Regressive Sin Taxes

Benjamin B. Lockwood and Dmitry Taubinsky*

June 19, 2015

Abstract

How should governments structure tax policy in the presence of both consumer mistakes and large wealth inequality? A common objection to “sin taxes”—commodity taxes intended to correct overconsumption of harmful goods—is that many such taxes are regressive, falling largely on low income consumers. This paper extends the literature on “internality taxes”—taxes intended to correct overconsumption due to consumer misoptimization—by studying the interaction between corrective and redistributive motives, and shows how these motives can either dampen or amplify each other. We derive general, elasticity-based formulas for optimal taxes, and we show that the optimal tax can be computed as a function of a few estimable sufficient statistics: the price elasticity of demand, and the covariances between consumer bias, demand elasticities, and consumers’ incomes. We show that these covariances determine whether the optimal sin tax rises, falls, or becomes negative (a “sin subsidy”) as redistributive motives increase. We also show that the dependence of the optimal tax on these covariances is unique to internality taxation, and does not apply to externality taxation (including targeted externalities). Finally, we extend our analysis to the optimal use of nonsalient tax instruments and to the role of persuasive advertising, such as graphic warning labels, and we present conditions under which such unconventional policy instruments are strictly superior to corrective taxes. Quantitatively, we apply our model to cigarette taxes, and we use numerical simulations to trace out optimal tax policy as a function of consumer bias, the strength of redistributive motives, and the demand elasticity. We show that for the range of elasticities typically documented in empirical work, the optimal cigarette tax remains positive, but is significantly dampened by redistributive concerns.

*We thank Raj Chetty, Stefano DellaVigna, Emmanuel Farhi, Xavier Gabaix, David Laibson, Matthew Rabin, Alex Rees-Jones, Emmanuel Saez, Florian Scheuer, Matthew Weinzierl, participants at the Behavioral Economics Annual Meeting, participants and the Berkeley Psychology and Economics lunch seminar, and participants at the George Mason Behavioral Economics Seminar for helpful comments and discussions.
1 Introduction

Tax policy is shaped by multiple social objectives that are not always aligned. One objective, emphasized by a growing body of work in behavioral public finance, is to address market failures due to consumer misoptimization. Governments tax goods such as cigarettes (Gruber and Kőszegi, 2004) and unhealthy foods (O’Donoghue and Rabin, 2006) that people are believed to consume too much, and create subsidies for goods such as energy-efficient products that people are believed to consume too little (Allcott and Taubinsky, forthcoming). However, many such policies are regressive. Cigarettes are overwhelmingly consumed by low-income earners (Gruber and Kőszegi, 2004), while energy-efficiency subsidies are overwhelmingly taken up by high income earners (Allcott, Knittel and Taubinsky, 2015). A government concerned with reducing income inequality – as in the large body of work in optimal tax theory initiated by Vickrey (1945) and Mirrlees (1971) – may thus face a tension between redistributive and corrective motives. In the absence of consumer mistakes (and externalities), goods that are preferred by low income earners – such as cigarettes – should be subsidized by a government attempting to reduce wealth inequality (Saez, 2002b).

How should a government concerned with both wealth inequality and consumer mistakes decide what taxes or subsidies to set for cigarettes, unhealthy foods, or energy efficient products? Although the objectives of correcting consumer mistakes and reducing wealth inequality have mostly been analyzed in isolation, many pressing policy questions involve an interplay between these two motives. The aim of this paper is to provide a bridge between the behavioral economics literature characterizing policy in the presence of consumer mistakes and the optimal tax literature (e.g., Mirrlees, 1971, Atkinson and Stiglitz 1976, Saez 2001, 2002) characterizing taxes in the presence of income inequality and redistributive motives. Our model nests the models considered in these two literatures, in particular building on and extending the crucial insights first advanced in Gruber and Kőszegi (2004).

The results in our paper can be categorized into five sets of contributions. The first is a simple “sufficient statistics” formula for the optimal tax rate on internality-producing goods which are consumed differentially across the income distribution. (For more on the sufficient statistics approach, see Saez 2001 and Chetty 2009 in public finance, and Chetty, Looney and Kroft, 2009; Mullainathan, Schwartzstein and Congdon, 2012; Allcott and Taubinsky, forthcoming; Farhi and Gabaix, 2015 in behavioral economics.) The formula shows that the optimal tax can be computed from a few high-level statistics: the price elasticity of demand for the good in question, and the covariances between the degree of overconsumption (or

\footnote{In a (government-provided) health-insurance setting, Baicker et al. (2015) also suggest lowering copays for certain chronic-disease medication that people may not be adhering to due to psychological biases.}
underconsumption), the elasticity of demand, and consumers’ incomes. The formula also clearly indicates whether, and to what extent, redistributive concerns counteract or magnify the corrective role of sin taxes. Second, we supplement this formula with sharp comparative statics results about how the optimal sin tax changes with the size of consumer bias, the strength of redistributive motives, and the price elasticity of demand. Third, we reexamine the common intuition that internalities behave like targeted externalities. Although this intuition may be a reasonable quantitative guide in some circumstances, we demonstrate that the importance of “correlated heterogeneity”—for example the covariance between demand elasticity and consumers’ incomes—is unique to internality taxes, and does not affect optimal externality taxes. Fourth, we demonstrate the applicability of our framework by presenting calibrations of optimal cigarette taxes, accounting for the high concentration of smokers among the poor. We show that under the most plausible parameter ranges, the optimal cigarette remains positive, but is substantially reduced in the presence of redistributive concerns. Fifth, we consider alternative instruments such as nonsalient taxes or cigarette warning labels, and we show how redistributive concerns can render such instruments strictly superior to conventional fully salient taxes—results which do not generally hold in the absence of redistributive motives.

Our model builds on Saez’s (2002a) extension of Atkinson and Stiglitz (1976) by considering an economy of consumers with heterogenous earnings abilities and tastes, who choose labor supply and a consumption bundle that exhausts their after-tax income. The policymaker chooses a non-linear income tax and a set of linear commodity taxes. 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 that assumption does not hold because of various possible psychological biases such as present bias, limited inattention, or incorrect beliefs. This approach nests O’Donoghue and Rabin (2006) and Allcott and Taubinsky (forthcoming) as special cases, and allows us to generate robust economic insights that apply to the variety of psychological biases that policymakers have invoked to justify specific policies such energy-efficiency subsidies, incandescent lightbulb bans, and taxes on unhealthy foods.

Our baseline analysis in Section 2 shows that the interaction between corrective and redistributive motives are nuanced: the two motives can both oppose and amplify each other. On the one hand, redistributive motives directly oppose corrective motives when the taxes are regressive. Just as Atkinson and Stiglitz (1976) and Saez (2002b) show that goods preferred by low income consumers should be subsidized and those preferred by high earners should be taxed, our general framework suggests that redistributive motives can push against corrective motives. At the same time, redistributive motives can also amplify
corrective motives: an inequality averse policymaker will be more concerned with low-income earners leaving money on the table than with high-income earners leaving money on the table. Thus in contrast to the Mullainathan, Schwartzstein and Congdon (2012) and Allcott and Taubinsky (forthcoming) result for the case of no income inequality, we show more generally that it is not only the average (marginal) level of misoptimization that matters, but whether this misoptimization is concentrated on the low-end or the high-end of the income distribution. We show that the estimable degree of “bias concentration” consists of the covariance between income and bias, as well as the covariance between income and the price elasticity of demand. We then show that the extent to which redistributive motives oppose or amplify corrective motives is determined by the price elasticity of demand. The lower the elasticity, the more redistributive motives oppose corrective motives; the higher the elasticity the more redistributive motives amplify corrective motives.

In addition to deriving the optimal tax formulas, in Section 2 we also derive comparative statics results about how the sign and magnitude of the optimal tax depends on the size of the bias and the price elasticity of demand. These results are novel not only because they treat issues related to psychological biases, but also because such comparative statics results are often difficult to prove and thus rare in models that allow for a nonlinear income tax and general ability distributions. In addition to providing sharp qualitative guidelines for tax policy, these results also help relate our work to the fundamental insights first suggested by Gruber and Kőszegi (2004) about how the incidence of cigarette taxes on low-income earners is shaped by their price elasticity of demand.

We end Section 2 with two extensions. In Section 2.7 we elucidate the key conceptual differences between our analysis of internalities and the recent work on externalities in the Mirrlees model (Jacobs and de Mooij, forthcoming). In Section 2.8 we consider a more general version of our model in which exogenous wealth shocks can have meaningful impacts on consumption of the sin good. We show that in this case, the optimal commodity tax formulas we derive in the simpler setting carry forward, except with the addition of a new term corresponding to the fiscal externality from labor supply distortions caused by the commodity tax.

In Section 3 we apply this framework to a classic “sin good”: cigarettes. We use data on smoking rates and elasticities by income to calibrate a model in which there are internalities from the health costs associated with smoking, and in individuals with lower income earning abilities prefer to smoke at higher rates than high earners. We compute the optimal cigarette

---

2 While focusing on internalities rather than externalities, our work complements Jacobs and de Mooij (forthcoming) by extending implications to preference heterogeneity, which appears to be an important factor in explaining cigarette consumption, for example, as discussed in 3.
tax for a range of degrees of behavioral bias. Consistent with the analytic results discussed in Section 2, we show that, relative to the optimal tax absent redistributive concerns, the optimal cigarette tax can be larger, smaller, or indeed negative (a cigarette subsidy) depending on the degree of bias and the elasticity of demand for cigarettes. For the most plausible parameter values, the optimal tax remains positive but substantially reduced.

Because our framework dispenses with the assumption that consumers optimize perfectly and respond only to prices, a second set of contributions in our paper is to consider non-standard policy instruments and non-standard responses to (seemingly) standard policy instruments. In Section 4.1 we consider the question of when the commodity tax should take the form of a salient commodity tax included in the price or a less salient commodity tax that a consumer only sees at the register. We show that the answer to this question depends on the covariance between earnings ability and attentiveness to the less salient tax. Restricting to the case of a regressive tax on sin goods, we first show that when all consumers are equally attentive to the less salient tax, the policymaker will never choose a less salient tax that is positive – he would instead prefer to use the more salient tax. And conversely, the policymaker should never use a salient subsidy for redistributive purposes – the less salient subsidy would be more efficient. We show, however, that the common intuition that corrective instruments should be maximally salient may be reversed when low-income earners are more attentive to the sales tax than high income earners. In a simplified version of our model, we show that for any level of bias, it can be optimal to tax with a less salient instrument if wealth inequality is sufficiently high and if high income consumers are sufficiently inattentive relative to low income earners.

In Section 4.2 we move on to study the efficiency persuasion tactics – such as graphic images depicting the adverse consequences of smoking – and we show that these tactics are most efficient precisely when standard commodity taxes are most regressive. Generally, we show that in the presence of redistributive concerns, being able to use the persuasion campaign lowers the optimal corrective sin tax. In fact, we show that when persuasion is sufficiently powerful and its total social cost is sufficiently low, the optimal policy mix consists of a high level of persuasion and a “sin subsidy”.

We conclude in Section 5 with a discussion of the behavioral statistics highlighted by our analysis. We argue that further empirical work on individual differences and correlated heterogeneity in behavioral biases is essential for providing robust guidelines for public policy.
2 Adding Redistributive Concerns to Models of Internal-ity Taxes

2.1 Model set up

We begin with a simplified framework that adds redistributive concerns to the types of models considered in Gruber and Kőszegi (2004, henceforth GK), O’Donoghue and Rabin (2006, henceforth OR), Mullainathan, Schwartzstein and Congdon (2012, henceforth MSC), Allcott, Mullainathan and Taubinsky (2014, henceforth AMT), Allcott and Taubinsky (forthcoming, henceforth AT). The GK, OR, MSC, AMT, and AT models consider a set up with quasilinear utility, and thus no effects of income on consumption of the good in question. Here, we will similarly begin with a situation in which changes in wealth do not change consumption of the sin good, and we show how adding redistributive concerns changes the basic conclusions. A key feature of the model considered in this section is that we are able to add redistributive concerns while keeping consumer behavior constant. The model here also complements the general treatment of the Ramsey model of taxation concurrently considered in Farhi and Gabaix (2015).

We consider individuals with one-dimensional types indexed by \( \theta \in \mathbb{R} \), distributed according to a cumulative distribution function \( F \). In period 0, individuals choose earnings \( z \), while in period 1 they choose a consumption bundle \((c_1, c_2)\) subject to the budget constraint \( c_1 + (p + t)c_2 \leq y \), where \( p \) is the relative price of \( c_2 \), \( t \) is the commodity tax on \( c_2 \), and \( y \) is after-tax income. In period 0, consumers choose according to a utility function \( \hat{U}(c_1, c_2, z, \theta) = c_1 + \hat{u}(c_2, \theta) - \hat{\psi}(z, \theta) \), while in period 1 consumers choose consumption according to a utility function \( U(c_1, c_2, \theta) = c_1 + u(c_2, \theta) \), with \( u \) not necessarily equal to \( \hat{u} \). For concreteness, we assume that the consumer is sophisticated in the sense that in period 0 he correctly anticipates his period 1 behavior. However, our results would not change if we allowed for full or partial naivete.

The policy maker disagrees with the individual about the benefits of \( c_2 \), and believes the consumer should instead seek to maximize \( V(c_1, c_2, z, \theta) = c_1 + v(c_2, \theta) - \psi(z, \theta) \). The policy maker sets a nonlinear income tax \( T(z) \) and a linear commodity tax \( t \) to maximize total welfare given by \( \int G(V(c_1, c_2, z, \theta)) \) subject to the revenue constraint \( \int (T(z(\theta)) + tc_2(\theta))dF \leq 0 \), where \( c_2(\theta) \) and \( z(\theta) \) denote a type \( \theta \)'s choice of \( c_2 \) and \( z \) in the presence of the taxes \( T(z) \) and \( t \). The important difference from the standard optimal tax set-up is that the planner maximizes a weighted average of utility functions \( V \) that may not correspond to the utility

---

3See also Baicker et al. (2015) for an analysis of optimal, government-mandated health insurance when consumers may misoptimize.
functions $U$ (or $\hat{U}$). Here, $V$ denotes the utility function according to which consumers would choose if they were not subject to various psychological biases.

The difference between $U$ (or $\hat{U}$) and $V$ captures a variety of different possible psychological biases. We now provide some examples:

1. **Present bias.** As in O’Donoghue and Rabin (2006), suppose that consumers underweight the future health costs of $c_2$ (e.g., potato chips) because of present bias. Suppose that true utility from $c_2$ is given by $u(c_2) = m(c_2) - h(c_2)$, where $h$ is the health cost realized in the future. Our framework can be seen as a reduced-form representation of a (slightly) more dynamic model in which consumers choose $z$ in period 0, choose $c_1$ and $c_2$ in period 1, and experience the health cost in period 3. A consumer with present-biased, $\beta - \delta$, preferences (Laibson, 1997) will have a utility function $\hat{U} = \beta \delta (c_1 + m(c_2)) - \beta \delta^2 h(c_2) - \psi(z, \theta)$ in period 0 and utility function $U = c_1 + m(c_2) - \beta \delta h(c_2)$ in period 1. The planner’s utility function is $V = \delta (c_1 + m(c_2)) - \delta^2 h(c_2) - \psi(z, \theta)$.

2. **Limited attention** to certain attributes of $c_2$, as in Gabaix and Laibson (2006); DellaVigna (2009); Gabaix (2014), and as documented by Allcott and Taubinsky (forthcoming) for energy-efficient appliances. Suppose that $c_2$ represents investment in energy-efficiency. Purchasing the more energy-efficient product generates immediate utility $m(c_2)$, but also generates energy-cost savings equal to $K(c_2)$. Inattentive consumers, however, underweight these future costs by some (potentially endogenous) $1 - \alpha$. Then $\hat{U} = c_1 + m(c_2) + \alpha K(c_2) - \psi(z, \theta)$ and $U = c_1 + m(c_2) + K(c_2)$.

3. **Incorrect beliefs**, as documented for food choices and energy-efficiency choices by Allcott (2013); Attari et al. (2010); Bollinger et al. (2011). Consumers may simply have incorrect beliefs about certain attributes of $c_2$, such as its calorie content or its energy efficiency. In the setup in example 1, consumers may believe that the health costs are $\tilde{h}(c_2)$ rather than $c_2$. In the setup from example 2, consumers may believe that the energy efficiency is $\tilde{K}(c_2)$ rather than $c_2$.

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.

In Section 2.8 we will consider the fully general case in which income levels can affect consumption of $c_2$ – we defer this more general model until later because the presence of income effects generates consumer behavior that is different from the behavior in the standard GK, OR, MSC, AMT, and AT models.
We assume that consumers’ preferences are such that the choice of \( c_2 \) is interior and thus characterized by the first order condition \( u_1(c_2, \theta) = p + t \). We also impose the mild technical condition that \( \psi_1(0, \theta) > 0 \). We assume that the optimal income tax gives rise to a one-to-one mapping between \( \theta \) and \( z \), and thus with some abuse of notation, we let \( c_2(z, t) \) denote the consumption bundle chosen by a type \( \theta \) individual who has after-tax income \( z - T(z) \) and faces a commodity tax \( t \). We let \( H(z) \) denote the distribution of incomes at the optimum, and we use \( C_2(t) = \int c_2(z, t) dH \) to denote aggregate consumption of \( c_2 \).

For a type \( \theta \) consumer, set \( \gamma(c_2, \theta) = u_1(c_2, \theta) - v_1(c_2, \theta) \) to denote the bias in valuing the marginal utility from \( c_2 \) in period 1. Here too we abuse notation to let \( \gamma(z, t) \) denote the bias of a consumer earning income \( z \) and consuming \( c_2(z, t) \). Throughout the analysis in this section, we will follow AMT and AT in defining the average marginal bias:

\[
\bar{\gamma}(t) = \frac{\int dc_2 dt \gamma(z, t)dH(z)}{\int dc_2 dt dH(z)}
\]

Intuitively, if \( dC_2 \) is the marginal change in total consumption of \( c_2 \) when \( t \) is perturbed, \( \bar{\gamma} \) is the average amount by which consumers over- or under-estimate the change in utility from that change in consumption. When no confusion results, we may eliminate the dependence on \( t \) for simplicity.

### 2.2 The case of no redistributive concerns

When \( G \) is linear, so that there are no redistributive concerns, the optimal tax system is straightforward:

**Proposition 1.** Suppose that \( G \) is linear. Then the optimal tax schedule \((T^*(z), t^*)\) must satisfy \( t^* = \bar{\gamma}(t^*) \).

Proposition 1 replicates known “sin tax” results from GK, OR, MSC, AMT, AT. The proposition shows that when consumers overconsume some good, an optimal tax on that good should be positive and set at the level of the average amount of marginal misperception. Conversely, when \( \gamma < 0 \), so that consumers underconsume the good (as for the case of energy efficiency upgrades in AMT and AT), then the optimal policy is to subsidize the good at the average level of marginal misperception. Following MSC, AMT and AT, the proposition generalizes GK by allowing heterogeneity and mistakes other than present bias, and generalize OR by allowing for mistakes other than present bias. The OR result that taxing sin goods is optimal even if some people optimize perfectly is an immediate corollary of our more general characterization of the optimal tax system: if \( \gamma(\theta) = 0 \) for some types but \( \gamma(\theta) > 0 \) for other types, then clearly \( \bar{\gamma} > 0 \) and thus a positive tax is always optimal.
Although Proposition 1 mostly replicates known results, there is one new insight contained in the proposition. Because we consider a two-period setting in which consumers first choose earnings $z$ and then choose their consumption of $c_1$ and $c_2$, our framework allows for the possibility that a time-inconsistent consumer might alter his choice of before-tax income $z$ because of the “mistakes” that he knows his period 1 self will make. This is in contrast to the static frameworks with fixed incomes considered in MSC, AMT, AT. Nevertheless, Proposition 1 shows that just like in the static frameworks in MSC, AMT and AT, the optimal commodity tax should still equal the marginal bias. Perhaps surprisingly, the period 0 preferences of the consumer do not alter the formula for the optimal commodity tax $t$. The intuition behind this result will be made clear in the derivations in the next subsection.

2.3 Adding redistributive concerns

We now move on to the case in which $G$ is strictly concave, so that there are redistributive concerns. Consider a perturbation that increases the commodity tax $t$ by $dt$. This reform has both mechanical effects and behavioral effects due to consumers’ response to the tax change. The reform mechanically raises additional revenue equal to $C_2dt$. Letting $g(z) = G'(V(z - T(z) - (p + t)c_2(z), c_2(z), z)/\lambda$ denote the social marginal welfare weight corresponding to a type earning income $z$, normalized by the marginal value of public funds $\lambda$, the corresponding mechanical welfare loss taxpayers is $-dt \int g(z)c_2(z,t)dH$. It also causes a loss in revenue as consumers substitute away from $c_2$ to $c_1$. Letting $dc_2(z,t)$ denote the change in $c_2$ purchased by a $z$-earner in response to the tax, the reduction in tax revenue from this behavioral response is equal to $tdC_2$. In neoclassical models, the substitution has no first-order effect on welfare due to the envelope theorem. Here, however, the planner disagrees with the consumer’s choice of $c_2$, and thus the planner’s marginal rate of substitution differs from the consumer’s by the factor $\gamma$. Thus the change in $c_2(z,t)$ generates a first-order welfare change equal to $-dt \int \gamma(z)g(z)\frac{dc_2(z,t)}{dt}dH$. Because this perturbation has no impact on the consumer’s optimal choice of $z$ (a consequence of our quasilinearity assumptions), the total welfare impact, normalized by the cost of public funds, is

$$\frac{dW}{dt} = \int \left(-g(z)c_2(z,t) - \gamma(z)g(z)\frac{dc_2(z,t)}{dt} + t \frac{dc_2(z,t)}{dt} + c_2(z,t)\right) dH(z)$$

(1)

Now let $\zeta(z,t) = \frac{dc_2(z,t)}{dp + t} \frac{p + t}{c_2(z,t)}$ denote the elasticity of demand for $c_2$ and let $\bar{\zeta}(t) = \frac{dC_2(t)}{dp + t} \frac{p + t}{C_2(t)}$ denote the aggregate elasticity. (Again, we may omit the dependence on $t$.) Let $\omega(z,t) =$

---

4Of course, because the choice of $z$ is still determined by $\hat{U}$, the nature of $\hat{U}$ might alter the optimal income tax, which then alters the social marginal welfare weights, and thus indirectly affects the optimal commodity $t$. 

9
\(\gamma(z)\zeta(z)\) denote how biased and how elastic a person is relative to the average. Under the optimal policy, equation (1) is equal to zero, that the optimal tax satisfies the following first-order condition:

\[
\begin{align*}
t &= \frac{\int \gamma(z)g(z)\frac{dc_2(z,t)}{dt}dH}{\int \frac{dc_2(z,t)}{dt}dH} + \frac{\int c_2(z,t)(g(z) - 1)dH}{\int \frac{dc_2(z,t)}{dt}dH} \\
&= \frac{\int \gamma(z)g(z)\zeta(z,t)\frac{\gamma_2(z,t)}{p+t}dH}{\int \frac{dc_2(z,t)}{dt}dH} + \frac{\int c_2(z,t)(g(z) - 1)dH}{\int \frac{dc_2(z,t)}{dt}dH} \\
&= \bar{\gamma} + \frac{Cov_H[g(z), \gamma(z)\zeta(z,t)c_2(z,t)]}{\zeta(t)C_2(t)} + \frac{Cov_H[g(z), c_2(z,t)]}{\zeta(t)C_2(t)} \\
&= \bar{\gamma} \left[ 1 + \frac{Cov_H[g(z), \omega(z,t)c_2(z,t)]}{\sigma(t)C_2(t)} \right] + (p + t) \frac{\rho(t)C_2(t)}{\zeta(t)C_2(t)} \\
\end{align*}
\]

where we have used the fact that \(\int g(z)dH = 1\), which follows from our assumption that there are no income effects on labor supply. Set \(\sigma(t) := Cov_H[g(z), \omega(z,t)c_2(z,t)]\) to denote the extent to which the average marginal bias is concentrated on the low income earners. Set \(\rho(t) \equiv Cov_H[g(z), c_2(z,t)]\) to denote the regressivity of the tax. In terms of these two indices of concentration and regressivity, we find that the optimal tax \(t^*\) satisfies

\[
t^* = \bar{\gamma} \left[ 1 + \frac{\sigma(t^*)}{\sigma(t^*)C_2(t^*)} \right] + (p + t^*) \frac{\rho(t^*)}{\zeta(t^*)C_2(t^*)} \\
\]

And under the simplifying assumption that \(\gamma\) and \(\zeta\) are homogeneous, we have that \(\sigma(t) = \rho(t)\) and thus that \(t^*\) must satisfy

\[
t^* = \bar{\gamma}(t^*) \left[ 1 + \frac{\rho(t^*)}{\rho(t^*)C_2(t^*)} \right] - (p + t^*) \frac{\rho(t^*)}{\zeta(t^*)C_2(t)} \\
\]

Focusing first on the simple case in equation (4) shows that the size of the corrective tax depends on two forces. The first component is the internality \(\gamma\), but now scaled up or down by the extent to which consumption of \(c_2\) is concentrated amongst the low income earners or the high income earners. When consumption of \(c_2\) is concentrated amongst the low income earners, the planner is more concerned about the internality. This is reflected by \(\rho > 0\), which inflates the corrective benefits. Conversely, when consumption of \(c_2\) concentrated amongst the high income earners the planner is less concerned about the internality, which
is reflected by $\rho < 0$.

The second component of (4) is the cost of imposing a regressive tax. When the low income earners prefer $c_2$, the tax is regressive and thus creates a redistributive cost, which again is reflected in $\rho > 0$. This force lowers the optimal tax and pushes it in the direction of being a subsidy when $\gamma$ is close to zero. Conversely, when the high income earners prefer $c_2$, the tax is progressive: there is a redistributive gain from taxation of $c_2$. This force increases the optimal commodity tax and pushes it in the direction of being positive even when $\gamma$ may be negative.

More generally, equation (3) shows that the corrective benefits depend not only on the extent to which consumption of $c_2$ is concentrated amongst low income earners, but also on the extent to which low income earners are more or less elastic than high earners. If the low income earners are particularly elastic to the tax, that further increases the gains from corrective taxation. This is captured in the first covariance term, which shows that it is not only the relationship between $g(z)$ and $c_2$ that matters, but also the relationship between $g(z)$ and $\omega(z)$.

Although (3) suggests that the optimal commodity tax $t$ is a monotonic function of bias, sharp comparative statics are not immediate because all terms in the formula in (3) are endogenous to the tax $t$. The optimal tax $t^*$ is defined implicitly as the fixed point of the expression in (3), and thus care must be taken in characterizing how the tax varies with the primitives of the model. In the next two subsections, we formalize the intuitions suggested by equation (3) for “sin goods” that consumers overvalue (i.e. $\gamma > 0$) and “virtue goods” that consumers undervalue (i.e. $\gamma < 0$).

### 2.4 Comparative Statics for “Sin Goods”

Let $t^{NR}$ denote the optimal tax when there are no redistributive motives, and let $t^G$ denote the tax with redistributive motives, captured by the function $G$. We now characterize how $t^G$ compares to $t^{NR}$ as a function of bias for the case in which bias is positive, i.e., $\gamma > 0$, and a positive commodity tax is regressive, i.e., $\rho > 0$.

**Proposition 2.** Suppose that $G$ is strictly concave, that $\gamma(c_2, \theta) > 0$ for all $c_2, \theta$, and that $c_2$ is decreasing in $z$. Holding constant $U$ and $\omega$, the following hold for an optimal tax policy $(t^G, T^G)$:

1. Suppose that $\zeta(z,t)$ is bounded from above and bounded away from zero. Then there exists $\gamma^* > 0$ such that $t^G < 0$ if $\gamma(c_2, \theta) < \gamma^*$ for all $c_2, \theta$.

2. Suppose $\{\zeta(z,t)\}_{z,t}$ and $\{\omega(z,t)\}_{z,t}$ are bounded away from zero. Then there exists a $\gamma^*$ such that $t^G > 0$ if $\gamma(\theta) > \gamma^*$ for all $\theta$. 

11
3. Suppose that the conditions of (2) hold and that, moreover, $\omega(z,t)$ are non-decreasing in $z$ for each $t$. Then for each $t'$ there exist $\gamma^\dagger > 0$ such that $t^G > t'$ if $\gamma(\theta) > \gamma^\dagger$ for all $\theta$.

4. Suppose that i) $\omega(z,t)$ is non-decreasing in $z$ for each $t$ ii) $\gamma(c_2, \theta) \equiv \gamma^\dagger \in \mathbb{R}^+$ iii) for each possible value of $\gamma^\dagger$, welfare is quasiconcave in $t$ when the income tax is given by $T^G - (t - t^G)C_2$, iv) $u_1$ is bounded. Then $t^G > t^{NR}$ for high enough $\gamma^\dagger$.

5. In contrast, if $\sigma(t) < 0$ for all $t$ then it always holds that $t^G < t^{NR}$.

Part 1 of the Proposition says that under an innocuous regularity condition bounding elasticity ratios, the optimal sin tax may be negative even when bias is positive. The intuition for this result can be obtained from (3) by setting $\bar{\gamma} = 0$. In verifying this result formally, the proof in the appendix shows that it is necessary to establish that $\bar{\gamma}/\rho \to 0$ as $\bar{\gamma} \to 0$ at the optimal tax policy.

Part 2 shows that under another set of mild regularity conditions, the optimal sin tax is positive for sufficiently high bias. And part 3 shows under stronger conditions that ensure $\sigma \geq 0$, the optimal tax can be arbitrarily high for sufficiently high values of bias. One theoretical nuance that complicates these comparative statics is that the optimal commodity tax need not necessarily be a monotonic function of $\bar{\gamma}$. Because high levels of bias may also increase redistributive concerns, and therefore $\rho$, higher $\bar{\gamma}$ does not universally imply high $t^G$.

The role of the regularity conditions in parts 2 and 3 is to ensure that for high commodity taxes and high bias, consumers do not become so inelastic to the commodity tax that the regressivity costs outweigh the bias correction benefits.

Part 4 shows that when $u_1$ is bounded – so that consumption of $c_2$ is zero for a high enough $t$ – redistributive concerns can make the optimal corrective tax higher than the Pigouvian benchmark. To obtain some intuition for this, note that when both $\bar{\gamma}$ and $t$ are high, it can simply be better to ban the product completely, so as to reduce the financial burden on the low-income earners.

Finally, part 5, which follows immediately from (3), simply says that when bias is concentrated on the high income earners, redistributive concerns always decrease the size of the optimal commodity tax.

### 2.5 Comparative statics for “Virtue Goods”

We now consider implications for “virtue goods” that consumers may not consume enough of. This may include healthy foods or investments in energy-efficient durables and appliances (AT, AMT). One potentially important distinction between this case and the case of sin
taxes is the relationship between $\sigma$ and $\rho$. With regressive sin goods that are more likely to be consumed by low-income earners, both $\sigma$ and $\rho$ are likely to be positive. Thus even though the tax is regressive, the bias is also more concentrated on the low income earners, which makes the policymaker put more weight on the internality. In contrast, with virtue goods that are more likely to be consumed by the high income earners, $\rho$ will be negative, but $\sigma$ may be negative as well because the higher income consumers also consume more of $c_2$. At the same time, the reason higher income consumer consume more of $c_2$ may be because they are less biased, which would push $\sigma$ back toward being positive. Overall, the sign of $\sigma$ is unclear for the case of virtue subsidies.

A simple condition that we will impose for some of the results is that there is a Laffer curve for taxing $c_2$. Letting $C_2(t^M)$ denote total consumption of $c_2$ as a function of $t$, we define:

**CONDITION A (LAFFER CURVE):** There exists a finite $t$ that maximizes $tC_2(t)$.

When Condition A holds, we will let $t^M$ denote the revenue-maximizing tax rate.

**Proposition 3.** Suppose that $G$ is strictly concave, that $\gamma(c_2, \theta) < 0$ for all $c_2, \theta$, and that $c_2$ is increasing in $z$. Holding constant $U$ as well as the concentration weights $\omega$, the following hold for an optimal tax policy $(t^G, T^G)$:

1. There exists $\gamma^\dagger < 0$ such that $t^G > 0$ if $\gamma(\theta) > \gamma^\dagger$ for all $\theta$.

2. Suppose $\{ |\zeta(z, t)| \}_{z, t}$ is bounded from below and from above, that condition A holds, and that $c_2(t^M, \theta) > 0$. Then for each $t'$ there exists a $\gamma^\dagger$ such that $t^G < 0$ if $\gamma(\theta) < \gamma^\dagger$ for all $\theta$.

3. In contrast, if $\sigma(t) < 0$ for all $t$ then it always holds that $t^G > t^{NR}$.

Proposition 3 is analogous to Proposition 2. Part 1 shows that when bias is concentrated on high earners, it may be optimal to tax a product that consumers under-consume, for sufficiently small bias. Part 2 shows that under some regularity conditions, it is optimal to have a positive subsidy, for sufficiently large bias. Finally, part 3 shows that when bias is concentrated on the high-income earners, the optimal subsidy with redistributive concerns is always smaller in magnitude than the Pigouvian benchmark without redistributive concerns.

### 2.6 The key role of the price elasticity of demand

We now focus on the role that the price elasticity of demand plays in the magnitude of the optimal commodity tax. Solving for $t^*$ in (2), it follows that
\[ t^G = \frac{\gamma \left[ 1 + \frac{\text{Cov}_H[g(z), \omega(z)c_2(z)]}{C_2(t)} \right] - \rho \frac{\text{Cov}_H[g(z), c_2(z)]}{|C_2(t)|}}{1 + \frac{\text{Cov}_H[g(z), c_2(z)]}{|C_2(t)|}} \]

Corrective Benefits

\[ = \frac{\gamma \left| \bar{\zeta} |C_2(t) + \sigma C_2(t) + \rho \right|}{|C_2(t) + \rho|} \]

Regressivity Costs

\[ = \frac{\text{Cov}_H[g(z), c_2(z)]}{C_2(t) + \text{Cov}_H[g(z), c_2(z)]} \]

(5)

When \( \zeta \) and \( \gamma \) do not vary with type, we have that

\[ t^G = \gamma |\bar{\zeta}| \frac{C_2(t) + \rho}{|C_2(t) + \rho|} - \frac{\rho \sigma}{|C_2(t) + \rho|} \]

(6)

The key insight in equation (5) is that while the corrective benefit term is proportional to the elasticity, the regressivity costs term is not. This suggests that the importance of correcting consumer bias, relative to regressivity costs of the commodity tax, is proportional to the price elasticity of demand. Intuitively, if consumers are not at all elastic to a regressive tax, then the tax only redistributes money from the poor to the rich, without correcting consumption of \( c_2 \). Thus in the extreme case in which \( |\bar{\zeta}(t)| \approx 0 \), it is clear that commodity tax should be used almost exclusively as a tool for redistribution, rather than as a tool for correcting consumption of \( c_2 \). More generally, formula (5) clarifies that outside of this extreme case, the role of consumer bias in shaping the optimal commodity tax is extremely sensitive to the elasticity.

We now formalize the intuitions obtained from (5). To formally discuss the role of the elasticity, we restrict to a family of utility functions \( U \) that keep certain statistics – such as baseline consumption when \( t = 0 \) – constant, and then discuss the implications of selecting utility functions with higher or lower elasticities. The function \( \hat{U} \) will be held constant for all comparative statics in this section.

**Proposition 4** (Implications for sin taxes). Fix a bias function \( \gamma \) with \( \gamma(\theta, c_2) > 0 \) for all \( \theta, c_2 \). Let \( \mathbb{U} \) denote a space of utility functions \( U \) such that the following are constant for all \( U \in \mathbb{U} \): i) \( c_2(\theta, 0) \), ii) \( u(c_2(\theta, 0), \theta) \), iii) the relative elasticity function \( e(\theta, t) = \zeta(\theta, t)/\bar{\zeta}(t) \).

Suppose also that \( c_2(\theta, t) \) is decreasing in \( \theta \) for all \( U \in \mathbb{U} \) and that \( e(\theta, t) \) is bounded. Then for each strictly concave \( G \) and a scalar \( t' > -\rho \) there exists a \( k > 0 \) such that \( t^G < t' \) for any \( U \in \mathbb{U} \) satisfying \( |\bar{\zeta}(t)| < k \) for all \( z, t \).
Proposition 4 shows that for a family of utility functions $U$ characterized by (i)-(iii), the optimal commodity tax can be made arbitrarily small for utility functions with sufficiently small elasticities. Although we have not been able to prove a converse for high elasticities, our simulations similarly suggest that the optimal corrective tax is increasing in the magnitude of the elasticity, and surpasses the Pigouvian benchmark for sufficiently high elasticities.

For regressive “virtue subsidies”, we prove a similar analogue of Proposition 4. The proposition similarly establishes that for a given family of utility functions $U$, it may be optimal to tax, rather than subsidize the under-consumed good if the elasticity is sufficiently close to zero.

**Proposition 5 (Implications for virtue subsidies).** Fix a bias function $\gamma$ with $\gamma(\theta, c) < 0$ for all $\theta, c$. Let $\mathbb{U}$ denote a space of utility functions $U$ such that the following are constant for all $U \in \mathbb{U}$: i) $c_2(\theta, 0)$, ii) $u(c_2(\theta, 0), \theta)$, iii) the relative elasticity function $e(\theta, t) = \zeta(\theta, t)/\bar{\zeta}(t)$. Suppose also that $c_2(\theta, t)$ is increasing in $\theta$ for all $U \in \mathbb{U}$ and that $e(\theta, t)$ is bounded. Then for each strictly concave $G$ there exists a $k > 0$ such that $t_G > 0$ for any $U \in \mathbb{U}$ satisfying $|\bar{\zeta}(t)| < k$ for all $z, t$.

2.7 Comparison to externality taxation

To compare our results to externality taxes, suppose instead that consumers maximize $U$, but that their total utility is given by $V = U - X(C_2)$, where $X(C_2)$ is the externality from total consumption of $C_2$. Then analogous to equation (1), we have that

$$dW = \int \left( -g(z)c_2(z, t) - g(z)X'(C_2) \frac{dC_2}{dt} + t \frac{dC_2}{dt} + C_2 \right) dH(z)$$

from which it follows that

$$t^* = \frac{X'(C_2)}{\text{Corrective Benefits}} + (p + t^*) \frac{\rho(t^*)}{\zeta(t^*) C(t^*)} \frac{1}{\text{Regressivity Costs}}$$

Note that contrary to (3), the first component of $t^*$ is simply the marginal damages from the externality-producing good. Whereas (3) shows that the weight given to the internality depends on whether consumption of $c_2$ is concentrated on the low-income or high-income portion of the population, (9) shows that the covariance between income and consumption of $c_2$ does not affect the weight given to the externality. Intuitively, this is because both low-income and high-income consumers have the same marginal impact on the externality damages that affect everyone in the population. The logic is different with internalities, because the internality generated by a low-income consumer impacts only that low-income
consumer. And because internalities have a higher impact on the welfare of low-income consumers than on the welfare of high-income consumers, it matters whether it is the low-income consumers or the high-income consumer who are the most biased.

The concepts generalize when externality production is heterogeneous and has heterogeneous welfare effects. Suppose that a type $\theta$ individual contributes $x(c_2, \theta)$ to the externality, so that $X(C_2) = \int x(c_2, \theta) dF$, and derives disutility $y(\theta)X(C_2)$ from consumption of $c_2$, for some some scalar $y(\theta)$. Then generalizing the Diamond (1973) formula, we have that

$$t^* = \bar{x} (1 + \text{Cov}[g(\theta), y(\theta)]) + (p + t^*) \frac{\rho(t^*)}{\zeta(t^*)C_2(t^*)}$$

(10)

where $\bar{x} = \int x(c_2, \theta) \frac{dc_2}{dt} dF$ is the average marginal contribution to the externality. The average marginal externality $\bar{x}$ is analogous to the average marginal bias $\bar{\gamma}$ in equation (3). Equation (10) appears more similar to (3), but with an important difference. When sensitivity to externalities is allowed to be heterogeneous, then as with internalities, it matters whether those most sensitive affected are the poor or the rich. However, the difference remains that with internalities – but not with externalities – the optimal commodity tax depends on how income covaries with both consumption of $c_2$ and the price elasticity of demand for $c_2$. Intuitively, for internalities it matters whether low-income earners are also those who are the most responsive to the the internality tax, since their utility is affected only by their own behavior. With externalities, it does not matter whether low income earners are most responsive to the tax, since their utility is affected by aggregate behavior.

The analysis here clarifies that in the standard case in which $y(\theta)$ is homogeneous, redistributive concerns always make the optimal externality-corrective taxes lower than the Pigouvian benchmark when the commodity tax is regressive. And more generally, greater redistributive concerns cannot make the policymaker more concerned about the externality – only more concerned about the regressivity of the commodity tax. In contrast, even when taxes are regressive, redistributive concerns can make a policymaker more concerned with the internality, and might even lead to a tax that is higher in magnitude than the average marginal bias.

2.8 The general case with income effects and non-separable preferences

In this section, we consider a more general setting in which an individual’s preference are not necessarily weakly separable in $c_1$, $c_2$ and $z$. We allow general functions $\hat{U}(c_1, c_2, z, \theta), U(c_1, c_2, z, \theta), V(c_1, c_2, z, \theta)$,
Our goals here are twofold: First, we examine the robustness of the conclusions from equation (5) about the interaction of corrective and redistributive motives. Second, although we do not derive closed-form solutions, we still implicitly characterize the optimal tax system in terms of measurable elasticities and a summary statistic of consumer bias. These formulas help clarify what key empirical parameters are needed to estimate the optimal policy.

For simplicity, we assume that $\hat{U}_1/\hat{U}_3 = U_1/U_3 = V_1/V_3$, meaning that consumers correctly tradeoff labor effort and consumption of $c_1$. This allows us to focus our analysis on incorrect tradeoffs between $c_1$ and $c_2$, without the additional complication that consumers may make suboptimal labor supply choices because of a general misoptimization with respect to labor and consumption tradeoffs.

Let $\xi(z, \theta, t)$ denote a type $\theta$’s earned income elasticity with respect to the net of tax rate $1 - T'$. We follow Jacquet et al. (2013) in defining this elasticity locally at the status quo tax regime. Let $\xi^c(z)$ denote the compensated elasticity, and let $\eta = \xi - \xi^c$ denote the income effect.

Analogously, let $\zeta^i(z, \theta, t)$ denote the elasticity of $c_i$ with respect to net of tax price. Here $\zeta^i(z)$ is the total elasticity: it captures how much a person currently at income $z$ will reduce his consumption of $c_2$, taking into account the fact that some of this reduction may be due to an income effect that results from this person lowering his income by some amount $dz$. Let $\zeta^{i,c}$ denote the compensated elasticity, and let $\varepsilon^i = \zeta^i - \zeta^{i,c}$ denote the income effect. The income effect is given simply by $\varepsilon^i(z) = \frac{dc^i}{dz}(p + t)$. Finally, let $\chi$ denote a type $\theta$’s earned income elasticity with respect to the net of tax price $p + t$ of $c_2$.

Consider now increasing the commodity tax by a small amount $dt$. First, this has a mechanical revenue effect given by $M = \int c_2(\theta, t) dF$. Second, this causes individuals to lower their income, creating a fiscal externality of size $I = \int \chi(\theta) \frac{z(\theta)}{p + t} T'(z(\theta)) dF$. Third, the tax decreases consumption of $c_2$. The revenue loss from this is given by $R = \int t(\theta) \frac{c^i(\theta,t)}{p + t} dF$.

Finally, set $\gamma(\theta) = U_2^\theta / U_1^\theta - V_2^\theta / V_1^\theta$ to be the difference in perceived versus experienced marginal rates of substitution between $c_2$ and $c_1$, where $c_1$ and $c_2$ are chosen to maximize $U$ at income $z(\theta)$ and tax $T(z(\theta))$. Set $g(\theta) = V_1^\theta (c_1, c_2, z)/\lambda$ where $c_1$ and $c_2$ are chosen to maximize the consumer’s decision utility $U$ at income $z(\theta)$ and tax schedule $T$.\footnote{Note that this definition is slightly different from conventional marginal social welfare weights, as it is not identical to the social utility of giving a marginal dollar to type $\theta$, however it serves as a useful summary of redistributive concerns in this context.} To abstract from issues related to consumers over- or under-working, and instead to focus solely on the incorrect tradeoffs between consumption of $c_1$ and $c_2$, we assume that $V_1/V_3 = U_1/U_3$ for all types $\theta$.

Then because $U_2/U_1 = p + t$ at the consumer’s perceived optimal choice of consumption,
the welfare effect from substitution, and from the utility loss to consumers paying the tax (normalized by the cost of public funds) is given by

\[
S = \frac{1}{\lambda} \int \left[ - \left( (p + t) V_1^\theta + V_2^\theta \right) \zeta(\theta) \frac{c_2}{p + t} - c_2 V_1^\theta \right] dF
\]

\[
= \frac{1}{\lambda} \int V_1^\theta \left[ - \left( (p + t) + \frac{V_2^\theta}{V_1^\theta} \right) \zeta(\theta) \frac{c_2}{p + t} - c_2 \right] dF
\]

\[
= \int g(\theta) \left[ - \left( \frac{U_2^\theta}{U_1^\theta} + \frac{V_2^\theta}{V_1^\theta} \right) \zeta(\theta) \frac{c_2}{p + t} - c_2 \right]
\]

\[
= - \int g(\theta) c_2(z, \theta, t) \left[ \frac{\gamma(\theta) \zeta(\theta)}{p + t} + 1 \right] dF
\]

Now at the optimum, \( M + I + R + S = 0 \), from which it follows that

\[
t = \frac{\int g(\theta) c_2(\theta, t) \left[ \frac{\gamma(\theta) \zeta(\theta)}{p + t} + 1 \right] dF - \int c_2(\theta) dF - \int \chi(\theta) \frac{z}{p + t} T'(z) dF}{\int \frac{dc_2(\theta, t)}{dt} dF}
\]

\[
= \frac{\int g(\theta) c_2(\theta, t) \gamma(\theta) \zeta(\theta) dF}{\int \zeta(\theta) c_2(\theta) dF} - (p + t) \int c_2(\theta)(1 - g(\theta)) dF + \int z(\theta) \chi(\theta) T'(z) dF
\]

\[
= \bar{g} \bar{\gamma} + \frac{\text{Cov}[g(\theta), \gamma(\theta) c_2(\theta) \zeta(\theta)]}{\int |\zeta(\theta)| c_2(\theta) dF} - (p + t) \frac{\text{Cov}[c_2(\theta), g(\theta)]}{|\zeta| C_2(t)} + C_2(\bar{g} - 1) + \int z(\theta) \chi(\theta) T'(z(\theta)) dF
\]

\[
= \bar{g} \gamma(t) \left[ \bar{g} + \frac{\text{Cov}[g(\theta), \omega(\theta) c_2(\theta)]}{C_2} \right] - (p + t) \frac{\text{Cov}[c_2(\theta), g(\theta)]}{|\zeta| C_2(t)} + C_2(\bar{g} - 1) + \int z(\theta) \chi(\theta) T'(z(\theta)) dF
\]

where \( \gamma(t) = \frac{\int \gamma(\theta) \frac{dc_2(\theta, t)}{dt}}{\int \frac{dc_2(\theta, t)}{dt} dF} \) denotes the average marginal bias, \( \bar{g} = \int g(\theta) dF \) denotes the average population average of \( g(\theta) \), and \( \omega(\theta) = \frac{\gamma(\theta) |\zeta(\theta)|}{\bar{\gamma} |\zeta|} \) denotes the how biased and how elastic a consumer is relative to the population. As before, we set \( \sigma(t) = \text{Cov}[g(\theta, t), \omega(\theta, t) c_2(\theta, t)] \) to denote the extent to which the misperception corresponding to a marginal change in consumption is concentrated on the low income earners. And we set \( \rho = \text{Cov}[c_2(\theta, t), g(\theta)] \) to denote the regressivity of the tax. The new term \( F E(t) = \int z(\theta) \chi(\theta) T'(z(\theta)) dF \) is the fiscal externality from the change in labor supply caused by an increase in the tax \( t \). Thus in terms of these two indices, we again have that

\[
t^* = \bar{g} \left[ \gamma(t) + \frac{\sigma(t)}{C_2} \right] - (p + t) \left[ \frac{\rho(t) + F E + \bar{c}_2(\bar{g} - 1)}{|\zeta| C_2} \right]
\]

The key difference from our baseline formula is that now it is not only the regressivity term \( \rho \) that matters, but also the fiscal externality \( F E \) from a change in the commodity
tax \( t \). Intuitively, the fiscal externality results when exogenous differences in income change consumption of \( c_2 \). If a consumer were to reduce his consumption of \( c_2 \) from \( c'_2 \) to \( c''_2 \) when he decreases his earnings from \( z' \) to \( z'' \), then the total taxes he pays would decrease by \((c'_2 - c''_2)t\). Thus when consumption of \( c_2 \) is rising in income, increasing the commodity tax \( t \) causes consumers to lower their labor supply.

To explore the connection to standard results on externalities, suppose consumption of \( c_2 \) produces a harmful externality, and let \( X(C_2) \) denote the impact that the externality has on each consumer’s utility. Then as before,

\[
t^* = \frac{X'(C_2)}{\lambda} - (p + t) \left[ \frac{\rho(t) + FE + \bar{c}_2(\bar{g} - 1)}{|\zeta|C_2} \right]
\]

Again, a key difference between externality taxes and internality taxes is the bias concentration term \( \sigma \), which shows that concerns for correcting internalities are amplified when the low income earners consume more of \( c_2 \) and are the most elastic. The difference is particularly transparent in the extreme case in which types are one-dimensional and decision utility \( U \) is weakly separable, so that \( \rho + FE + \bar{c}_2(\bar{g} - 1) = 0 \)\. In this case, the optimal commodity tax is always set equal to the Pigovian benchmark \( X'(C_2) \). In contrast, the optimal internality tax can be higher or lower than the Pigovian benchmark, depending on whether low income earners’ consumption of \( c_2 \) is relatively more responsive to the commodity tax \( t \).

### 3 Numerical analysis

We now explore the quantitative implications of the above results in a the context of a common application of sin taxes: cigarette consumption. Our approach to this topic will be a simplified one—rather than modeling a full dynamic model of cigarette consumption with addiction and present bias, as in Gruber and Kőszegi (2001, 2004), we adopt the approach of Saez (2002a) and O’Donoghue and Rabin (2006), in which consumption of the sin good imposes a separable (perhaps delayed) cost which is not internalized by the consumer.

Although the time separable model of consumption utility in this paper is perhaps most appropriately applied to unhealthy, nonaddictive products like sugary drinks or “junk food”, we focus here on cigarettes for three reasons. First, there is extensive literature on smoking patterns and elasticities (e.g., and, crucially, the heterogeneity of those patterns across the

---

6This fact is a variation of the well-known Atkinson-Stiglitz Theorem. Note, however, it is not clear what is the “right” analog of the Atkinson-Stiglitz assumptions in our setting. Because there is a wedge between \( \hat{U}, U \) and \( V \), it is not clear which of these utility functions “should” be homogeneous. In general, because all existing work that estimates bias at the individual level finds substantial heterogeneity in biases, the assumption of homogeneous \( U \) and \( \hat{U} \) seems implausible in our setting.
income distribution. This literature is largely made possible by the extensive variation in cigarette taxes over time and across states – see, for example, Evans et al. (1999) and Adda and Cornaglia (2006) for aggregate elasticity estimates, and Gruber and Kőszegi (2004) and Goldin and Homonoff (2013) for estimates by income. One aim of this paper is to highlight the value of collecting such data for other sin goods, but rather than conjecture about such parameters for other goods, we prefer to draw evidence from a domain that has already received substantial empirical focus. Second, and related to the widespread variation in taxes, cigarettes are already widely recognized as a “sin good” with taxes that are driven by internalities. Indeed, the externalities from cigarette consumption are typically found to be much smaller than internalities, and much smaller than prevailing excise taxes. (See Chaloupka and Warner (2000) for an extensive review of this literature.)

Finally, there is evidence that smoking preference is correlated with income earning ability, independent of income itself. As noted in the introduction, and discussed extensively in Saez (2002a), commodity taxes are useful instruments for redistribution if the variation in consumption across the income distribution is due to differences in preferences across abilities, rather than to income effects – otherwise the classic result from Atkinson and Stiglitz (1976) implies that all redistribution can be achieved through the nonlinear income tax. Most generally, if the negative relationship between smoking and income were driven by income effects, and not by preferences, cigarettes would need to be an inferior good – yet as shown below, rich smokers do not consume fewer cigarettes than poor smokers. More concretely, see Weiser et al. (2010) for evidence on the correlation between IQ scores and cigarette takeup among adolescents, which remains negative even after controlling for parents’ socioeconomic status.

We use the data set from Goldin and Homonoff (2013), drawn from the Behavioral Risk Factor Surveillance System telephone survey in years 1984–2000, to calibrate smoking rates and elasticities across the income distribution. Figure 1 displays the share of smokers and the number of cigarettes smoked per day among smokers, both as a function of income. Data for the year 2000, which we use to calibrate our model, is plotted in bold. These patterns suggest that the primary dimension of heterogeneity in smoking across the income distribution is on the extensive margin (whether an individual smokes at all) rather than the number of cigarettes consumed. This is also the primary source of variation in cigarette consumption over time—whereas the share of smokers declined by 37% from 1984 to 2000, the number of cigarettes consumed among smokers fell by less than 10%. The variation in smoking rates across incomes is substantial, and is consistent with more recent survey data reported by Gallup, displayed in Figure 2.
Figure 1: Smoking patterns as a function of income across years. Data are from Goldin and Homonoff (2013), drawn from the Behavioral Risk Factor Surveillance System. These schedules are constructed using kernel regressions. The left panel regresses an indicator for whether an individual is a smoker (smokes at least one cigarette per day) on income percentile in each year. The right panel regresses the number of cigarettes smoked per day on income among smokers.

Figure 2: Smoking prevalence across income, according to Gallup.

To fit the patterns in Figure 1, we adopt a slightly modified version of the general setup presented in Section 2. Specifically, we assume two types of consumers—those who have a taste for cigarettes (“smokers”, denoted by an indicator variable \( s = 1 \)), and those who do not \( (s = 0) \), with quasilinear demand for cigarettes among smokers. We let \( c_1 \) denote all consumption other than cigarettes and \( c_2 \) denote cigarettes (or the sin good of interest), both measured in dollars. As in the standard Mirrlees (1971) model of redistributive income taxation, we further assume individuals have heterogeneous wages \( \theta \), and generate earnings.
by supplying labor effort $\ell = z/\theta$. Individual decision utility thus takes the following form:

$$ U(c_1, c_2, \ell, s) = \begin{cases} 
  c_1 - v(\ell) & \text{if } s = 0 \\
  c_1 + u(c_2) - v(\ell) & \text{if } s = 1.
\end{cases} \tag{11} $$

Experienced utility differs from decision utility only for smokers, who experience an additional (uninternalized) cost $\gamma c_2$. This cost is assumed to be linear so that the optimal sin tax absent redistributive considerations is simply $\gamma$. Thus experienced utility takes the form

$$ V(c_1, c_2, \ell, s) = \begin{cases} 
  c_1 - v(\ell) & \text{if } s = 0 \\
  c_1 + u(c_2) - \gamma c_2 - v(\ell) & \text{if } s = 1.
\end{cases} \tag{12} $$

In what follows, we use $v(\ell) = \frac{e^{1+1/e}}{1+1/e}$, where $e$ is the constant compensated elasticity of labor supply, as in the Type I utility function from Saez (2001), which we set equal to 0.33, the preferred value in Chetty (2012). This specification avoids complications from income effects on the choice of labor supply, which are not of central concern for the results which follow. The distribution of wages is assumed to be continuous, denoted $F(\theta)$ with density $f(\theta)$, and is calibrated to match the empirical income distribution.

We also use $u(c_2) = k \left( c_2^{1-1/\zeta} - 1 \right)$, where $\zeta$ denotes the absolute value of the constant price elasticity of demand for cigarettes, with $u(c_2) = k \log(c_2)$ when $\zeta = 1$. The constant $k$ controls the level of spending on cigarettes, which we set to generate $1200 in spending on cigarettes per year among smokers, equal to approximately one pack per day, in the presence of a 34% tax (the mean cigarette excise tax in the US in 2000), consistent with Figure 1.

The planner’s objective is to select a nonlinear income tax $T(z)$ and linear tax on cigarettes $t$, in order to maximize a weighted sum of experienced utility. Consistent with the literature on preference heterogeneity and redistribution, we assume that differences in preferences alone do not merit redistribution (see Fleurbaey and Maniquet (2006) and Lockwood and Weinzierl (2015)). Thus we write the marginal social welfare weights $g(\theta)$ as a decreasing function of ability $\theta$, but not of smoking taste $s$, with $\int g(\theta)dF(\theta) = 1$. For numerical simplicity, we here assume marginal social welfare weights are a function of type directly, rather than being determined endogenously by the concave transformation $G$ as

---

7 Specifically, we use the density of incomes as measured by the Current Population Survey for years 2003–2006, among single individuals and heads of households over 25 years old who earn at least 50% of household income. Due to the sparse coverage of CPS at the top of the income distribution, we follow Mankiw et al. (2009) and smooth paste a Pareto tail with parameter $\alpha = 2$ (consistent with Saez (2001)) above the 97th percentile. Following Saez (2001) and Saez (2002b), we back out the ability distribution so that the empirical income distribution would arise under a flat tax rate (equal to 30%). This avoids numerical complications from bunching at kinks in a piecewise linear tax function, which are not of central interest for our purposes.
in Section 2. As before, the planner’s objective is to maximize $g(\theta)V(c_1, c_2, z/\theta, s)$, where the choice variables $c_1$, $c_2$, and $z$ are chosen optimally by each individual taking $T$ and $t$ as given, subject to a resource constraint that total income and commodity taxes sum to zero. (Imposing an exogenous government expenditure requirement affects only the lump sum grant in this specification and can thus be ignored for our purposes.)

3.1 Baseline results

The optimal income tax function satisfies the standard first-order condition from Diamond (1998) and Saez (2001) (in the specification without income effects)—the schedule of marginal tax rates is plotted in Figure 3. To make our results comparable to other familiar results in the optimal tax literature, we set marginal social welfare weights $g(\theta)$ to equal $g(\theta) = c_1^{-\nu}$ among $s = 0$ consumers at the optimum, so that the parameter $\nu$ governs the redistributive taste of the planner, with $\nu = 1$ in our baseline specification. Under the quasilinear specification in (11) and (12), the schedule of optimal marginal income tax rates does not depend on the share of smokers, nor on how smoking varies with income.

![Figure 3](image)

Figure 3: The left panel displays the schedule of consumption (post-tax income) as a function of pre-tax earnings at the optimum (the dashed 45° line represents the case of no taxes for comparison). The right panel plots the schedule of optimal marginal income tax rates.

In this setup, $\sigma(t) = \rho(t) = C_2(t)(E[g|s = 1] - 1)$, and thus expression 3 takes a particularly simple form:

$$t = \gamma E[g|s = 1] - \frac{1 + t}{\zeta} (E[g|s = 1] - 1)$$

(13)

The optimal tax on $c_2$ is plotted as a function of the degree of bias $\gamma$ for a range of demand elasticities in Figure 4. The maximum value $\gamma = 1$ corresponds to the greatest degree of bias
considered by O’Donoghue and Rabin (2006), in which the present bias parameter $\beta = 0.9$ and the future health cost of consuming a sin good is equal to ten times its pre-tax price. By construction, the optimal tax absent redistributive concerns is simply equal to $\gamma$, represented by the dashed $45^\circ$ line through the origin. The line for $\zeta = 0.35$, corresponding to the preferred intensive margin elasticity estimate from Goldin and Homonoff (2013) (see Table 5 of that paper), is plotted in bold. Consistent with Proposition 2, part 1, the tax is negative (a subsidy) for sufficiently low levels of bias, and rises with bias. Conditional on a given degree of bias, the tax is lower when demand for $c_2$ is less elastic. When $\zeta = 0.05$, a subsidy is optimal for the entire range of bias considered. In contrast, for sufficiently high elasticity, and at high degrees of bias, the tax is higher than the non-redistributive benchmark. This illustrates 2, part 4: regressivity of sin taxes may raise the size of the optimal tax if elasticities are sufficiently high, although these simulations suggest such a result occurs only at elasticities significantly higher than those typically found in the literature on cigarette consumption.

Together, these results suggest that for modest to large degrees of bias, accounting for redistributive motives tends to reduce the optimal tax substantially, relative to the non-redistributive benchmark. For example, if smokers ignore about $1 in future health costs for every $1 worth of cigarettes consumed ($\gamma = 1$) then the optimal tax is 75%, rather than 100%, of the cost of a pack of cigarettes. Finally, the wide dispersion of the lines in Figure 4 emphasize the importance of accurately measuring the demand elasticity of sin goods – for example, a difference in elasticity of 0.35 versus 0.1 determines whether cigarettes should be subject to a substantial tax or a substantial subsidy for moderate degrees of bias.

Figure 4: Optimal linear tax on $c_2$ as a function of the ignored health cost $\gamma$, for a range of price elasticities of demand $\zeta$. 

24
Figure 5 displays the optimal commodity tax on $c_2$ as a function of $\gamma$ for a range of degrees of inequality aversion. The baseline case of $\nu = 1$ is plotted in bold—all are computed using the baseline elasticity of $\zeta = 0.35$. The schedule of income tax rates for each value of $\nu$ is plotted in the right panel. This figure highlights the key role of redistributive concerns in determining the optimal tax for goods which are preferred by the poor. Even with the relatively mild decline in smoking across the income distribution (Figure 1) a high aversion to inequality of $\nu = 4$ can reduce the optimal tax by between $0.25$ and $0.50$ for every $1$ of cigarette expenditures. On the other hand, if redistributive motives are mild ($\nu = 0.25$) the optimal tax is only slightly lower than the non-redistributive benchmark. Indeed, the quantitative implications of inequality aversion for the optimal commodity tax when $\gamma = 1$ are quantitatively similar to those for optimal marginal income tax rates – raising $\nu$ from 1 to 2 reduces the optimal cigarette tax by about 10 percentage points, and raises optimal marginal income tax rates by a comparable amount across the income distribution.

Finally, Figure 6 plots the gains in social welfare generated by implementing the optimal tax, and the naive “Pigouvian” tax, relative to not having any tax on $c_2$. The optimal commodity tax generates gains over the Pigouvian tax equal to about $1.10$ per person. Although this may not seem large, it can be quite substantial as a share of the potential gains from taxing cigarettes at all. As noted above, the average excise tax on cigarettes in 2000 was $0.34$. If this were calibrated based on estimates that average marginal bias was $\bar{\gamma} = 0.34$, with no account of redistributive concerns, then Figure 6 indicates the (substantially lower) optimal cigarette tax would generate $1.27$ per person in social welfare gains, rather than only $0.18$ – that is, the welfare gains from the prevailing tax could be raised seven-fold by accounting for inequality aversion. Importantly, at low levels of bias (perhaps less relevant for cigarettes than for junk foods or other mildly unhealthy goods consumed by the poor) the Pigouvian benchmark tax is actually harmful.
Figure 5: Optimal linear tax on $c_2$ as a function of the ignored health cost $\gamma$, for a range of redistributive tastes $\nu$, with $\zeta = 0.35$. Optimal marginal income tax rates in each case are shown in the right panel.

Figure 6: Welfare gains (measured in dollars of public funds) from implementing the optimal commodity tax on $c_2$, and from implementing the suboptimal “Pigouvian” corrective tax equal to $\gamma$, as a function of the bias $\gamma$.

3.2 Heterogeneous demand elasticity

In the above analysis we assumed a constant price elasticity of demand for cigarettes of 0.35, consistent with the intensive margin elasticity estimated in Goldin and Katz (2007), and does not vary measurably with income. However other sources have estimated demand elasticities which decline in magnitude with income—for example, Gruber and Kőszegi (2004) use the
Consumer Expenditure Survey to compute elasticities which fall (in absolute value) from 1.1 in the poorest quartile of consumers to 0.38 in the top quartile.\(^8\) Although our goal is not to adjudicate between alternative elasticity computations, in light of the derivations in Section 2, we are interested in the implications of elasticities which covary with marginal social welfare weights.

Therefore we re-run the specification from the preceding section with the modification that \(\zeta\) is set equal to 1.1 for the bottom quartile of earners, 0.70 for the second quartile, 0.53 for the third quartile, and 0.39 for the fourth.\(^9\) This is admittedly a somewhat ad-hoc method of incorporating elasticities which vary with incomes, as it assumes that variation is driven by variation in the utility function with \(\theta\). Yet it allows us to avoid the numerical complexities of income effects on \(c_2\), which complicate the optimal income tax and which, in any case, appear inconsistent with the fairly constant level of cigarette consumption among smokers of all incomes.

The left panel of Figure 7 displays the optimal commodity tax on \(c_2\) in this calibration, for a range of redistributive tastes (the optimal income tax schedules are as in Figure 5). The right panel displays the optimal tax \(t\) when the elasticity is constant at the average level from the left panel, \(\zeta = 0.66\). This calibration generates optimal commodity tax rates that rise more quickly with the degree of bias \(\gamma\), and we see levels of \(t\) that exceed the Pigouvian benchmark (the dashed line) for \(\gamma > 0.6\) even under our baseline redistributive tastes of \(\nu = 1\). This stark result is driven in part by positive covariance between welfare weights and the demand elasticity, as emphasized by the importance of \(\omega\) in 3. It is also driven by the fact that in this specification, a larger tax reduces the relationship between cigarette consumption and income – since low income consumers are more elasticity, a higher tax reduces their consumption by more, thereby reducing the regressivity cost of cigarette taxation through a reduction in \(\rho\). However \(\rho\) remains positive even at the highest level of tax shown in Figure 7, indicating that the higher-than-Pigouvian commodity taxes are driven by progressivity alone.\(^10\)

\(^8\)Gruber and Kőszegi (2004) report levels of cigarette expenditures (3.2% in the bottom quartile and 0.4% in the top) which are somewhat higher than those in the Goldin and Katz (2007) data, 2.5% and 0.2% respectively.

\(^9\)As before, we calibrate \(k\) so that annual cigarette expenditures is $1200 across the income distribution when \(t = 0.34\). This implies that when \(t \neq 0.34\), cigarette consumption varies with income quartile.

\(^10\)To be precise, \(\rho/C_2\), which is not sensitive to the units on \(C_2\), declines from 0.093 when \(\gamma = 0\) to 0.023 when \(\gamma = 1\) in the simulation with \(\nu = 1\).
Figure 7: The left panel displays the optimal linear tax on $c_2$ as a function of the ignored health cost $\gamma$, for a range of redistributive tastes $\nu$, with $\zeta$ declining in income, as in Gruber and Kőszegi (2004). The right panel shows the optimal taxes arising when demand elasticities are constant, at the mean elasticity ($-0.66$) reported in Gruber and Kőszegi (2004).

3.3 Heterogeneous bias

We also explore the quantitative implications of heterogenous bias $\gamma$ across the ability distribution. Since we do not possess data on the degree of bias by income level, this analysis is necessarily hypothetical – however this can provide useful guidance on the importance of determining the extent to which such heterogeneity exists. These simulations, shown in the left panel of Figure 8, are constructed by assuming that bias decreases linearly across the income distribution, with $\gamma = 0$ for top earners. We continue to plot the mean level of bias on the horizontal axis, so for example the point 0.6 on the horizontal axis $\gamma = 1.2$ for the lowest income smokers and $\gamma = 0$ for the highest income smokers, with $\gamma$ interpolated linearly across intervening percentiles. We show these results for a range of redistributive tastes $\nu$. The right panel of Figure 8 reproduces Figure 5 for comparison. The figure shows that heterogeneity in bias can be quite important for the optimal commodity tax, particularly at high levels bias, where the optimal tax is higher due to this heterogeneity. Intuitively, for a given average marginal bias $\bar{\gamma}$, the bias is more important to correct if it is concentrated among low income consumers. In the notation from Section 2, this heterogeneity raises $\sigma$, amplifying the corrective component of the optimal tax. Importantly, both panels of Figure 5 are consistent with the same observed cigarette consumption behavior, highlighting the importance of identifying how behavioral biases vary with income.
Figure 8: The left panel displays the optimal linear tax on \( c_2 \) as a function of the ignored health cost \( \gamma \) when \( \gamma \) is heterogeneous, decreasing linearly across income percentiles (from twice the mean at the lowest income, to \( \gamma = 0 \) at the highest income) for a range of redistributive tastes \( \nu \). The right panel shows the the optimal taxes arising when \( \gamma \) is constant across incomes, reproducing Figure 5.

4 Extensions

4.1 Tax salience

So far, we have assumed that the commodity tax is fully salient. However, recent empirical work suggests that taxes that are not included in posted prices may be ignored by consumers (Chetty et al., 2009; Finkelstein, 2009; Goldin and Homonoff, 2013; Feldman and Ruffle, 2015). An intuition informally suggested in the literature on tax salience is that taxes that are used for corrective purposes, such as reducing cigarette consumption, should be made maximally salient (Feldman and Ruffle, 2015).

In this section, we thus consider a policymaker who can choose between either a fully salient commodity tax \( t \) that is included in the posted prices and a less salient commodity tax \( t_\phi \) that is not included in posted prices and is paid by consumers at the register. We assume that the policymaker can choose either a salient or a non-salient commodity tax, but cannot use a combination.\(^{11} \) For the case of cigarette taxes, for example, the government can either include the tax in the price, or simply charge it at the register. Although state governments in the Unites States also use sales taxes, those are constrained to be the same for a wide variety of commodities and thus cannot target any one specific good in question. Although our analysis could be generalized to consider an optimal mix of both \( t \) and \( t_\phi \), we

\(^{11}\)See Goldin (2015) for a model (of otherwise optimizing consumers and no redistributive concerns) in which the policymaker can combine tax instruments of differing salience to raise revenue in the least distortionary way possible.
suspect that our starting is a reasonable one because of political economy constraints – a politician may have trouble explaining to the public why he chose to break up an otherwise simple tax into shrouded and unshrouded subcomponents.

We begin with the case in which attention to the less salient tax is homogeneous. When the less salient tax is used, consumers choose $c_2$ to maximize $U = z - (p + \phi t_\phi)c_2 + u(c_2) - \psi(z/\theta)$, where $\phi \in [0, 1]$ is consumers’ attention to the less salient commodity tax $t_\phi$. Because of quasilinearity, salience does not affect consumers’ optimal choice of income $z$. Consumers’ choice of $z$ satisfies the first-order condition $\frac{1}{\theta} \psi'(z/\theta) = 1$.

For simplicity, we will focus in this section on regressive sin taxes. We will assume that $\gamma(\theta, c_2) > 0$ for all $\theta, c_2$ and that consumption of $c_2$ is decreasing in $\theta$. Analogous results would hold for the case of regressive subsides for goods that people underconsume. Proposition 6 below characterizes conditions under which the policymaker will choose the more or less salient commodity tax.

**Proposition 6.** Suppose attentiveness $\phi$ is homogeneous. Then:

1. A positive, less-salient commodity tax is never optimal. That is, the optimal policy cannot have $t_\phi > 0$.

2. A negative, fully salient commodity tax is never optimal. That is, the optimal policy cannot have $t < 0$.

Proposition 6, part 1, says that if the policymaker sets a positive tax, then it always has to be the more salient tax. That is, the policymaker should not try to reduce consumption of $c_2$ with the less salient policy instrument. Intuitively, this is because the less salient tax $t_\phi$ would have to be larger than $t$ to achieve the same change in behavior. However, because taxes are regressive, it is not desirable to have a higher tax.

Conversely, part 2 of the proposition says that if the policymaker chooses to subsidize $c_2$, then he should do so with the less salient tax instrument. Intuitively, this is because a less salient subsidy can achieve the same redistributive properties as the more less salient subsidy, while having a smaller impact on behavior. A simple corollary of part 2 of the proposition above, combined with part 1 of Proposition 2, is that if consumer bias is sufficiently small, than the optimal policy must involve a subsidy in the form of the less salient tax instrument:

**Corollary 1.** Under the assumptions of Proposition 2, part 1, there is $\gamma^\dagger > 0$ such that if $\gamma(\theta, c_2) < \gamma^\dagger$ then the planner uses the less salient commodity tax, setting $t_\phi < 0$. 
4.1.1 Heterogeneous attention

We now consider the case in which attentiveness to the tax is heterogeneous. We model salience as being a function of both ability \( \theta \), as well as income \( z \). Both could plausibly affect attentiveness to taxes. On the one hand, higher ability types may be more cognitively skilled to correctly react to taxes. On the other hand, the more wealth a person has, the smaller his marginal utility from money, and thus the less important it is for this person to exert cognitive effort to pay attention to a not fully salient commodity tax.

When \( \phi \) depends on \( z \), consumers’ attentiveness is then partly endogenous to the income tax, which the policymaker must take into account. But now while \( \phi(\theta, z) \) is a function of both \( \theta \) and \( z \), note that because there is a bijection between \( \theta \) and \( z \), we can simply describe tax attentiveness via an indirect function \( \hat{\phi}(z) \) that maps income \( z \) to tax attentiveness.\(^{12}\)

While it is hard to separately identify how \( \phi \) depends on both \( \theta \) and \( z \), the optimal commodity tax \( t_\phi \) can still be expressed compactly in terms of \( \hat{\phi}(z) \). To see this, note that salience bias operates much like overvaluation bias: at a commodity tax \( t_\phi \), a consumer earning \( z \) overestimates the marginal utility from consuming \( c_2 \) by \( (1 - \bar{\phi}(z))t_\phi \) due to salience bias. Thus the total amount by which the consumer overestimates the marginal utility from \( c_2 \) is \( \hat{\gamma}(z) = (1 - \bar{\phi}(z))t_\phi + \gamma(z) \). The optimal tax \( t_\phi \) can then be expressed in terms of \( \hat{\gamma}(z) \) by replacing \( \gamma(z) \) in equation (2) with \( \hat{\gamma}(z) \).

We now illustrate, by way of a simple two-type example, that with heterogeneous salience, it may be optimal to choose a positive, but non-salient sin tax. Suppose that there are just two ability types \( \theta_L \) and \( \theta_H \), and let \( z_L \) and \( z_H \) denote income in the two-type model.

**Proposition 7.** In the two-type model, let \( z^*_L \) and \( z^*_H \) denote the labor income choices that would result when the policymaker chooses an optimal policy \((T, t)\) that restricts to only using the salient commodity tax. And suppose that \( c_2 > 0 \) for both types at this policy. Then using the salient commodity tax is suboptimal if \( \bar{\phi}(z^*_L) \) is sufficiently close to 1, \( \bar{\phi}(z^*_H) \) is sufficiently close to zero, and \( \theta_H \) is sufficiently high.

The intuition behind the Proposition follows in three steps. First, for \( \theta_H \) sufficiently high, the social benefit of giving the high type more \( c_1 \) or \( c_2 \) approaches zero. This means that for \( \theta_H \) sufficiently high, the benefits from correcting type \( H \) consumers’ consumption of \( c_2 \) approach zero, and to a first-order, the policymaker wants to set taxes so as to maximize the revenue he can obtain from these type \( H \) consumers. But plainly, the less salient tax is a more effective way to raise revenue from type \( H \) consumers. At the same time, if \( \bar{\phi}(z^*_L) \) is close to zero, the policymaker can more effectively raise revenue from type \( L \) consumers. As a result, we find that for \( \theta_H \) sufficiently high, it may be optimal to choose a positive, but non-salient sin tax.

\(^{12}\)Here, we continue assuming that salience does not affect period 0 income choices. These are still governed by the FOC \( \frac{1}{\bar{\theta}}\psi'(z/\bar{\theta}) = 1 \).
sufficiently high, then the less salient tax is only slightly less effective at changing type \( L \)'s behavior than the more salient tax.

### 4.1.2 Numerical results for salience

For this analysis we return to the baseline (constant elasticity) specification from subsection 3.1 above. In Figure 9, we consider three different assumptions about salience, in addition to the baseline (plotted in bold). The left panels show the optimal commodity tax on \( c_2 \) for three cases with an average value of \( \phi \) equal to 0.5. In the first, all consumers are equally inattentive. In the second, the bottom quartile of earners are fully attentive (\( \phi = 1 \)), the top quartile are completely inattentive (\( \phi = 0 \)), and the second and third quartiles have \( \phi = 0.75 \) and \( \phi = 0.25 \), respectively, consistent with the finding in Goldin and Homonoff (2013) that low income consumers respond more to cigarette taxes which are not included in the posted price.\(^{13}\) Finally, the last specification reverses this relationship, so that low income consumers are fully inattentive and high income consumers are fully attentive. Although extreme, this specification qualitatively corresponds to the finding in Taubinsky and Rees-Jones (2015) that higher income consumers pay more attention to sales taxes on common household commodity items, at least in part because they are more financially sophisticated and are better at calculating taxes.

We display these results for two degrees of redistributive preference—the baseline \( \nu = 1 \) and a higher value \( \nu = 4 \). The right panels of Figure 9 shows the difference in welfare resulting from the use of the non-salient tax instrument relative to the baseline (fully salient) instrument. This difference is measured in dollars of public funds per person.

As is evident from the left panels, lower salience raises the size of the optimal subsidy when bias is low (indicated by the more negative intercept). Moreover, when the optimal policy is a subsidy, lower saliences makes nonsalient taxes superior instruments for raising welfare, as indicated by the positive welfare gains for low levels of bias in the right panels. These two results demonstrate Corollary 1.

Heterogeneous attention has large implications for the shape of the optimal tax, however. When low income earners are more attentive, the tax is higher than under uniform attention for all degrees of bias. For high bias, this reflects the higher public priority of correcting mistakes of poor consumers than of rich consumers. When high earners are more attentive,

\(^{13}\)Specifically, Goldin and Homonoff (2013) estimates intensive margin elasticities among the top 75% of the income distribution equal to \(-0.31\) for the excise tax (assumed fully salient) and 0.18 for the sales tax (possibly not salient, as sales taxes are typically not included in posted prices). Among the bottom quartile of consumers, they find elasticities of \(-0.3\) (excise) and \(-0.59\) (sales)—see Table 6 and footnotes 33 and 34 in that paper. We assume that the elasticity on the less salient instrument must lie between zero and the elasticity with respect to the fully salient (excise) tax, and thus we approximate their data by assuming salience of 1 for the bottom quartile of earners and 0 for the rest.
the effect is reversed: for high degrees, the optimal tax is lower than under uniform attention, reflecting the lower priority of correcting rich consumers’ mistakes, combined with the desire not to levy heavy taxes on inattentive poor consumers. Finally, the bottom row (with $\nu = 4$) demonstrates Proposition 7 for high levels of bias: a nonsalient tax may be preferable to a salient one, even when the tax is positive and substantial, if attention is inversely correlated with income. Intuitively, when the tax is positive and large in the presence of substantial bias, a less salient tax instrument may be optimal as it allows the high tax to raise substantial revenues from rich consumers, which then raise the lump sum grant for redistributive benefit. Although the nonsalient instrument leads to greater consumption of $c_2$ among the rich, the resulting internalities have only small social welfare costs due to the low welfare weight on high income consumers.
Figure 9: The left panels display the optimal linear tax on $c_2$ under different assumptions about tax salience. The parameter $\phi$ represents the share of the tax that is perceived by consumers. The bold line represents the optimal tax under our baseline scenario, with $\phi = 1$ for all consumers. The next three lines plot the optimal tax when the tax is partially ignored. In each case, the average value of $\phi$ is 0.5. The first has $\phi = 0.5$ for all consumers. The second assumes low income consumers are more attentive, consistent with the results presented in Goldin and Homonoff (2013). The third assumes that attention rises with income. The right panel shows the social welfare gains from implementing the optimal tax in each case, relative to the optimal fully salient tax. (Thus the bold line is mechanically zero in the right panels.) The top row shows these results under our baseline assumption of inequality aversion—with $\nu = 1$, while the bottom shows results with higher inequality aversion of $\nu = 4$. All are computed assuming the baseline elasticity $\zeta = 0.35$.

4.2 Non-financial instruments

We now consider the effects of “persuasion” – such as graphic imagery about the health costs of smoking – that might change consumers’ consumption of $c_2$. As in Section 4.1, we continue restricting to the case in which low income earners have a higher preference for $c_2$, 

34
Suppose that decision utility in periods 0 and 1 is now given by 
\[ \hat{U} = c_1 + \hat{u}(c_2, \theta) - \psi(z/\theta) - sc_2 \] and 
\[ U = c_1 + u(c_2, \theta) - \psi(z/\theta) - sc_2, \]
where \( s \) is the size of the persuasion campaign. The policy maker maximizes the weighted average of utility functions 
\[ V = c + m(c_2, \theta) - \psi(z/\theta) - \gamma(c_2, \theta)c_2 - nsc_2, \]
where \( n \leq 1 \) parametrizes the nudge aspect of the psychological tax. When \( n = 0 \), the policy is a pure nudge in the sense that it changes demand without affecting people’s utility. When \( n = 1 \) the policy is a pure psychological tax.

As usual, the policy maker maximizes \( \int G(V) \). The impact of increasing the tax \( t \) by some small amount \( dt \) is as before. Consider now the impact of increasing the psychological tax \( s \) by the same amount \( ds \). Noting that \( \frac{dc_2}{ds} = \frac{dc_2}{dt} \) by construction, the perturbation \( dt \) changes utility by \( \int [nc_2(z) + s(1-n)]g(z)\frac{dc_2}{dt}dH(z) \). Second, this perturbation decreases revenue by \( t\frac{dc_2}{dt} \). Assuming that the psychological tax costs the government \( \kappa s \) to implement, the net welfare effect is

\[
\frac{dW}{ds} = \int \left[ -nc_2(z)g(z) - \gamma(z)g(z)\frac{dc_2}{dt} + t\frac{dc_2}{dt} + s(1-n)g(z)\frac{dc_2}{dt} \right] dH(z) - \kappa
\]

Thus

\[
\frac{dW}{ds} - \frac{dW}{dt} = \int [(1-n)g(z)c_2(z) - c_2(z)] dH(z) - \kappa
\]

\[ = Cov_H [g(z), c_2(z)] + C_2 - n \int g(z)c_2(z) - C_2 - \kappa \] (14)

\[ = Cov_H [g(z), c_2(z)] - n \int g(z)c_2(z) - \kappa \] (15)

Equation (15) shows that whether persuasion is more beneficial than taxation depends on three terms. The first term, \( Cov_H [g(z), c_2(z)] \), is the regressivity of the tax. When the tax is not regressive, so that tax revenues are recycled to consumers in a way that does not impede redistributive goals, persuasion cannot improve upon the tax. The more regressive the tax, however, the higher the relative benefits to persuasion, because persuasion does not a impose a relatively higher burden on the low income earners than on the high income earners. The

14See also Goldin and Lawson (2015) for a complementary analysis of the optimal combination of a tax and a nudge for a population of active and passive savers.
The second and third terms are simply the social cost of persuasion: the psychological cost that it imposes on consumers and the implementation cost that it imposes on the policymaker. The higher the social cost of persuasion, the lower its impact relative to the commodity tax. We formalize these insights in the proposition below:

**Proposition 8.** Suppose that $\gamma(\theta, c_2) > 0$ for all $\theta, c_2$.

1. Suppose that $G$ is linear or that preferences for $c_2$ do not vary by type. Then the optimal tax system sets $t^G = \bar{\gamma}$ and $s^G = 0$. Psychological taxes are strictly suboptimal when there are no redistributive motives or when there is no preference heterogeneity.

2. Suppose that $G$ is strictly concave and that $c_2$ is decreasing in $z$. Then the optimal policy sets $s^G > 0$ for low enough $n$ and $\kappa$.

3. Under the additional conditions that $\gamma(\theta, c_2)$ is bounded and non-increasing in $\theta$, and that $|G''|$ is bounded away from zero, the optimal policy sets $s^G > 0$ and $t^G < 0$ for low enough $\eta$ and $\kappa$.

Part 1 of the proposition states that when taxes are not regressive, the optimal policy mix should not rely on non-financial instruments. The intuition is that when the tax is not regressive, its revenues are recycled perfectly to consumers in a way that does not increase wealth-inequality, and thus it is a costless way to change behavior. In contrast, persuasion imposes both a psychological cost on consumers and an implementation cost on the policymaker, and thus it is always a costly way to change behavior.

Part 2, however, shows that because a regressive tax is no longer “costless,” it may be optimal to use some persuasion when its total social cost is not too high. Part 3, in fact, shows that under some additional regularity assumptions, if persuasion is sufficiently cheap then it is best to correct behavior with persuasion, and then set a negative tax to redistribute wealth from high income earners to low income earners.

### 5 Conclusion

In this paper, we have analyzed optimal taxation in the presence of both consumer misoptimization and redistributive concerns. Hotly debated policies, such as cigarette taxes or energy efficiency subsidies, must address both consumer misoptimization and the redistributive goals of the government. The policy debates are often polarizing: some claim that regressive taxes are unfair and thus should be eliminated, while others argue for high taxes, focusing mostly on the corrective and revenue-raising aspects of the taxes. This paper provides a framework for tractably considering both motives. Theoretically and numerically,
we show that both motives matter, and we characterize conditions under which each motive is likely to matter the most. Our simulation analysis shows that while the optimal tax is positive for many of the commonly made assumptions about the magnitude of bias, redistributive concerns significantly dampen the size of the optimal corrective tax.

Our framework also clarifies that corrective and redistributive motives do not simply oppose each other. Because an inequality averse government should care more about the poor leaving money on the table than the rich, redistributive motives can amplify corrective motives. We show that the extent to which these two motives reinforce each other can be expressed as an estimable covariance between a person’s income and 1) his elasticity to the sin tax and 2) how biased he is relative to the average consumer. We show that such correlated heterogeneity can significantly change the magnitude of the optimal corrective tax and, for high levels of bias, can generate an optimal tax that is higher than what it would be in the absence of redistributive motives.

In addition to providing a tractable framework for studying corrective and redistributive motives jointly, our work thus underscores the importance of further empirical work on individual differences. Our generally-applicable, quantifiable formulas show that it is not only how biased people are “on average” that matters; it matters who is biased. Whether the mistake is being made by low income or high income consumers, whether the mistake is being made by those more or less elastic to the tax instrument, and whether the low income consumers are relatively more or less elastic to the tax instrument are critical questions for generating robust policy recommendations.
References


A Dynamic model with addiction

Here we show how our model can capture basic dynamic considerations. Suppose that consumers live for $K$ periods $k = 1, \ldots, K$. In each period, a consumer chooses income $z_k$ and pays a tax $T(z_k)$. After-tax income is thus $z_k - T(z_k)$. Note that here, we are assuming (realistically) that the income tax $T$ cannot be age-dependent. Similarly, we are assuming that the commodity tax $t$ cannot be age-dependent. The consumer chooses a consumption bundle $(c_{k1}, c_{k2})$ subject to the budget constraint $c_{k1} + (p + t)c_{k2} \leq z_k$. Note that by assumption, we abstract here from saving and borrowing, as well as time-dependent earnings ability.

The consumer’s true, period $k$ flow utility is given by $V(c_{k1}, c_{k2}, S, z, \theta)$, where $S$ is the stock of $c_2$ consumption that evolves according to $S_{k+1} = (1 - d)(c_{k2} + S_k)$.

Let $U^k(c_{k1}, c_{k2}, S_k, z_k, \theta)$ be the utility function that the consumer maximizes in period $k$. Note that this taxes into account the consumer’s dynamic considerations. The $U$ could be the objective function maximized by either a sophisticated or naive present-biased consumer, for example.

The planner maximizes $\int G \left( \sum_{k=1}^{K} \hat{V}^{k-1}V^k(c_k^1(\theta), c_k^2(\theta), S_k(\theta), z_k(\theta), \theta) \right) dF(\theta)$, where $V^k$ is the period $k$ flow utility and $\mu(\theta)$ is the welfare weight on type $\theta$ in period $k$. In each period $k$ let $V^k(c_{k1}, c_{k2}, S_k, z_k, \theta)$ denote the true utility that results from a choice of $(c_{k1}, c_{k2}, z_k)$ in period $k$, taking into account the how the consumer will choose in periods $k+1, \ldots, K$. As before, set $\gamma(c_2, S, \theta) = V^2_k/V^1_k - U^2_k/U^1_k$.

Then calculations analogous to those in Section 2 show that

$$t = \frac{\bar{\gamma}(\theta)}{\mu(\theta)} \left[ \hat{g} + \frac{Cov[g(\vartheta), \omega(\vartheta)c_2(\vartheta)]}{C_2} \right] - (p + t) \frac{Cov[c_2(\vartheta), g(\vartheta)] + C_2(\bar{g} - 1) + \int z(\vartheta)\chi(\vartheta)T'(z(\vartheta))d\mathcal{F}}{|\xi|C_2}$$

where $\vartheta = (k, \theta)$ is the age-dependent type encoding both the consumer’s age and intrinsic type $\theta$, $\mathcal{F}$ is the distribution of $\vartheta$, where $g(\vartheta) := G^k_1$, and where $\bar{\gamma}$ is the average marginal bias with respect to age-dependent types. This shows that all of the core economic concepts from the static framework carry over to dynamic case.
B Proofs of Propositions (sketches)

Lemma 1. Let $y(\theta)$ and $z(\theta)$ denote a type $\theta$’s post-tax and pre-tax earnings at the optimal tax system. Then $z(\theta) - z(\hat{\theta}) > y(\theta) - y(\hat{\theta}) > (\log(\theta) - \log(\hat{\theta}))\psi'(0)$.

Proof. The first order condition that $y(\theta)$ must satisfy is $y'(\theta) = \frac{1}{\theta} \psi'(z(\theta)/\theta)$. Thus

$$y(\theta^*) - y(\theta) = \int \frac{1}{\theta} \psi'(z(\theta)/\theta)$$

$$\geq \int \frac{1}{\theta} \psi'(0) = (\log(\theta^*) - \log(\theta))\psi'(0)$$

Proof of Proposition 2

Proof. Part 1: We have that

$$\frac{dW}{dt} = \int \left( -g(z)c_2(z,t) - \gamma(z)g(z) \frac{dc_2(z,t)}{dt} + t \frac{dc_2(z,t)}{dt} + c_2(z,t) \right) dH(z)$$

$$= \int \left( -g(z)c_2(z,t) + \gamma(z)g(z) \frac{\zeta|c_2}{p+t} - t \frac{|\zeta|c_2}{p+t} + c_2(z,t) \right) dH(z)$$

(16)

By assumption, there is a finite $A$ such that $\zeta(c_2, \theta)/\zeta(c_2, \theta) < A$ for all $c_2, \theta$.

Thus by assumption, $\int g(z)|\zeta(z)|c_2(z)dH(z) \leq A \int |\zeta(z)|c_2(z)dH(z)$. Thus if $\gamma(z, c_2) \leq \gamma^\dagger$, then the optimal commodity tax must satisfy $t < A\gamma^\dagger$.

Now let $W(t, \gamma)$ denote the highest possible welfare, conditional on a bias function $\gamma$ and a commodity tax $t$. Now by assumption, the derivative $W_t(0, 0) < 0$. Now by continuity, for any interval $[0, a]$ and any $\epsilon > 0$ there exists a $\gamma^\dagger$ such that $|W(t, \gamma) - W(t, 0)| < \epsilon$ for all $t \in [0, a]$ and all $\gamma$ such that $\gamma(z, c_2) < \gamma^\dagger \forall z, c_2$. This implies that there exists a $\gamma^\dagger$ small enough such that $W_t(t, \gamma) < 0$ for all $t \in [0, A\gamma^\dagger]$ and all $\gamma$ bounded above by $\gamma^\dagger$. But since the optimal commodity tax satisfies $t < A\gamma^\dagger$, it thus follows that $t^G < 0$ for $\gamma^\dagger$ small enough.

Part 2: Now if $|\zeta(z, t)|$ and $\omega(z, t)$ are bounded away from zero, then there exists a high enough $\gamma^\dagger$ such that $|\zeta|c_2(\gamma^\dagger/(p + t)) > 1$ for all $z$ and $t \leq 0$. By (16), it thus follows that when $\gamma(c_2, \theta) \geq \gamma^\dagger$, we must have $\frac{dW}{dt} > 0$ for all $t \leq 0$.

Part 3: Now note that for $\gamma(\theta) \geq \gamma^\dagger > t'$, we have that
\[
\frac{dW}{dt} = \int \left( -g(z)c_2(z,t) + (\gamma(z) - t)g(z) \frac{\zeta_c(z,t)}{p + t} + t g(z) \frac{\zeta_c(z,t)}{p + t} - t \frac{\zeta_c(z,t)}{p + t} \right) dH(z)
\]

\[
= \int \left( c_2(z,t) - g(z)c_2(z,t) + (\gamma(z) - t)g(z) \frac{\zeta_c(z,t)}{p + t} \right) dH(z) + \frac{t}{p + t} \text{Cov}[g(z), |\zeta(z)|c_2] \quad (17)
\]

Under these same conditions, there exists a large enough \( \gamma^\dagger \) such that \( (\gamma^\dagger - t) \frac{\zeta_c}{p + t} > 1 \) for all \( t \leq t' \) and all \( z \). Thus \( \frac{dW}{dt} > 0 \) for all \( t \leq t' \) as long as \( \text{Cov}[g(z), |\zeta(z)|c_2(z,t)] \geq 0 \).

**Part 4:** Suppose, for the sake of contradiction, that \( t \leq \gamma^\dagger \). Using the computations from part 3, we have that

\[
\frac{dW}{dt} = \frac{t|\tilde{z}|}{p + t} \text{Cov}[g(z), \omega(z)c_2(z,t)] - \text{Cov}[g(z), c_2(z,t)] + \int (\gamma^\dagger - t)g(z) \frac{\zeta_c(z,t)}{p + t} dH(z). \quad (18)
\]

Now at \( t = \gamma^\dagger \), we have that \( \frac{dW}{dt} = \frac{t|\tilde{z}|}{p + t} \text{Cov}[g(z), \omega(z)c_2(z,t)] - \text{Cov}[g(z), c_2(z,t)] \). Note, however, that because \( c_2(z,t) \to 0 \), while \( \frac{d\psi_2(\theta,t)}{dt} \) is bounded away from zero for \( t < m_1(0, \theta) \), it follows that \( |\tilde{z}| \to \infty \) on the interval \( (0, m_1(0, \theta)) \) as \( t \to m_1(0, \theta) \). It thus follows that for \( \gamma^\dagger \) sufficiently close to \( m_1(0, \theta) \), \( \frac{dW}{dt} > 0 \) at \( t = \gamma^\dagger \). By quasiconcavity, this implies that we cannot have \( t^G \leq \gamma^\dagger \), a contradiction.

**Part 5:** Obvious.

\[\Box\]

**Proof of Proposition 3**

**Proof.** **Part 1.** Analogous to the proof of part 1 of Proposition 2.

**Part 2.** Let \( A > a > 0 \) be such that \( A > |\zeta(c_2, \theta)| > a \) for all \( c_2, \theta \). Now clearly, the optimal commodity tax cannot be greater than \( t^M \). Then from (16), we have that for \( t \in [t', t^M] \) and for \( \gamma \leq \gamma^\dagger < 0 \)

\[
\frac{dW}{dt} < \int \left( \gamma^\dagger g(z) \frac{ac_2(t^M, \theta)}{p + t} - t' \frac{Ac_2(z,t)}{p + t'} + c_2(z,t) \right) dH(z)
\]

\[
< \frac{\gamma^\dagger c_2(t^M, \theta)}{p + t} \frac{t'}{p + t'} Ac_2(t') + C_2(t')
\]

Thus for \( \gamma^\dagger \) sufficiently negative, it follows that \( \frac{dW}{dt} < 0 \) for all \( t \in [t', t^M] \).

**Part 3.** Obvious.

\[\Box\]
Proof of Proposition 4

Proof. Take $\bar{e}$ and $\bar{c}$ such that $e(\theta, t) \in (\bar{c}, \bar{e})$ for all $\theta, t$. Take $\gamma_{\max} = \max_{\theta, t} \gamma(\theta, c(\theta, t))$. Then from (16) it follows that

$$
\frac{dW}{dt} < \int \left( \gamma_{\max} g(\bar{z}) \frac{\bar{c}c(\bar{z}, t)}{p + t} - |\bar{c}(t)| \frac{e(\bar{z}, 0)c(\bar{z}, 0)}{p + t} + c(\bar{z}, 0) \right) dH(z) < 0
$$

This shows that regardless of what value $|\bar{c}(t)|$ takes on, we must have $t^G < \frac{\gamma_{\max} \bar{c}}{\bar{e}}$. Let $\bar{t}^*$ denote this upper bound. Now let $g_0(z)$ denote the social marginal welfare weights when the commodity tax is set to $t = 0$ and the optimal income tax is set optimally. By conditions (i) and (ii) of the proposition, the function $g_0$ is constant over all $U \in \mathbb{U}$. Now because $\text{Cov}[g_0(z), c(\bar{z}, 0)] < 0$ by assumption, there exists a low enough $k$ such that

$$
\int \left( -g_0(z)c(\bar{z}, 0) + |\bar{c}(t)| \gamma(z) g_0(z) \frac{e(\bar{z}, 0)c(\bar{z}, 0)}{p + t} - t |\bar{c}(t)| \frac{e(\bar{z}, 0)c(\bar{z}, 0)}{p + t} + c(\bar{z}, 0) \right) dH(z) < 0
$$

if $|\bar{c}(t)| < k$ for all $t \in [t', \bar{t}^*]$. At the same time, we have that $\max_{z, t \leq \bar{t}^*} \{|c(\bar{z}, t) - c(\bar{z}, 0)|\} \to 0$ as $|\bar{c}(t)| \to 0$. To see this, note that

$$
c(\bar{z}, t) - c(\bar{z}, 0) = \int \frac{|\bar{c}(z, t)|c(\bar{z}, t)}{p + t} \leq c(\bar{z}, 0) \int \frac{|\bar{c}(z, t)|}{p + t} dH(z) \leq k\bar{e} \int \frac{1}{p + t} dH(z)
$$

and thus $c(\bar{z}, t) - c(\bar{z}, 0)$ converges uniformly to 0 as $k \to 0$. From this, it follows that for each $\epsilon > 0$ there exists a small enough $k$ such that

$$
\left| \int \left( -g(z)c(\bar{z}, 0) + |\bar{c}(t)| \gamma(z) g(z) \frac{e(\bar{z}, 0)c(\bar{z}, 0)}{p + t} - t |\bar{c}(t)| \frac{e(\bar{z}, 0)c(\bar{z}, 0)}{p + t} + c(\bar{z}, 0) \right) dH(z) \right| \leq \left| \int \left( -g(z)c(\bar{z}, t) + |\bar{c}(t)| \gamma(z) g(z) \frac{e(\bar{z}, t)c(\bar{z}, t)}{p + t} - t |\bar{c}(t)| \frac{e(\bar{z}, t)c(\bar{z}, t)}{p + t} + c(\bar{z}, t) \right) dH(z) \right| + \epsilon
$$

if $|\bar{c}(t)| < k$ for all $t \in [t', \bar{t}^*]$. But now for $\epsilon$ small enough, it follows that

$$
\int \left( -g(z)c(\bar{z}, t) + |\bar{c}(t)| \gamma(z) g(z) \frac{e(\bar{z}, t)c(\bar{z}, t)}{p + t} - t |\bar{c}(t)| \frac{e(\bar{z}, t)c(\bar{z}, t)}{p + t} + c(\bar{z}, t) \right) dH(z) < 0
$$

for all $t \in [t', \bar{t}^*]$, from which the statement of the proposition follows.

\[\square\]
Proof of Proposition 5

Proof. The proof is identical to the proof of Proposition 4. First, for a bounded bias function there exists some $t^*$ such that for any $\zeta$, the optimal tax is $t^G \geq t^*$. Second, almost identical reasoning shows that there exists a small enough $k$ such that $\frac{dW}{dt} > 0$ if $|\tilde{\zeta}(t)| < k$ for all $t \in [t^*, 0]$.

\[
\Delta(z) := [-\phi t^* c_2^*(z) - (1 - \phi) t^* C_2^*] - (1 - \phi) t^*[c_2^*(z) - C_2^*]
\]

Now by construction $\int \Delta(z) dH(z) = 0$, but $\Delta(z)$ is also decreasing in $z$. Thus this change in policy is preferable because it distributes more resources to the lower income earners.

Proof of Proposition 6. Part 1. Suppose, for the sake of contradiction, that the optimal policy sets $t^*_\phi > 0$. Let $T^*$ denote the corresponding income tax, let $c^*_2(z)$ denote consumption of $c_2$ by a consumer earning $z$ at this tax policy, and let $C^*_2$ denote the total consumption of $c_2$ that occurs at this tax policy. Now consider instead a policy that sets a fully salient commodity tax $t = \phi t^*_\phi$, and sets an income tax $T$ given by $T(z) = T^*(z) - (1 - \phi) t^*_\phi C^*_2$. By construction, a type $\theta$ consumer chooses the same level of both $z$ and $c_2$ under these two tax policies. Thus the only difference is to each consumer’s consumption of $c_1$. Compared to the first policy, the change in $c_1$ consumption of a consumer earning $z$ is given by

\[
\Delta(z) := [-\phi t^* c_2^*(z) - (1 - \phi) t^* C_2^*] - (1 - \phi) t^*[c_2^*(z) - C_2^*]
\]

Now by construction $\int \Delta(z) dH(z) = 0$, but $\Delta(z)$ is also decreasing in $z$. Thus this change in policy is preferable because it distributes more resources to the lower income earners.

Part 2. Suppose, for the sake of contradiction, that the optimal policy sets $t^* < 0$. Let $T^*$ denote the corresponding income tax, let $c^*_2(z)$ denote consumption of $c_2$ by a consumer earning $z$ at this tax policy, and let $C^*_2$ denote the total consumption of $c_2$ that occurs at this tax policy. Now consider instead a policy that sets the less salient commodity tax $t_\phi = t^* / \phi$, and sets an income tax $T$ given by $T(z) = T^*(z) - (1 - 1/\phi) t^* C^*_2$. By construction, a type $\theta$ consumer chooses the same level of both $z$ and $c_2$ under these two tax policies. Thus the only difference is to each consumer’s consumption of $c_1$. Compared to the first policy, the change in $c_1$ consumption of a consumer earning $z$ is given by

\[
\Delta(z) := [-t^* c_2^*(z) / \phi - (1 - 1/\phi) t^* C_2^*] - (1 - 1/\phi) t^*[c_2^*(z) - C_2^*]
\]

Now by construction $\int \Delta(z) dH(z) = 0$, but $\Delta(z)$ is also decreasing in $z$. Thus this change in policy is preferable because it distributes more resources to the lower income earners.
Proof of Proposition 8

Proof. Part 1. This is clear from equation (15), which shows that \( \frac{dW}{ds} - \frac{dW}{dt} < 0 \) for any \( s > 0 \).

Part 2. Suppose, for the sake of contradiction, that the optimal policy sets \( s^G = 0 \) and sets a commodity tax \( \hat{t} \). Now by assumption, \( \text{Cov}[g(z), c_2(z)] > 0 \) at \( \hat{t} \). Thus by (15), \( \frac{dW}{ds} - \frac{dW}{dt} > 0 \) at this tax policy for low enough \( n \) and \( \kappa \), from which it follows that \( s^G = 0, t^G = \hat{t} \) cannot be optimal, thus generating a contradiction.

Part 3. As in the proof of Proposition 4, the boundedness assumption on \( \gamma \) implies that there is a \( \bar{t}^* \) such that \( t^G \leq \bar{t}^* \). Now set \( \hat{\sigma}(t) = \text{Cov}[g(z, t), c_2(z)] \), where \( g(z, t) \) and \( c_2(z, t) \) are defined as follows: given commodity tax \( t \), let \( T_t \) denote the optimal income tax, and let \( g(z, t) \) and \( c_2(z, t) \) denote the corresponding social marginal welfare weights and consumption choices. As we will show below, \( \hat{\sigma}(t) \) is bounded away from zero on the interval \([0, \bar{t}^*]\). Call the lower bound \( \bar{\sigma} \). Next, because \( \int g(z)c_2(z)dH = \text{Cov}[(g(z), c_2(z)] + C_2(z) \), we have that for all \( t \in [0, \bar{t}^*] \),

\[
\frac{dW}{ds} - \frac{dW}{dt} = (1 - \eta)\hat{\sigma}(t) - \eta c_2(z, t) - \kappa = (1 - \eta)\hat{\sigma}(t) - \eta c_2(z, 0) - \kappa
\]

is positive for \( \eta \) and \( \kappa \) small enough. Thus there exist small enough \( \eta \) and \( \kappa \) such that that \( \frac{dW}{ds} - \frac{dW}{dt} > 0 \) for all potential optimal taxes \( t \in [0, \bar{t}^*] \).

We now show that \( \hat{\sigma}(t) \) is bounded from below on the interval. To see this, notice that \( c_2(z, t) \) is increasing in \( z \) and does not depend on the optimal income \( T \). For a given \( t \), \( \hat{\sigma} \) is minimized when \( g(z, t) \) is decreasing as little as possible in \( z \). Now since \( g(z, t) = G'(z - T(z) - (p + t)c_2(z, t) - \gamma(z, t)) \), the assumptions in the proposition imply that

\[
g_z(z, t) \leq (1 - T' - (p + t)\frac{d}{dz}(c_2 + \gamma))G''(z - T(z) - (p + t)c_2 - \gamma) \leq -(p + t)\frac{d}{dz}(c_2)G''(z - T(z) - (p + t)c_2 - \gamma).
\]

Now define \( a < 0 \) such that \( G''(x) < a \) for all \( x \). Next, take some \( \theta^* > \theta, \) and let \( b = \max_{z' \leq \theta^*} \left\{ \frac{dc_2(z,t)}{dz} |_{z=\theta^*} \right\} \). Note that \( b \) is well-defined because \( \frac{dc_2(z,t)}{dz} |_{z=\theta^*} \) is continuous in \( z' \) and \( t \). It thus follows that \( g_z(z, t) \leq -(p + t)ab \) when \( \theta \leq \theta^* \) and \( t \in [0, \bar{t}^*] \). We now have that
\[ \hat{\sigma}(t) \leq F(\theta^*) \text{Cov}[g(z(\theta), t), c_2(z(\theta), t) | \theta \leq \theta^*] \]
\[ \leq F(\theta^*) \text{Cov}[g(z(\theta), t) - (p + t)abz(\theta), c_2(z(\theta)) + bc_2z(\theta)] \]
\[ = F(\theta^*)(p + t)|a|b^2 \text{Var}[z(\theta) | \theta \leq \theta^*] \]

Now from lemma 1 it follows that \( \text{Var}[z(\theta) | \theta \leq \theta^*] \) is bounded away from zero for all \( t \in [0, \bar{t}^*] \), from which it follows that \( \hat{\sigma}(t) \) is bounded away from zero on the interval. 

\[ \square \]