_i(z)}{\partial z} + \eta_z \bar{c}_i\right) dH(z)}^{\text{Direct effect on labor supply}} \\ &+ \overbrace{\sum_j \int \left(\zeta_z^c z \frac{\partial \bar{c}_i(z)}{\partial z} + \eta_z \bar{c}_i\right) \frac{\eta_j(g\gamma_j - t_j)}{p+t_j} dH(z)}^{\text{Indirect effect on consumption through changes in } z} \\ &= \sum_j \int \left(\bar{c}_i(z)(1-g(z)) + \left(\frac{\partial}{\partial t_i} h_j + \frac{\eta_j \bar{c}_i}{p+t_j}\right)(t_j - g(z)\gamma_j)\right) dH(z) \\ &- \int \int_{s \geq z} (1-\hat{g}(s))ds \left(\frac{\partial \bar{c}_i(z)}{\partial z} + \frac{\eta_z \bar{c}_i}{z\zeta_z^c}\right) dH(z) \end{aligned}$$

The welfare impact of an income tax perturbation $dT(z) = dt\bar{c}_i(z)$ is

$$
\begin{aligned}
\frac{dW}{dT} &= \sum_j \int \left(\bar{c}_i(z)(1-g(z)) - \frac{\eta_j \bar{c}_i}{p+t_j}(t_j - g(z)\gamma_j) \right) dH(z) \\
&- \int_\Theta \frac{T'(z)}{1-T'} \left(\zeta_z^c z \bar{c}_i'(z) + \eta_z \bar{c}_i(z) \right) dH(z) \\
&+ \sum_j \int \left(\zeta_z^c z \bar{c}_i'(z) + \eta_z \bar{c}_i \right) \frac{\eta_j (g\gamma_j - t_j)}{p+t_j} dH(z) \\
&= \sum_j \int \left(\bar{c}_i(z)(1-g(z)) - \frac{\eta_j \bar{c}_i}{p+t_j}(t_j - g(z)\gamma_j) \right) dH(z) \\
&- \sum_j \int \int_{s \geq z} (1-\hat{g}(s)) ds \left(\bar{c}_j'(z) + \frac{\eta_z \bar{c}_i}{\frac{\eta_z \bar{c}_i}{z \zeta_z^c}} \right) dH(z)
\end{aligned}
$$

Setting $\frac{dW}{dT} = 0$ and subtracting that from $\frac{dW}{dt}$ shows that

$$
\frac{dW}{dt_i} = \sum_j \int \left(\bar{c}_i(z)(1-g(z)) + \frac{\partial}{\partial t_i} h_j(t_j - g(z)\gamma_j) \right) dH(z) \tag{24}
$$

$$
- \int \int_{s \geq z} (1-\hat{g}(s)) ds \left(\frac{\partial \bar{c}_i(z)}{\partial z} - \bar{c}_j'(z) \right) dH(z) \tag{25}
$$

We can transform equation (25) using integration by parts as in the proof of Proposition 1, exactly as we did in equation (18). This yields the desired result.

C Additional theoretical results

C.1 For discussion of term with conditional covariances

As is apparent from Equation (16), this derivation of the optimal commodity tax using the perturbation approach can accommodate heterogeneity conditional on income. Although Assumption A simplifies the form by assuming that the income-conditional covariances which appear there are equal to zero, that restriction is not necessary for the basic approach, and here we relax it by writing the necessary conditional for the optimal tax with flexible multidimensional heterogeneity:

$$
\begin{aligned}
t = \bar{\gamma}\left(\bar{g} + \sigma\right) &- \frac{p+t}{\bar{\zeta}^c} Cov\left[\hat{g}(z), \tilde{\phi}(z)\right] \\
&- \frac{p+t}{\bar{\zeta}^c} \int_0^\infty \frac{1}{1-T'(z)} \left(g(z) \Sigma^{(z)}_{\eta\gamma, c_2/C_2} - t \Sigma^{(z)}_{\eta, c_2/C_2} \right) dH(z) \\
&- \frac{p+t}{\bar{\zeta}^c} \int_0^\infty \left(\frac{1}{z \bar{\zeta}_z^c(z) C_2} \right) \left(\int_{s \geq z} (1-\hat{g}(s)) ds \left(\Sigma^{(z)}_{\zeta_z^c, \eta} \right) + T'(z) \Sigma^{(z)}_{\eta_z, c_2} \right) dH(z)
\end{aligned} \tag{26}
$$

Equation 26 resembles Equation (3) in Proposition 1, with additional terms corresponding to the

income conditional covariances between income effects, elasticities, and consumption concentration ($c_2(\theta)/C_2$). (Note that although this expression allows for more general multidimensional heterogeneity, it still requires that the first-order approach characterizes the optimal tax system—i.e., heterogeneity must not lead to discontinuous jumping of earnings or consumption bundles in response to small tax perturbations.) As with Proposition 1, this equation is implicit, and thus it does not represent a closed form solution for the optimal commodity tax—however it could be solved numerically via fixed point iteration, like the simulations in Section 6. Additionally, it provides a heuristic guide to the likely comparative static effects of the income-conditional covariance terms on the optimal tax. For example, focusing on the final term, suppose $\Sigma^{(z)}_{\eta_z,c_2} > 0$ for all z, indicating that labor supply income effects are smallest in magnitude (least negative) among those who consume the most c_2 at each income—then the optimal commodity tax will tend to be reduced. Intuitively, the joint reform of a higher t and a regressivity-offsetting income tax reform confers a greater benefit on those who consume little c_2 at each income. If those individuals exhibit larger labor supply responses to such benefits, then the negative fiscal externality (through the income tax) of this joint reform is larger, suggesting a lower optimal income tax. Similar reasoning can be applied to the other covariance terms.

C.2 Optimal taxes when $\eta_z = 0$ but η is not negligible

Corollary 6. *Suppose that labor supply income effects are negligible $\eta_z \approx 0$. Then*

$$t \approx \frac{\bar{\zeta}^c\bar{\gamma}(\bar{g}+\sigma) + Cov[\tilde{\phi}(z), \eta g(z)\bar{\gamma}(z)]}{\bar{\zeta}^c + Cov\left[g(z)+\eta, \tilde{\phi}(z)\right]} - p\frac{Cov\left[g(z), \tilde{\phi}(z)\right]}{\bar{\zeta}^c + Cov\left[g(z)+\eta, \tilde{\phi}(z)\right]}$$

$$\frac{T'(z)}{1-T'(z)} \approx \tfrac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z) + \frac{1}{\zeta^c_z z h(z)}\left[\int_z^\infty \frac{g(x)\bar{\gamma}(x)-t}{p+t}\eta(x)(1+\eta_z(x))dH(x) + \int_z^\infty (1-g(x))dH(x)\right]$$

Proof. We set $\eta_z = 0$ in equation (3), with

$$\hat{g}(z) = g(z) - (1+\eta_z)\frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z) + \eta_z\frac{T'}{1-T'}$$

to get

$$\begin{aligned} t &= \bar{\gamma}(\bar{g}+\sigma) - \frac{p+t}{\bar{\zeta}^c}Cov\left[g(z) - \frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z), \tilde{\phi}(z)\right] \\ &= \bar{\gamma}(\bar{g}+\sigma) - \frac{p+t}{\bar{\zeta}^c}Cov\left[g(z), \tilde{\phi}(z)\right] + \frac{1}{\bar{\zeta}^c}Cov\left[(g(z)\bar{\gamma}(z)-t)\eta(z), \tilde{\phi}(z)\right] \end{aligned} \tag{27}$$

Solving equation (27) for t yields the expression in the Corollary above. □

C.3 Solving the differential equation for the optimal income tax

We now solve the differential equation for the optimal income tax. To that end, define $Y(z) = \int_z^\infty -\frac{\eta_z T'(x)}{1-T'(x)} h(x)dx$. Then the equation comes down to

$$\frac{1}{\eta_z h(z)} Y' - \frac{1}{zh(z)\zeta^c} Y = \frac{1}{zh(z)\zeta^c} \int_z^\infty (1-\hat{g}(x))h(x)dx + \frac{\bar{\gamma} g\eta - t}{p+t}$$

or

$$Y'(z) - B(z)Y(z) = B(z)A(z) + D(z)$$

where $A(z) = \int_z^\infty \left(1 - g(x) + \frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(x)(\eta_z(x)+1)\right) h(x)dx$ and $B(z) = \frac{\eta_z}{z\zeta_z^c}$, $D(z) = \frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z)\eta_z h(z)$. This gives us a first order differential equation, which can be solved for Y

$$\begin{aligned} Y &= -\frac{\int_{x=z}^\infty e^{\int^x -B(x')dx'} \left(A(x)B(x) + D(x)\right) dx + K}{e^{\int^z -B(x')dx'}} \\ &= -\frac{\int_{x=z}^\infty e^{\int^x -B(x')dx'} D(x)dx}{e^{\int^z -B(x')dx'}} - \frac{\int_{x=z}^\infty e^{\int^x -B(x')dx'} A(x)B(x)dx + K}{e^{\int^z -B(x')dx'}} \end{aligned}$$

where K is a constant of integration. But because $Y(\infty) = 0$, we must have $K = 0$. Thus

$$\begin{aligned} Y' &= D(z) - \frac{B(z)\int_{x=z}^\infty e^{\int^x -B(x')dx'} D(x)dx}{e^{\int^z -B(x')dx'}} \\ &- \frac{-A(z)B(z)e^{\int^z -B(x')dx'} + B(z)\int_{x=z}^\infty e^{\int^x -B(x')dx'} A(x)B(x)dx}{e^{\int^z -B(x')dx'}} \end{aligned}$$

Now integration by parts shows that

$$\begin{aligned} \int_{x=z}^\infty e^{\int^x -B(x')dx'} A(x)B(x)dx &= -A(x)e^{\int^x -B(x')dx'}|_{x=z}^{x=\infty} - \int_{x=z}^\infty (1-\tilde{g}(x))h(x)e^{\int^x -B(x')dx'}dx \\ &= A(z)e^{\int^z -B(x')dx'} + \int_{x=z}^\infty (1-\tilde{g}(x))h(x)e^{\int^x -B(x')dx'}dx. \end{aligned}$$

Thus

$$Y' = D(z)e^{\int^z B(x)dx} + B(z)\int_{x=z}^\infty e^{\int_{x'=z}^{x'=x} -B(x')dx'}(D(x) + (1-\tilde{g}(x))h(x)dx$$

and so

$$\frac{T'}{1-T'} = \frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z) + \frac{1}{zh(z)\zeta_z^c}\int_z^\infty \left(1 - g(x) - \frac{g(x)\bar{\gamma}(x)-t}{p+t}\eta(x)\eta_z(x)\right) e^{\int_{x'=z}^{x'=x} -\eta_z/(z\zeta_z^c)dx'} h(x)dx.$$

This formula is the same formula as Saez (2001), except with 1) $g(z)$ replaced by $g(x)+\frac{g(x)\bar{\gamma}(x)-t}{p+t}\eta(x)\eta_z(x)$, and also the presence of the term $\frac{g(z)\bar{\gamma}(z)-t}{p+t}\eta(z)$.

D Non-differentiable labor supply choices

We assume that types θ are unidimensional. Let $\hat{U}(y,z,\theta)$ denote a type θ's indirect utility from after-tax income y and before-tax income z.

We solve for $\chi(\theta)$. We need for $d\theta > 0$ and approaching 0,

$$\lim_{d\theta\to 0} \hat{U}(y(\theta+d\theta) - dtc_2(\theta+d\theta, z(\theta+d\theta)) + \chi(\theta+d\theta), z(\theta+d\theta), \theta+d\theta) - \hat{U}(y(\theta) - dtc_2(\theta+d\theta, z(\theta)) + \chi(\theta), z(\theta$$
$$= \lim_{d\theta\to 0} \hat{U}(y(\theta+d\theta), z(\theta+d\theta), \theta+d\theta) - \hat{U}(y(\theta), z(\theta), \theta+d\theta)$$

We now compute $\chi(\theta+d\theta) - \chi(\theta)$. We Taylor expand to get that

$$\begin{aligned}
&(-dtc_2(\theta+d\theta, z(\theta+d\theta)) + \chi(\theta+d\theta))\hat{U}_1(y(\theta+d\theta), z(\theta+d\theta), \theta+d\theta) \\
&+(dtc_2(\theta+d\theta, z(\theta)) - \chi(\theta))\hat{U}_1(y(\theta), z(\theta), \theta+d\theta) \\
&=O((dt)^2)
\end{aligned}$$

and thus that

$$\begin{aligned}
&\lim_{d\theta\to 0} \chi(\theta+d\theta) - \chi(\theta) \\
&= \lim_{d\theta\to 0} -\chi(\theta)\frac{A(\theta,\theta+d\theta)}{\hat{U}_1(y(\theta+d\theta), z(\theta+d\theta), \theta+d\theta)} \\
&+dt\left[c_2(\theta+d\theta, z(\theta+d\theta)) - c_2(\theta+d\theta, z(\theta)) + c_2(\theta+d\theta, z(\theta))\frac{A(\theta,\theta+d\theta)}{\hat{U}_1(y(\theta+d\theta), z(\theta+d\theta)), z(\theta+d\theta); \theta+d\theta)}\right]
\end{aligned} \tag{28}$$

where

$$A(\theta,\theta+d\theta) = \hat{U}_1(y(\theta+d\theta), z(\theta+d\theta); \theta+d\theta) - \hat{U}_1(y(\theta), z(\theta); \theta+d\theta)$$

Note that when z is differentiable at θ, $\lim_{d\theta\to 0} A(\theta,\theta+d\theta) = [U_{11}y'(\theta) + U_{12}z'(\theta)]\, d\theta$, or $[(1-\tau)U_{11} + U_{12}]z'(\theta)d\theta$.

Now from the first order condition $(1-\tau)U_1 + U_2 = 0$, an exogenous change in income y on labor earnings z is

$$\frac{dz}{dy} \propto (1-\tau)U_{11} + U_{12}$$

while the compensated response with respect to $(1-\tau)$ is

$$\frac{dz^c}{d(1-\tau)} \propto U_{1.}$$

Now since $\zeta_z^c = \frac{dz^c}{d(1-\tau)} \cdot \frac{1-\tau}{z}$ and since $\eta_z = \frac{dz}{dy} \cdot (1-\tau)$, we have that at all points of differentiability,

$$\lim_{d\theta \to 0} \frac{A(\theta, \theta + d\theta)}{\hat{U}_1(y(\theta + d\theta, z(\theta + d\theta)), z(\theta + d\theta); \theta + d\theta)} = \frac{\eta_z(\theta)}{z(\theta)\zeta_z^c(\theta)} z'(\theta) d\theta$$

When c_1, c_2, z are differentiable at θ, we can divide both sides by $d\theta$ and arrive at the differential equation

$$\chi'(\theta) + \chi(\theta) \frac{\eta_z(\theta)}{z(\theta)\zeta_z^c(\theta)} z'(\theta) = \frac{\partial c_2(z, \theta)}{\partial z} z'(\theta) + \frac{\eta_z(\theta)}{z(\theta)\zeta_z^c(\theta)} c_2(z, \theta) z'(\theta) \tag{29}$$

We can solve the first-order ODE (29) in any region in which $z(\theta)$ is differentiable. We can then solve the general equation (28) by setting proper initial conditions for the differential equation solutions so as to guarantee that the pasted together differential equation solutions satisfy the general equation at points of non-differentiability.

The solution is particularly simple when there are no labor supply income effects. In this case, it follows that $A(\theta, \theta + d\theta) = 0$, even at discontinuities, and thus

$$\lim_{d\theta \to 0} [\chi(\theta + d\theta) - \chi(\theta)] = \lim_{d\theta \to 0} [c_2(\theta + d\theta, z(\theta + d\theta)) - c_2(\theta + d\theta, z(\theta))]$$

Now for $z \in \lim_{d\theta \to 0} [z(\theta), z(\theta + |d\theta|)]$, define

$$\omega(z) = \lim_{d\theta \to 0} \frac{c_2(\theta + |d\theta|, z(\theta + |d\theta|)) - c_2(\theta + |d\theta|, z(\theta))}{z(\theta + |d\theta|) - z(\theta)}$$

Then as before, it follows that if $z^*(\theta) = z$, then $\chi(\theta) = c_2(0) + \int_{x=0}^{z} \omega(z) dz$. Note that in cases in which z exhibits a jump at point θ, $\omega(z)$ is simply the slope of the line connecting $(z(\theta), c_2(z(\theta)))$ to $(z(\theta + d\theta), c_2(z(\theta + d\theta)))$.

In the general case, suppose that $z(\theta)$ is continuous on $[z_0, z_1], (z_2, z_3], (z_4, z_5], \ldots$ In between the intervals there are gaps in the earnings distribution, and there is possibly bunching at the endpoints. Now for $z \in [z_{2i}, z_{2i+1}]$, analogous computations show that $\chi(z)$ must satisfy

$$\chi(z) = c_2(z_{2i}) + \underbrace{\int_{x=z_i}^{z} \omega(x) dx}_{\text{amount of } c_2 \text{ variation explained by income effects}} + \underbrace{\int_{z_i}^{z} w(x, z) \frac{\eta_z}{\bar{\zeta}_z^c x} (c_2(x) - \phi(x)) dx}_{\text{labor supply income effect adjustment}}. \tag{30}$$

or alternatively, in terms of types such that $z(\theta)$ is continuous except at $\theta_1, \theta_2, \ldots$, where $z(\theta)$ exhibits a jump.

$$\chi(\theta) = c_2(\theta_i) + \underbrace{\int_{\vartheta=\theta_i}^{\theta} \omega(\vartheta)x'(\vartheta)d\vartheta}_{\text{amount of } c_2 \text{ variation explained by income effects}} + \underbrace{\int_{\vartheta=\theta_i}^{\vartheta=\theta} w(x(\vartheta), z(\vartheta))\frac{\eta_z}{\bar{\zeta}_z^c x}(c_2(\vartheta) - \phi(\vartheta))d\vartheta}_{\text{labor supply income effect adjustment}}. \tag{31}$$

Next, for $z_{2i-1} = z^*(\theta)$ and $z_{2i} = \lim_{d\theta\to} z^*(\theta + |d\theta|)$, we determine $\chi(\theta_i + d\theta) - \chi(\theta_i)$ to satisfy (28). In particular, we have

$$\lim_{\theta\to\theta_i^+} \chi(\theta) = c_2(z(\theta), \theta) - [(c_2(\theta_i), \theta) - \chi(\theta_i)] \frac{A(\theta_i, \theta)}{\hat{U}_1(y(\theta), z(\theta), z(\theta); \theta)}.$$

E Analytic derivations for numerical simulations

The first-order condition governing choice of soda consumption is

$$-(1+t)\left(1 + v_c'(s, c, \theta)\right) + v_s'(s, c, \theta) = 0$$

with

$$v_s'(s, c, \theta) = a(c, \theta)s^{-k}$$

and

$$v_c'(s, c, \theta) = a_c'(c, \theta)\left(\frac{s^{1-k}}{1-k}\right)$$

The expressions above are all conditional on type θ. For some of the following expressions, we also need the partial derivative of v with respect to type:

$$v_\theta'(s, c, \theta) = a_\theta'(c, \theta)\left(\frac{s^{1-k}}{1-k}\right)$$

Second-order condition (SOC) for interior solution:

$$v_{ss}''(s, c, \theta) - 2v_{sc}''(s, c, \theta)(1+t) + v_{cc}''(s, c, \theta)(1+t)^2 < 0$$

with

$$v_{ss}''(s, c, \theta) = -a(c, \theta)ks^{-k-1}$$

$$v_{sc}''(s, c, \theta) = a_c'(c, \theta)s^{-k} - a(c, \theta)s^{-k}\log(s)$$

$$v_{cc}''(s, c, \theta) = a_{cc}''(c, \theta)\left(\frac{s^{1-k}}{1-k}\right)$$

Compensated demand response of soda:

$$\left.\frac{ds}{dt}\right|_c = \frac{1 + v_c'(s, c, \theta)}{SOC}$$

Now with the above expressions, can solve for the $a(c,\theta)$ which generate the observed pattern of soda consumption levels and desired pattern of elasticities. As described in the text, we assume $a(c,\theta)$ is the product of two functions, each of which is a univariate function: $a(c,\theta) = a_c(c)a_\theta(\theta)$, so that $a'_c(c,\theta) = a'_c(c)a_\theta(\theta)$, $a''_{cc}(c,\theta) = a''_c(c)a_\theta(\theta)$, $a''_{c\theta}(c,\theta) = a'_c(c)a'_\theta(\theta)$, etc.

For pure preference heterogeneity, a is a function of θ only; for pure income effects, they are functions of c only. For intermediate values, there's a free degree of freedom determining how much a and k are changing c vs. θ—that degree of freedom maps monotonically to the aggregate statistic Φ as defined in the text.

The cross-sectional derivative of s with respect to income z is can be decomposed into an income effect and a preference heterogeneity effect. The income effect is the change in soda consumption which would come from giving someone $1 - T'(z)$ more of post-tax consumption, while holding their type θ fixed:

$$\psi(\theta) = \frac{1 - T'(z)}{SOC}\left((1+t)v''_{cc}(s,c,\theta) - v''_{sc}(s,c,\theta)\right)$$

The "type effect" is the variation in soda consumption that comes solely from the type difference (preference heterogeneity) across earners, holding their income constant:

$$\frac{\partial s}{\partial \theta}\frac{\partial \theta}{\partial z} = \frac{1}{SOC}\left((1+t)v''_{c\theta}(s,c,\theta) - v''_{s\theta}(s,c,\theta)\right)\frac{\partial \theta}{\partial z}$$

with

$$v''_{c\theta}(s,c,\theta) = a''_{c\theta}(c,\theta)\left(\frac{s^{1-k}}{1-k}\right)$$

and

$$v''_{s\theta}(s,c,\theta) = a'_\theta(c,\theta)s^{-k}$$

The total cross-sectional variation in soda consumption is the sum of these two effects

$$\frac{ds}{dz} = \frac{1}{SOC}\left[\left(1 - T'(z)\right)\left((1+t)v''_{cc}(s,c,\theta) - v''_{sc}(s,c,\theta)\right) + \frac{\partial \theta}{\partial z}\left((1+t)v''_{c\theta}(s,c,\theta) - (1+t)v''_{c\theta}(s,c,\theta)\right)\right]$$

Labor supply FOC:

$$\left(1 - T'(z)\right)\left(1 + v'_c(s,c,\theta)\right) - \frac{\Psi'\left(\frac{z}{w_\theta}\right)}{w_\theta} = 0$$

Solve for w:

$$w_\theta = \left(\frac{z^{1/\varepsilon}}{\left(1 - T'(z)\right)\left(1 + v'_c(s,c,\theta)\right)}\right)^{\frac{\varepsilon}{1+\varepsilon}}$$

And for z (note c and $T'(z)$ are endogenous to z, but by updating those in term during the numerical

iteration, we'll converge to the solution):

$$z = w_\theta^{1+\varepsilon} \left(\left(1 - T'(z)\right) \left(1 + v_c'(s,c,\theta)\right) \right)^\varepsilon$$

And labor supply second-order condition SOC_z:

$$\left(1 - T'(z)\right)^2 v_{cc}''(s,c,\theta) - \left(1 + v_c'(s,c,\theta)\right) T''(z) - \frac{\Psi''\left(\frac{z}{w_\theta}\right)}{w_\theta^2} < 0$$

Define income tax perturbation that is relevant for computing labor responses. Let

$$\hat{T}(z; z_\theta, \rho, \tau) = T(z) + \tau(z - z_\theta) - \rho$$

so that τ represents a local perturbation of the marginal tax rate, and ρ represents an increase in income, both of which preserve the tax function curvature at z_θ. Then the income effect is $\frac{dz}{d\rho}$ evaluated at $\rho = 0$ and $\tau = 0$. Income effect on labor supply:

$$\eta_z(\theta) = \frac{\left(1 - T'(z)\right) v_{cc}''(s,c,\theta)}{SOC_z}$$

and compensated ETI:

$$\zeta_z^c(\theta) = \frac{1 + v_c'(s,c,\theta)}{SOC_z} \frac{1 - T'(z_\theta)}{z_\theta}$$

Bias:

$$\gamma(\theta) = \frac{U_s}{U_c} - \frac{V_s}{V_c} = \frac{\tilde{\gamma}}{1 + v_c'(s,c,\theta)}$$

The marginal social utility of net consumption c_{net} is:

$$\begin{aligned}\frac{dV(c,s,z,\theta)}{dc_{net}} &= \left(1 + v_c'(s,c,\theta)\right) - \frac{\psi(\theta)}{1 - T'(z)}(1+t)\left(1 + v_c'(s,c,\theta)\right) + \left(\frac{\psi(\theta)}{1 - T'(z)}\right) v_s'(s,c,\theta) - \tilde{\gamma}\left(\frac{\psi(\theta)}{1 - T'(z)}\right) \\ &= 1 + v_c'(s,c,\theta) - \tilde{\gamma}\left(\frac{\psi(\theta)}{1 - T'(z)}\right)\end{aligned}$$

And in this specification, there is no misoptimization wedge between c and z, though that does imply a wedge between c_{net} and z.

Since utility is nearly quasilinear, we introduce declining Pareto weights to generate a redistributive motive. These are denoted $\alpha(\theta)$, and we set them so that $\alpha(\theta)$ is declining linearly with income percentile, equaling zero at the top. (The results are not particularly sensitive to the precise pattern of Pareto weights, provided they are exogenous to the equilibrium.) Marginal social utility from additional c:

$$g(\theta)\lambda = \alpha(\theta)\left(1 + v_c'(s,c,\theta)\right)$$

Marginal value of a marginal dollar of income tax reduction for θ, accounting for fiscal externalities

and biases from misallocation:

$$\hat{g}(\theta) = g(\theta) + \psi(\theta)\left(\frac{1}{1-T'(z_\theta)} + \eta_z(\theta)\right)(t - \gamma(\theta)g(\theta)) + \eta_z(\theta)T'(z)$$

Marginal value of public funds satisfies

$$\lambda = \int_\Theta \hat{g}(\theta)\lambda d\mu(\theta)$$

Solving for λ:

$$\lambda = \frac{\int_\Theta \alpha(\theta)\left(1 + v_c'(s,c,\theta)\right)\left(1 - \psi(\theta)\left(\frac{1}{1-T'(z_\theta)} + \eta_z(\theta)\right)\gamma(\theta)\right)d\mu(\theta)}{1 - \int_\Theta \left(\psi(\theta)\left(\frac{1}{1-T'(z_\theta)} + \eta_z(\theta)\right)t + \eta_z(\theta)T'(z_\theta)\right)d\mu(\theta)}$$

Optimal income tax:

$$\int_z^\infty (1 - \hat{g}(\theta(y)))dF(y) + f(z)\left(\left.\frac{dz_\theta}{d\tau}\right|_c\right)\left(T'(z) + \psi(\theta)\left(t - g(\theta)\gamma(\theta)\right)\right) = 0$$

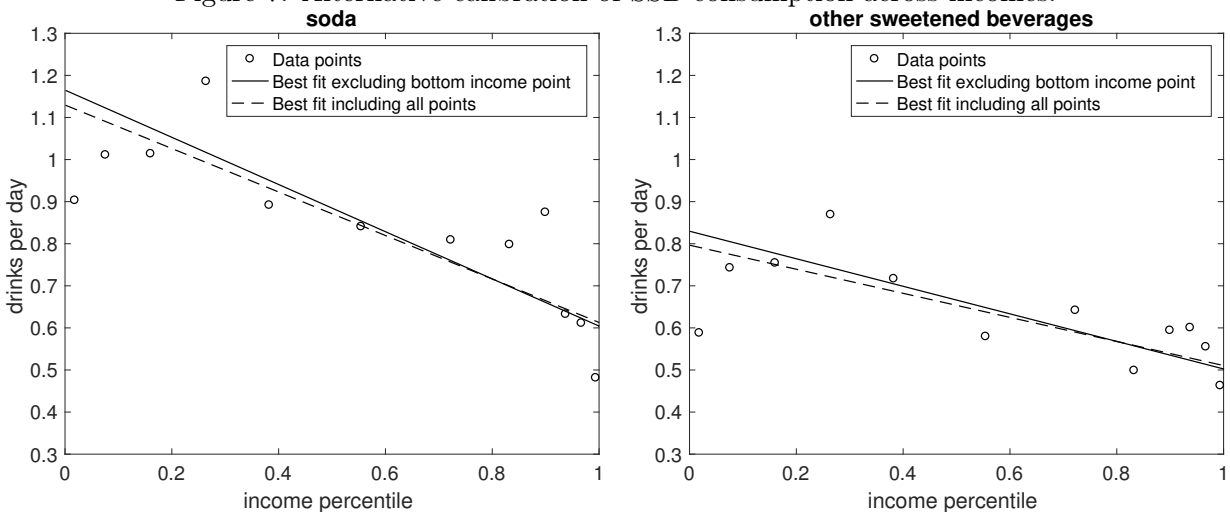
$$T'(z) = -\frac{1}{f(z)\left(\left.\frac{dz_\theta}{d\tau}\right|_c\right)}\int_z^\infty (1 - \hat{g}(\theta(y)))dF(y) - \psi(\theta)\left(t - g(\theta)\gamma(\theta)\right)$$

F Alternative numerical results and robustness checks

F.1 Alternative calibration of sweetened beverage consumption

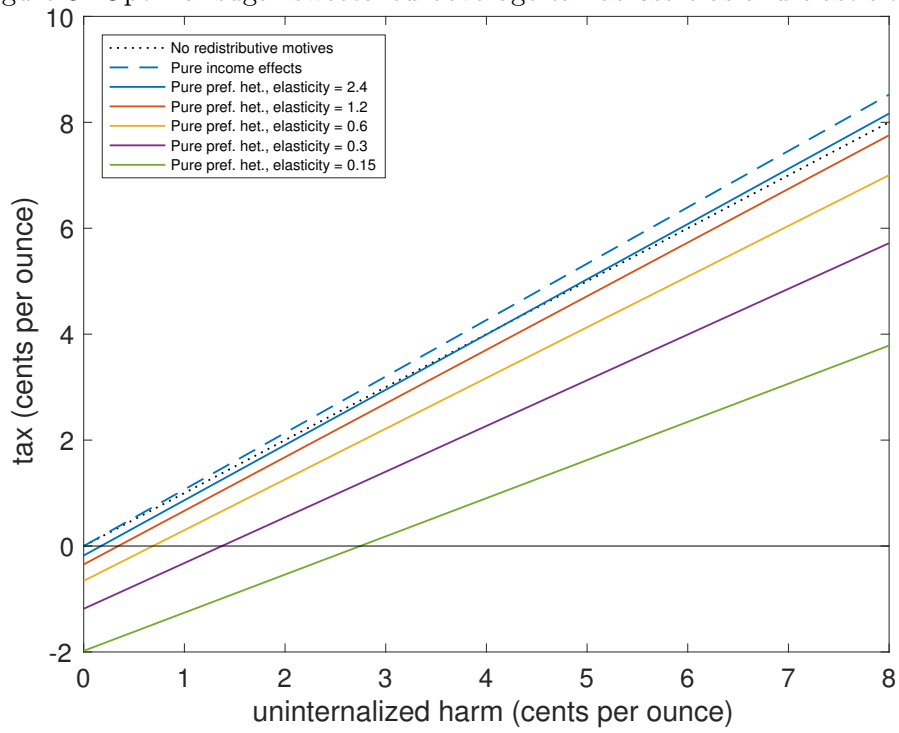
The baseline calibration of sugar-sweetened beverage consumption across incomes excludes the data points for the lowest income bracket from the National Health and Nutrition Examination Survey, which are outliers relative to neighboring income brackets, raising the possibility that they reflect income misreporting. Figures 7 and 8 check the robustness of this restriction. Figure 7 plots the alternative schedule of SSB consumption across incomes which results when the lowest income bracket is included. Figure 8 reproduces Figure 2 from the text under this alternative calibration assumption, and shows that the optimal tax magnitudes are not sensitive to this calibration decision.

Figure 7: Alternative calibration of SSB consumption across incomes.



This figure plots the number of soda drinks and other sugar-sweetened beverages (SSBs) consumed across the income distribution, according to survey data drawn from the National Health and Nutrition Examination Survey in 2009–2010. Whereas the baseline calibration in the text excludes the lowest income bracket, which reports a lower level of SSB consumption than neighboring brackets, the dotted line in this specification shows the effect of including that point in the calibration.

Figure 8: Optimal sugar-sweetened beverage tax across bias and elasticities



This figure reproduces Figure 2 from the text, using the SSB consumption calibration which includes the data point for the lowest income bracket (the dotted line in Figure 7).

F.2 Alternative redistributive preferences

In the baseline simulations, we use Pareto weights equal to $\alpha(\theta) = 1/c_{US}(\theta)$, where $c_{US}(\theta)$ is θ's net income under the US status quo income tax (approximated by a flat tax for simplicity). This generates weights comparable to those which arise under optimal tax models with logarithmic utility from consumption. To explore the effect of the strength of redistributive preferences, in Figure 9 we plot optimal soda taxes when Pareto weights are instead given by $\alpha(\theta) = c_{US}(\theta)^{-0.25}$ and $\alpha(\theta) = c_{US}(\theta)^{-4}$, representing substantially weaker and stronger redistributive preferences, respectively.

F.3 Optimal tax computed using the simplification in Corollary 1

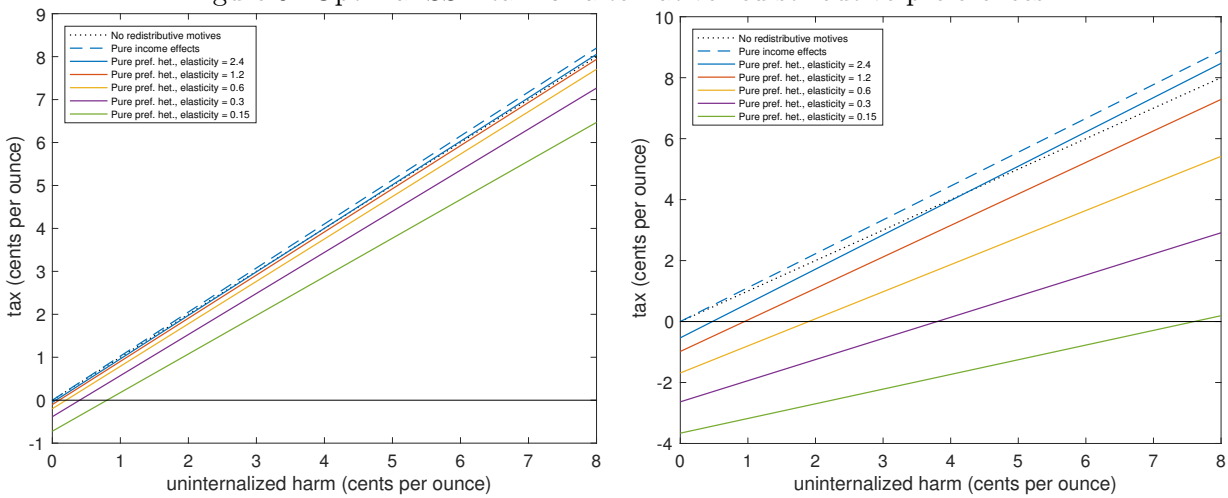
Figure 10 reproduces Figure 3 from the body of the paper, but using the simplification in Corollary 1, wherein the optimal tax is computed replacing the weights $\hat{g}$ with g, which exclude income effects. Comparing the figures verifies that the approximation is quite accurate.

F.4 Tax differences under alternative elasticity values

Figure 11 displays the equivalent of Figure 5 from the body of the paper, which shows the change in total SSB taxes and income taxes paid across the income distribution (relative to the constrained optimum without an SSB tax), here shown for higher and lower values of the SSB demand elasticities.

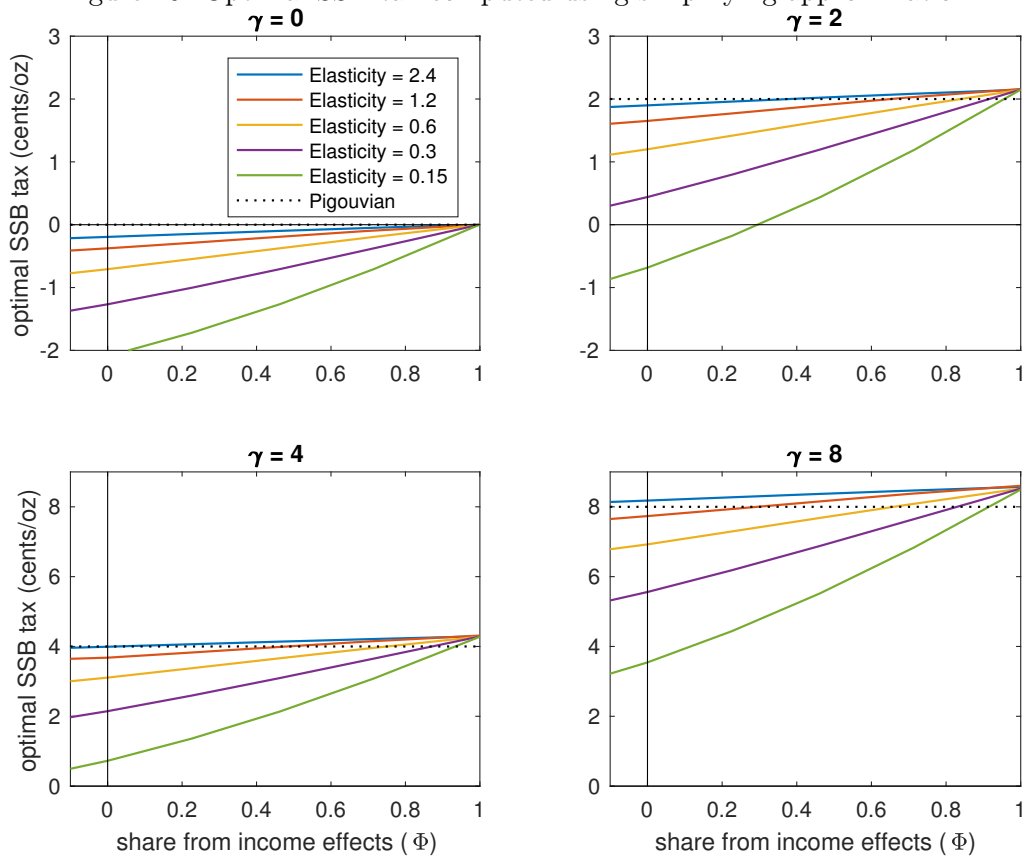
Appendix Figures

Figure 9: Optimal SSB tax for alternative redistributive preferences



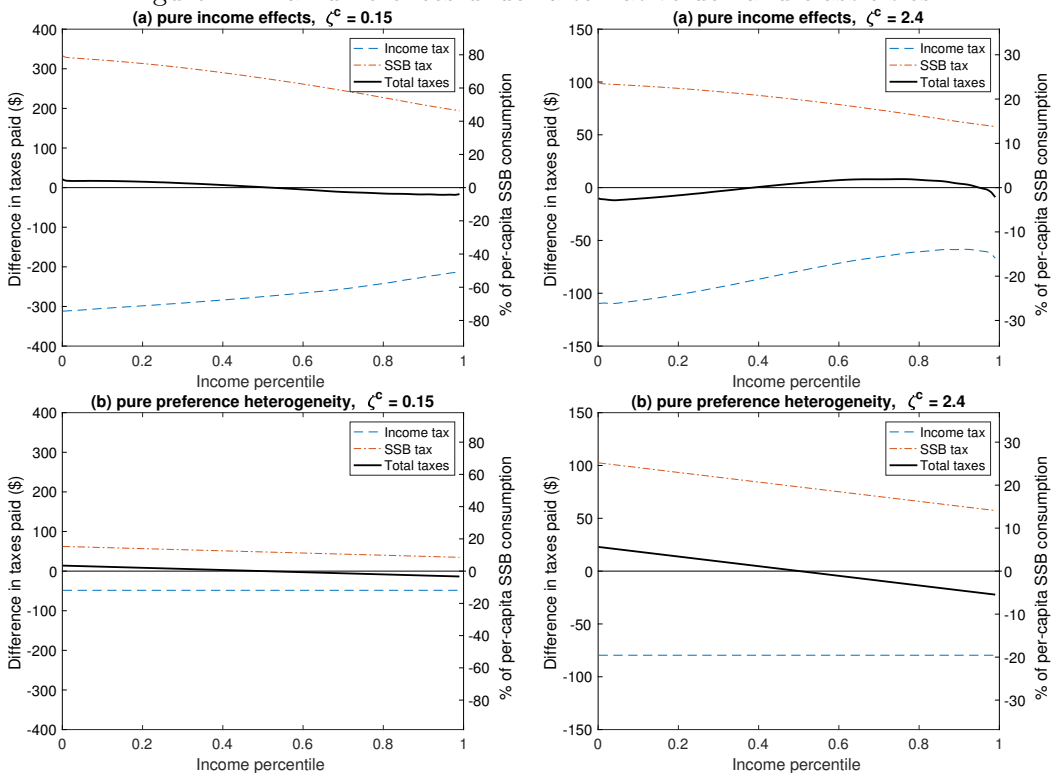
This figure plots optimal soda taxes for alternative redistributive preferences, with declining Pareto weights equal to $\alpha = c_{US}^{-0.25}$ and $\alpha = c_{US}^{-4}$.

Figure 10: Optimal SSB tax computed using simplifying approximation



This figure reproduces Figure 3 from the paper, but with a simplification that replaces the weights $\hat{g}$ with g.

Figure 11: Tax differences under alternative demand elasticities



This figure reproduces Figure 5 from the text, but with alternative demand elasticities for sugar-sweetend beverages.