Do Distributional Concerns Justify Lower Environmental Taxes?

Ashley C. Craig[1], Thomas Lloyd[2], and Dylan T. Moore[3]

[1]*Australian National University*
[2]*University of Michigan*
[3]*University of Hawai'i at Mānoa*

March 2026

Abstract

How should taxes on externality-generating activities be adjusted if they are regressive? In our model, the government raises revenue using distortionary income and commodity taxes. If more or less productive people have identical tastes for externality-generating consumption, the government optimally imposes a Pigouvian tax equal to the marginal damage from the externality. This is true regardless of whether the tax is regressive. But, if regressivity reflects different preferences of people with different incomes rather than solely income effects, the optimal tax differs from the Pigouvian benchmark. We derive sufficient statistics for optimal policy, and use them to study carbon taxation in the United States. Our empirical results suggest an optimal carbon tax that is remarkably close to the Pigouvian level, but with higher carbon taxes for very high-income households if this is feasible. When we allow for heterogeneity in preferences at each income level as well as across the income distribution, our optimal tax schedules are further attenuated toward the Pigouvian benchmark.

JEL codes: H21, H23, Q52, Q54

Keywords: optimal taxation, externalities, corrective taxation, income taxation, climate change, carbon tax, commodity taxation, redistribution

Declarations of interest: None.

Acknowledgements: We thank Kai Song (Australian National University) for excellent research assistance. We are grateful to Charlie Brown, James Hines, Matthew Lilley, Ilias Matterne, Matthew Shapiro, Itai Sher, Damián Vergara, and Celine Wasem for comments and suggestions, as well as seminar participants at the University of Michigan, U. Mass. Amherst, the University of Melbourne, the Australian National University (ANU), the 2025 IIPF Annual Congress, and the 2025 National Tax Association Annual Meeting. Financial support from the University of Michigan and the ANU is gratefully acknowledged. This research was approved by the institutional review board at the Australian National University (H/2024/1175) and exempted at the University of Michigan (HUM00261013).

1 Introduction

Human-generated carbon emissions are widely accepted to be causing large social costs, both now and in the future (Intergovernmental Panel on Climate Change, 2023). Economists view this as a classic case for government intervention, because there is an externality: Consumers and producers do not bear the full cost of the emissions they generate.

Most economists also agree that the solution is to impose a "Pigouvian" tax on goods that generate negative externalities—in this case, goods whose production or consumption generates carbon emissions (Akerlof et al., 2019). Setting this tax equal to the marginal damage to others from emissions ensures that consumers and producers pay the true social costs of their choices (Pigou, 1920; Sandmo, 1975). This Pigouvian tax would be optimal if non-distortionary lump sum compensation of winners and losers were possible. However, when redistribution is costly, it is much less clear how a carbon tax should be set given the potential tradeoff between externality correction and distributional impacts.

The fact that carbon taxes are regressive is an important reason why they are unpopular (Hassett et al., 2009; Grainger and Kolstad, 2010; Mathur and Morris, 2014; Metcalf, 2009; Williams III et al., 2015; Pizer and Sexton, 2019). There is widespread concern about their distributional impacts (Dietz and Atkinson, 2010; Dechezleprêtre et al., 2025). As we confirm in our data, this is justified. A uniform tax on all carbon-generating activities would lead poorer households to pay more as a share of their incomes (or expenditures) than richer households. This raises the question of whether corrective taxes should be lowered relative to the Pigouvian benchmark for distributional reasons.

We answer that question. Specifically, we ask how the optimal carbon tax changes if society values equality but compensation can only occur using efficiency-reducing income and commodity taxes. The distributional impact of a carbon tax cannot readily be dismissed in this setting. Imposing the tax produces winners and losers. Compensating the losers by taxing the winners could lead to a greater loss in efficiency than the gain from correcting the externality in the first place. In other cases, the combined package of carbon taxation and compensation could lead to net efficiency gains. Only in knife-edge cases, such as with strong restrictions on household utility functions, is the carbon tax optimally set at the Pigouvian level of "marginal damage" (Kaplow, 2012).

Our contribution is to provide a practical and general characterization of how a corrective tax should be adjusted in light of its distributional impact. We start with a model in which each household chooses its income and carbon intensity. Building on work by Ferey et al. (2024), we use our model to derive sufficient statistics formulas that characterize the optimal trade-off between externality correction and redistribution. Heterogeneity in

preferences for carbon-intensive consumption plays a pivotal role, both within and across income levels. Taking into account both types of heterogeneity, we apply our framework empirically to study the optimal taxation of electricity, natural gas, heating fuels, and gasoline consumption in the United States. Our results suggest that the optimal carbon tax on these goods remains remarkably close to the Pigouvian prescription.

We note that our focus on whether the carbon tax should be changed for distributional reasons is distinct from questions of how much to compensate households, which others discuss in detail (Grainger and Kolstad, 2010; Goulder et al., 2019). Our framework does incorporate flexible compensation, but the pattern of that compensation has no direct connection to the question of whether the carbon tax should differ from the Pigouvian level. Rather the optimal level of the carbon tax hinges on whether it is more efficient to reduce inequality by lowering the carbon tax versus increasing the income tax.

Preference heterogeneity is important because it determines the relative cost of redistribution via income versus carbon taxes. If the carbon tax is set at the Pigouvian level, households allocate their after-tax income efficiently because the externality is exactly internalized. Starting from this level, raising the carbon tax lowers labor supply, while reducing income taxes to compensate increases labor supply. If all households have identical preferences for carbon intensity, the two labor supply effects cancel out. The reason for this is that the reform exactly compensates every household for the impact of the carbon tax not only at their present income, but also at every other level of income they might consider. There is therefore no net impact on individuals' labor supply decisions, and no reason to deviate from purely Pigouvian taxation.

Heterogeneity in preferences breaks the neutrality of the combined reform. This is easiest to see in the simpler case in which households with different income levels vary in their preferences, but households with the same income are equally carbon-intensive. If higher-income households have a relative preference for carbon intensity, so that they would be more carbon-intensive at any given level of disposable income, raising the carbon tax and lowering the income tax yields a net increase in labor supply. In this case, raising the carbon tax redistributes more efficiently than income taxation at the margin. The optimal tax on carbon is therefore greater than the marginal damage. By similar logic, if lower productivity types have more carbon-intensive preferences, lowering the carbon tax below the Pigouvian level becomes an efficient way to redistribute.

Our next step is to extend the model to allow for multidimensional heterogeneity in the sense that carbon intensity is allowed to vary across households with the same level of income. The multidimensional case is more complex because individuals with the same

taxable income may differ in their responsiveness to taxation and their preferences for the carbon-intensive good. Furthermore, adjustments to the income tax can no longer exactly compensate all households for changes to the carbon tax. Nonetheless, we show how to derive an empirically tractable optimal tax condition. While the lessons from the unidimensional case carry over qualitatively, a particularly important result is that variation in the causal effect of higher income on consumption choices tends to dampen any deviation of the optimal tax from its Pigouvian level. This force stems from an inability of the planner to set different tax rates to target different types within the population. It pushes the optimal carbon tax closer to the Pigouvian benchmark, all else equal.

Another way of seeing this is that information about carbon intensity is only useful for redistribution to the extent that carbon intensity is a "tag" in the sense of Akerlof (1978). This is the case if it provides direct information about a household's type beyond what is revealed by an individual's choice of income. For example, if carbon intensity were an immutable characteristic that co-varied with income, then taxing (or subsidizing) carbon-intensive goods would allow some redistribution without efficiency costs.[1] In turn, this would provide a reason other than externality correction to tax (or subsidize) carbon-intensive consumption. Conversely, if all variation in carbon intensity comes from income effects, then taxing carbon-intensive goods has no advantage over income taxation for the purpose of redistribution, so there is no reason to deviate from the Pigouvian level of carbon taxation.

Assuming that differences in preferences do justify deviating from the Pigouvian level of taxation of the externality, distributional concerns lead the government to trade off externality correction and redistribution. As a result, our model highlights that the optimal tax depends on how responsive households are to the carbon tax and income tax. If the income tax is highly distortionary, it can be optimal to use the carbon tax to achieve distributional aims, even if there is a large cost of setting a tax that is different from the Pigouvian level. Using the carbon tax to redistribute is also attractive if doing so has little behavioral impact on carbon intensity. Alternatively, if changing the carbon tax sharply changes carbon emissions, using the carbon tax for redistribution is less desirable.

Unlike the classic Pigouvian case without distributional concerns, the optimal externality correction is not necessarily the same at all levels of consumption. To the extent possible, using means-tested vouchers or non-linear electricity and gas pricing, the optimal carbon tax schedule may be progressive or regressive. There are many real-life examples of non-

[1] Mankiw and Weinzierl (2010) study the example of taxing an individual's height, which is close to this ideal. See also Parsons (1996); Kaplow (2006); and Cremer et al. (2010b). Note that "tagging" logic applies even for endogenous choices as long as those choices reveal information about an individual's type.

linear pricing and taxation schemes such as electricity tiered pricing, congestion taxes, and emission-based vehicle registration fees or subsidies (see e.g., Zhang and Qin, 2015; ACEA, 2020). Whether such non-linearity is optimal depends on the rate at which the preference for carbon intensity changes with income in different parts of the distribution. It does not depend on whether the carbon tax is regressive or progressive in the sense that poorer or richer households pay more as a share of their incomes.

Our final step is to apply our framework to carbon taxation in the United States. Our model highlights the need to empirically isolate differences in preferences between people with different levels of income. This is distinct from income effects, which can also drive differences in carbon intensity along the income distribution.[2] Indeed, it is clearly possible for any given cross-sectional relationship between income and carbon intensity to be explained exclusively by income effects, by preference heterogeneity, or by some combination of the two. The fact that poorer households are more carbon-intensive, and carbon taxes correspondingly regressive, does not therefore suffice to tell us whether the carbon tax should deviate from the Pigouvian benchmark.

We again start with the simpler case of unidimensional heterogeneity, in which preferences vary across income levels but not within. Our model provides a convenient way to identify heterogeneity in preferences for people with different status quo incomes, following Ferey et al. (2021). We compare two statistics. First, we measure the cross-sectional relationship between income and consumption of carbon-intensive goods. This captures both income effects and preference heterogeneity. Second, we measure income effects by quantifying the causal impact of higher income on a given household's consumption of carbon-intensive goods. Finally, we subtract income effects from the cross-sectional relationship to yield the component that arises solely from preferences.

We implement this framework for the United States using survey evidence from 46,918 households in the Consumer Expenditure Surveys (CEX) and 1,448 respondents from a custom online survey. Our analysis focuses on the four expenditure categories that account for the largest share of carbon emissions aggregated into a single "dirty good": gasoline, natural gas, electricity, and heating fuels. In our survey, we ask respondents for information about their income, overall expenditure, and how they split that expenditure over these and other categories. We then ask how their expenditure on each of these categories would change if they had more to spend.

[2]The relevant distinction for our analysis is whether differences in choices by people with different incomes simply reflect differences in the amount they have to spend (income effects), or factors that are independent of income, which would drive different choices even conditional on income. The latter category includes true "preferences", but also habits or constraints that arise due to factors such as where one lives.

Our first step is to measure the cross-sectional relationship between household taxable income and consumption of carbon-intensive goods. While higher-income households consume a greater (absolute) quantity of these goods, they allocate a smaller share of their total consumption to carbon-intensive goods, reflecting an increasing but concave relationship between income and carbon-intensive good consumption.

Next, we measure respondents' causal consumption response to a small and permanent increase in household income across the income distribution. Surprisingly, the results indicate that a pattern of income effects in carbon intensity closely matches the cross-sectional variation in consumption in the Consumer Expenditure Survey. We note that this is not at all mechanical or obvious. Indeed, these causal consumption responses are within-individual effects, which are compared to variation across individuals. The result suggests that cross-sectional variation in consumption of these carbon-intensive goods is largely explained by income effects, and not by differences in preferences.

The interpretation of these survey responses hinges on the reliability of stated preferences. Growing evidence suggests that stated preferences perform well in general (List et al., 2006; Fuster and Zafar, 2023). Yet, reliability inevitably depends on the context and methodology. In our setting, we have a way of verifying the reliability of our survey results. Because we measure status quo allocations across different types of goods in both our survey and in the diary-based Consumer Expenditure Survey, we are able to verify that at least the cross-sectional relationship between income and carbon-intensive good consumption is very similar across both sources. This lends credibility to our survey instrument and sample. We also note that we calibrate the hypothetical income increase to be familiar and relevant for respondents—comparable in size to a small permanent cut in average tax rates or typical lump-sum stimulus checks.

Our empirical results allow us to estimate the optimal tax on carbon-intensive goods. Differences in preferences justify only small deviations from the standard Pigouvian prescription of setting the tax equal to marginal damage. Our estimates suggest that the Pigouvian level of the tax is about 40 cents per dollar of expenditure (assuming a Social Cost of Carbon of $200 per ton). Taking into account differences in preferences, we find that the optimal linear tax would be just 0.2 cents lower than this. The very small magnitude of this adjustment is robust to a wide range of potential parameterizations, reflecting our core empirical result that the causal and cross-sectional relationships between income and carbon intensity match closely on average. Allowing for heterogeneity within as well as across income levels further compresses the optimal tax toward its Pigouvian level.

To the extent that taxes on externality-generating goods can be made non-linear—which has

some precedents—there are somewhat larger deviations from the Pigouvian prescription. For a non-linear tax, we compare the cross-sectional and causal relationships between income and consumption choices at each point in the income distribution. For example, consider households with incomes around $100,000. We find that the estimated causal consumption response is slightly larger than the slope of the cross-sectional relationship between income and carbon intensity. This implies that those earning around $100,000 have a decreasing preference for carbon-intensive goods as we move up the income distribution. In turn, this justifies a carbon tax that is around 1.5 cents per dollar of expenditure lower than the Pigouvian level. By contrast, households with incomes closer to $250,000 would face carbon taxes of 43 cents, which is 7.5 percent higher than the Pigouvian level.

Our paper provides an empirical and theoretical framework that could be applied more broadly to environmental taxes, or any other tax designed to correct an externality. Consistent with the principle of additivity (Sandmo, 1975; Kopczuk, 2003), the total tax on any externality-generating activity is a combination of a "Pigouvian" correction and a distributional component. We provide a way to measure the distributional component and show how it depends on causal and cross-sectional sufficient statistics. The optimal tax on the externality may deviate positively or negatively from the Pigouvian prescription. In some cases, the total tax could even flip sign relative to the externality. In our case, we find that distributional concerns have only a minor impact on the optimal carbon tax.[3] However, applying our framework to other taxes may yield different results.

Connections in the Literature

Externalities have long been recognized as providing a motivation for government intervention. Corrective taxation as a solution dates back to Pigou (1920). Economists have studied wide varieties of externalities in practice, including environmental taxes (Weitzman, 1974; Andersson, 2019), occupation choice (Lockwood et al., 2017), congestion (Walters, 1961), labor market distortions (Craig, 2023), control of a pandemic (Craig and Hines, 2020), rent-seeking (Rothschild and Scheuer, 2016), and positional externalities (Aronsson and Johansson-Stenman, 2008).

A subset of this literature incorporates distributional concerns. Kopczuk (2003) shows how the optimal tax on an externality can be written as the sum of an externality correction component and a separate term capturing other objectives, building on Sandmo (1975). Kaplow (2012) studies a special case without preference heterogeneity, showing that the optimal tax is equal to marginal damage in this case. Pirttilä and Tuomala (1997) allow for

[3]We show in Appendix sections A.6 and A.7 that this is also true for the individual components of our carbon-intensive aggregate good: electricity, natural gas and heating fuels, and gasoline.

interactions with labor supply. Jacobs and van der Ploeg (2019) focus on linear taxation. Cremer et al. (1998) and Cremer et al. (2003) consider optimal taxation and externality correction with a small number of types, demonstrating the importance of homogeneity and separability in preferences, highlighting the potential for the optimal tax on the externality-generating good to have the opposite sign from the externality. Ahlvik et al. (2024) focus on different types of equity considerations and differences in cost burdens. They use a mechanism design approach to study the interaction of income taxation and externality correction, including an empirical exercise with much stronger restrictions on individual heterogeneity than ours. Jacobs and de Mooij (2015) ask whether the Pigouvian correction should be adjusted for the marginal cost of public funds. Belfiori et al. (2024) focus on compensation in a macroeconomic model without heterogeneity in preferences for carbon intensity. Micheletto (2008) studies the correction of externalities that are not "atmospheric" so that externality correction is imperfectly targeted.

Methodologically, the most closely related work is by Allcott et al. (2019) and Ferey et al. (2024). These papers focus primarily on "internality" correction, and savings taxation when the government values equality and taxes are distortionary. The authors highlight the core insight that causal income effects can be subtracted from cross-sectional variation in consumption patterns to establish the degree to which preferences vary across income levels. Our contributions are to: (i) apply these ideas to shed light on how externality-correcting taxes should change with preference heterogeneity, including multidimensional heterogeneity; and (ii) study carbon emissions by households, which is one of the most important examples of an externality. These papers and ours build on previous work (Saez, 2002; Piketty and Saez, 2013; Diamond and Spinnewijn, 2011; Gauthier and Henriet, 2018); and connect to concurrent theoretical analysis by Doligalski et al. (2025).

Empirically, recent studies highlight significant heterogeneity in consumer preferences for carbon-intensive goods vs. cleaner alternatives. For example, Lyubich (2024) decomposes the unobserved heterogeneity in US household carbon consumption and finds, by and large, a greater role for person effects (up to 50% of the variance decomposition) relative to place effects (up to 23%). Similarly, Dorsey and Wolfson (2024) provide evidence of across-income preference heterogeneity in solar panel uptake.

Our focus is on heterogeneity between households and the importance of the distortionary costs of redistributive taxation. We do not incorporate general equilibrium effects or dynamics, which can provide important additional motives for deviating from standard Pigouvian taxation. This has been considered in other work (Cremer and Gahvari, 2001; Cremer et al., 2010a; Douenne et al., 2025; Barrage, 2020; Bierbrauer, 2024; Fried et al., 2024).

Similarly, we do not consider multinational issues (see Aronsson and Blomquist, 2003) or explicitly model multiple generations (Fried et al., 2018). Our work complements Fricke et al. (2025), who study the optimal rebate when carbon emissions are capped.

Roadmap for the Paper

In Section 2, we introduce a model of optimal taxation with flexible tax instruments, an externality-generating good, and heterogeneity across but not within levels of income. We show how the optimal taxes on income and the externality interact, and derive empirically estimable sufficient statistics to characterize optimal taxation. Section 3 extends our results to multidimensional heterogeneity. Section 4 then applies these results to carbon taxation in the United States. Section 5 concludes.

2 Externality Correction With Taste Heterogeneity

Consider a population of individuals indexed by type $w \in W$, where w has differentiable cumulative distribution F and W is compact. Individual w chooses taxable income $z \in [0, \overline{z}]$. Income has differentiable cumulative distribution $H_z(z)$ and density $h_z(z)$. Income is subject to nonlinear income taxation with tax schedule $T_z\left(z\right)$. She allocates her after-tax income, $z - T_z(z)$ between a numeraire consumption good, c, and an externality-generating good, x. A tax, $T_x(x)$ must be paid on consumption of the externality-generating good. We constrain $T_z\left(z\right)$ and $T_x(x)$ to be twice continuouslydifferentiable.

We focus in this section on the case in which heterogeneity is unidimensional in the sense that $w \in \mathbb{R}$ and utility is $u(c, x, z \mid w)$. This implies that there is no variation in preferences conditional on w. However, individuals with different w may have different preferences over the two consumption goods. Specifically, if two individuals with different w were forced to have the same z and $z - T(z)$ they may prefer different consumption bundles. In Section 3, we extend our results to allow for multidimensional heterogeneity.

Separable Benchmark. As an important benchmark, we will consider the case, as in Atkinson and Stiglitz (1976), where preferences over consumption goods are weakly separable from labor supply, and homogeneous across agents, including those of different types. In this special case, it is well-known that we obtain the standard Pigouvian result that it is optimal to tax the externality-generating good at a constant rate equal to the marginal external damage associated with consuming the good (Kaplow, 2012).

2.1 Individual Maximization

Each agent maximizes her utility subject to her budget constraint and the tax system set by the social planner, producing indirect utility $v(w)$.

$$v(w) = \max_{z,x,c} u(c, x, z \mid w) \tag{1}$$
$$\text{s.t.} \quad c + x + T_x(x) = z - T_z(z)$$

We assume that $u(c, x, z \mid w)$ is twice continuously differentiable with bounded first derivatives, increasing in c and x, and decreasing in z. We further assume that $u(c, x, z \mid w)$ is strictly concave in c, x, and z.

We will often find it useful to approach (1) as a two stage maximization problem. First, a type w chooses their income, $z(w)$. Second, they decide how to split the result disposable income between the two consumption goods, leading to choices $c(w)$ and $x(w)$. Specifically, a type w agent chooses taxable income:

$$z(w) \equiv \arg\max_{z} \{u(z - T_z(z) - x(w; z) - T_x(x(w; z)), x(w; z), z \mid w)\} \tag{2}$$

where:

$$x(w; z) \equiv \arg\max_{x} \{u(z - T_z(z) - x - T_x(x), x, z \mid w)\} \tag{3}$$

is the amount of the externality-generating good a type w agent would choose if his or her taxable income were z. The amount of the externality-generating good that a type w agent actually chooses is $x(w) \equiv x(w; z(w))$.

This approach allows us to write first-order conditions characterizing a type w agent's choices as a function of both tax schedules:

$$1 - \mathcal{T}_z'(z) - \frac{\partial x(w; z)}{\partial z}\left(1 + \mathcal{T}_x'(x)\right) = -\frac{u_z + \frac{\partial x(w;z)}{\partial z} u_x}{u_c} \tag{4}$$

$$\text{and} \qquad 1 + \mathcal{T}_x'(x) = \frac{u_x}{u_c}. \tag{5}$$

For the sake of brevity we have suppressed the dependence of the marginal utility (e.g. u_c) terms on type (w) and/or on income (z). These two conditions can be combined to yield a reduced first order condition for the optimal choice of income: $1 - \mathcal{T}_z'(z) = -\frac{u_z}{u_c}$, but this masks important intuition as we highlight below.

Equation 4 highlights something important: an agent's anticipation of their own behavior at stage two enters their first stage decision: For each additional dollar of income they consider earning, they evaluate how it would change their disposable income and how much they would spend on good x, given their preferences and the tax rate. In this way, the agent can be thought of as choosing their level of taxable income subject to an *effective net of income tax rate* of $1 - \mathcal{T}_z'(z) - \frac{\partial x(w;z)}{\partial z}\mathcal{T}_x'(x)$. Thus, when a rational agent sees their purchasing power decline due to commodity taxes, this reduces their marginal benefit of

earning income. The size of this reduction depends on what fraction of their marginal dollar of disposable income will be spent on the dirty good. Critically for the analysis below, this marginal spending behavior may differ across people.

2.2 Social Welfare Maximization

A welfarist social planner chooses twice-differentiable tax functions, $\mathcal{T}_z$ and $\mathcal{T}_x$. It does so to jointly to maximize social welfare, $\boldsymbol{\mathcal{W}}$, while raising enough revenue to cover an exogenous revenue requirement, $\mathcal{R}$.

$$\max \boldsymbol{\mathcal{W}} = \int \gamma(w) v(w) dF - \mathcal{D}(\overline{x}) \tag{6}$$

$$\text{s.t.} \quad \mathcal{R} = \int \left(T_z(z(w)) + T_x(x(w)) \right) dF(w) \tag{7}$$

$$\text{and} \quad \overline{x} = \int x(w) dF$$

Here, $\gamma(w)$ are Pareto weights; and $\mathcal{D}(\overline{x})$ is a separable atmospheric externality, which is an increasing function of aggregate consumption of the externality-generating good.

This formulation is equivalent to having the harm of aggregate consumption of the externality-generating good entering individuals' utility in a separable way. To see the equivalence between the two ways of defining social welfare, consider $\tilde{u} \equiv u(c, x, z \mid w) - \widetilde{x}(\bar{x}; w)$, then $\tilde{v}(w) = v(w) - \widetilde{x}(\bar{x}; w)$ and $\mathcal{D}'(\overline{x}) = \int \gamma(w) \widetilde{x}'(\bar{x}; w) dF$. Our formulas are also unaffected if we define welfare using a social welfare function, $\mathcal{W} \equiv \int W\left(v\left(w\right)\right) dF - \mathcal{D}(\overline{x})$, rather than Pareto weights. However, the distinction between the two approaches would matter for non-marginal tax changes.

2.3 Regularity Assumptions

For all tax schedules we consider, we assume that the optimal choices of c, x, and z are globally unique for every individual, with the relevant second-order conditions holding strictly. We also assume there is a smooth and strictly monotonic mapping from w to $c(w)$, $x(w)$, and $z(w)$. This is equivalent to the approach of Jacquet and Lehmann (2021) of making a single-crossing assumption and verifying monotonicity. These assumptions are standard and are required to ensure that choices move smoothly as the tax system changes, but imply restrictions on the set of tax schedules under consideration.

2.4 Characterizing Tax Reforms

We characterize the welfare effects of reforming the tax system using a formalization of the perturbation method of Saez (2001) via Gateaux derivatives, an approach discussed more

thoroughly in Golosov et al. (2014), Jacquet and Lehmann (2021), and Gerritsen (2024). Given an initial tax system with the tax schedules, $\mathcal{T}_x(x)$ and $\mathcal{T}_z(z)$, we ask how welfare and revenue change when we perturb them in the direction of $\tau_x(x)$ and $\tau_z(z)$ respectively, with $\tau_x(x)$ and $\tau_z(z)$ bounded and continuously differentiable. Letting $\kappa \in \mathbb{R}$ index the magnitude of the perturbation, we can define the reformed tax schedules as follows:

$$T_z(z;\kappa) \equiv \mathcal{T}_z(z) + \kappa\tau_z(z) \qquad \text{and} \qquad T_x(x;\kappa) \equiv \mathcal{T}_x(x) + \kappa\tau_x(x). \tag{8}$$

All our results about efficient and optimal tax policy will be based on examining the first-order effect of such perturbations on the social planner's Lagrangian,

$$\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = \left.\frac{\partial \mathcal{W}}{\partial \kappa}\right|_{\kappa=0} + \frac{1}{\lambda}\left.\frac{\partial \mathcal{R}}{\partial \kappa}\right|_{\kappa=0},$$

starting from the initial tax system. Here, λ is the multiplier on the government's budget constraint. For any initial tax system, and proposed set of reforms, if $\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} > 0$, then movement in that direction is welfare-improving. At the optimal tax system, no such improvements can be possible and thus, at the optimum, we must have:

$$\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = 0 \quad \text{for all } \tau_z(\cdot) \text{ and } \tau_x(\cdot). \tag{9}$$

2.5 Distribution Neutral Reforms

To begin our investigation of the optimal tax system, consider a special type of tax reform: one which is *distribution neutral*.

Let $\hat{w}(z)$ be the type of agent that chooses income z. In other words, it is the inverse the taxable income choice function, $z(w)$. This allows us to define $\hat{x}(z) \equiv x(\hat{w}(z); z)$, the dirty good consumption of an agent with taxable income z.

A distribution neutral tax reform is one where for a given change in the dirty good tax liability of agents with taxable income z, $\tau_x(\hat{x}(z))$, we also change their income tax liability by a precisely offsetting amount, such that the agent can still afford their original bundle. That is to say, the income tax reform is defined by:

$$\tau_z(z) = -\tau_x(\hat{x}(z)). \tag{10}$$

Notice, this implies marginal income tax rate changes of:

$$\tau_z'(z) = -\hat{x}'(z)\,\tau_x'(\hat{x}(z)), \tag{11}$$

where $\hat{x}'(z)$ is the slope of the observed cross-sectional relationship between dirty good consumption and taxable income.

Distribution neutral reforms are of special interest because of the relative simplicity of their impact on social welfare. For any such reform, there is no first-order impact on agent utility, nor is there a mechanical impact on revenue. All effects of the reform arise because behavioral responses affect the externality and government revenue.

This also implies that for any distribution neutral reform, if $\frac{\partial \mathcal{L}}{\partial \kappa}\big|_{\kappa=0} > 0$, then a reform in that direction is a Pareto improvement: Positive gains in government revenue net of externality costs can be obtained without changes to individual agent utility, which allows a small transfer to be made that makes all agents strictly better off. Put another way, for any Pareto efficient tax system, it must be the case

$$\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = 0 \quad \text{for all } \tau_z(\cdot) \text{ and } \tau_x(\cdot) \text{ satisfying } \tau_z(z) = -\tau_x(\hat{x}(z)) \tag{12}$$

2.6 Behavioral Responses

By construction, any distribution-neutral reform produces a change in relative prices without a change in their purchasing power. Thus, only compensated behavioral responses will be relevant to the analysis. To establish what kinds of responses will occur, we can revisit the agent's first-order conditions.

Consumption of the Dirty Good (Direct). The reform to T_x changes the stage 2 first-order condition of an agent with taxable income z in the following way:

$$1 + \mathcal{T}_x'(\hat{x}(z)) + \kappa \tau_x'(\hat{x}(z)) = \frac{u_x}{u_c}. \tag{13}$$

Holding constant the agent's choice of z, this generates a behavioral response of:

$$\left.\frac{\partial \hat{x}(z)}{\partial \kappa}\right|_{z,\,\kappa=0} = -\frac{\varepsilon_{x|z}(z)\,\hat{x}(z)}{1+\mathcal{T}_x'(\hat{x}(z))}\tau_x'(\hat{x}(z)) \tag{14}$$

where:

$$\varepsilon_{x|z}(z) \equiv -\frac{1+\mathcal{T}_x'(\hat{x}(z))}{\hat{x}(z)}\left.\frac{\partial \hat{x}(z)}{\partial \mathcal{T}_x'}\right|_z^c \tag{15}$$

is the compensated elasticity of spending on x with respect to its tax-inclusive price.

Income. Turning our attention to the stage 1 first-order condition, we have the following perturbed first-order condition:

$$1-\mathcal{T}_z'(z)+\underbrace{\kappa\hat{x}'(z)\,\tau_x'(\hat{x}(z))}_{\text{Reform to } T_z}-x'_{\text{inc}}(z)(1+\mathcal{T}_x'(x)+\underbrace{\kappa\tau_x'(\hat{x}(z))}_{\text{Reform to } T_x})=-\frac{u_z+x'_{\text{inc}}(z)u_x}{u_c}. \quad (16)$$

where $x'_{\text{inc}}(z) \equiv \frac{\partial x(\hat{w}(z);z)}{\partial z}$ is the causal effect of increasing the taxable income of a type $\hat{w}(z)$ agent on their choice of dirty good consumption.

Notice, in the two-stage formulation of the agent's problem, that this first-order condition depends on the reforms to both tax schedules. This is because, when responding to the reform at stage 1, the agent anticipates how the reform to the tax on the dirty good will impact their marginal return to earning income, given that some of their income will be spent on that good. After rearranging the left hand side of the equation, the reformed effective marginal income tax rate can be rewritten as:

$$1-\mathcal{T}_z'(z)-x'_{\text{inc}}(z)\left(1+\mathcal{T}_x'(x)\right)+\kappa\left(\hat{x}'(z)-x'_{\text{inc}}(z)\right)\tau_x'(\hat{x}(z)). \quad (17)$$

This formulation highlights that heterogeneity in preferences leads to a distinction between the causal effect of additional income on dirty good consumption, and the cross-sectional relationship between income and dirty good consumption. Following Ferey et al. (2024), we let $x'_{\text{het}}(z)$ be the component of the cross-sectional relationship that is driven by heterogeneity in tastes:

$$x'_{\text{het}}(z) \equiv \hat{x}'(z)-x'_{\text{inc}}(z). \quad (18)$$

As equation 16 shows, the change in the effective marginal income tax rate induced by the reform reform is directly proportional to this preference heterogeneity effect. As a result, the behavioral response of taxable income induced by the reform is:

$$\left.\frac{\partial z}{\partial\kappa}\right|_{\kappa=0}=\frac{z\varepsilon_z(z)}{1-\mathcal{T}_z'(z)}x'_{\text{het}}(z)\tau_x'(\hat{x}(z)) \quad (19)$$

where:

$$\varepsilon_z(z)\equiv\frac{1-\mathcal{T}_z'(z)}{z}\left.\frac{\partial z}{\partial\left(1-\mathcal{T}_z'(z)\right)}\right|_z^c \quad (20)$$

is the compensated elasticity of taxable income with respect to the net-of-income-tax rate. Note that formal definitions of these elasticities are included in the proof of Proposition 1.

Consumption of the Dirty Good (Indirect). To complete our enumeration of behavioral responses, we consider the subsequent impact of the induced change in income on consumption of x. For an agent earning income z, this is determined by the causal effect of income on the dirty good,

$$\left.\frac{\partial \hat{x}(z)}{\partial z}\right|_{\kappa=0}\left.\frac{\partial z}{\partial \kappa}\right|_{\kappa=0} = x'_{\text{inc}}(z)\frac{z\varepsilon_z(z)}{1-\mathcal{T}'_z(z)}x'_{\text{het}}(z)\tau'_x(\hat{x}(z)), \tag{21}$$

resulting in a total effect of the reform on dirty good consumption of:

$$\left.\frac{\mathrm{d}x}{\mathrm{d}\kappa}\right|_{\kappa=0} = \left.\frac{\partial \hat{x}(z)}{\partial \kappa}\right|_{z,\,\kappa=0} + \left.\frac{\partial x}{\partial z}\right|_{\kappa=0}\left.\frac{\partial z}{\partial \kappa}\right|_{\kappa=0}. \tag{22}$$

2.7 A Key Sufficient Statistic: The "Taste Elasticity"

The formulation above highlights the key role of heterogeneity in preferences. Specifically, our optimal tax conditions will feature an important additional sufficient statistic: the "taste elasticity", $\eta^{\text{Taste}}_{x,z}(z)$. This is the elasticity of consumption of x with respect to income that occurs only through heterogeneity in agent tastes:

$$\eta^{\text{Taste}}_{x,z}(z) \equiv \frac{zx'_{\text{het}}(z)}{\hat{x}(z)}. \tag{23}$$

The taste elasticity captures any differences in consumption of x across income that occur not due to the causal effect of higher income, but due to differences in the preferences of agents who locate at different parts of the income distribution.

In the Atkinson and Stiglitz (1976) benchmark in which preferences are identical across agents and separable from labor supply, $\eta^{\text{Taste}}_{x,z}(z) = 0$. In this special case, both the income and indirect consumption effects (equations 19 and 21) are zero. The logic of the two-stage decision process above sheds light on why. Recall that a distribution neutral reform uses the income tax to exactly offset the financial impact of the altered tax on the dirty good. If all agents have the same preferences, this also means that every agent's evaluation of every counterfactual income choice is also unchanged, because all agents would make the same consumption decisions if they had the same income level. This implies that the reform induces no change to any agent's choice of income.

More generally, differences in tastes across types can lead to a positive or negative value of $\eta^{\text{Taste}}_{x,z}(z)$, depending on whether higher or lower types have a relative preference for the dirty good. This implies potential labor supply impacts from the distribution neutral reform. To see why, recall the two stage decision process. An agent of type w_0 and income z_0 considers whether to increase her income slightly to z_1. If $\eta^{\text{Taste}}_{x,z}(z) < 0$, she would consume less of x given income z_1 than type w_1 agents (who do have income z_1). Now consider a reform that increases $T_x(\hat{x}(z_1))$ while lowering $T_z(z_1)$ just enough to leave the utility of type w_1 unchanged at z_1. This same reform must raise the utility that w_0 would

receive at z_1, making that higher income more attractive. Conversely, if $\eta^{\text{Taste}}_{x,z}(z) < 0$, the same distribution neutral reform reduces the incentive to increase one's income.

2.8 Testing Local Pareto Efficiency: A Specific Reform

To characterize the set of Pareto efficient tax systems, consider a specific distribution neutral tax reform where we only alter the marginal commodity tax rate for agents earning taxable income z, increasing it:

$$\tau'_x(\hat{x}(z')) = \begin{cases} 1 & \text{if } z' = z \\ 0 & \text{otherwise} \end{cases}. \tag{24}$$

Distribution neutrality then requires an offsetting reduction of those same agents' marginal income tax rates. Specifically, $\tau'_z(z') = -\hat{x}'(z')$ if $z' = z$ and $\tau'_z(z') = 0$ otherwise.

2.9 Welfare Effects of the Reform

We obtain the welfare effect of a reform by multiplying each behavioral response (equations 14, 19, and 21) by its marginal impact on social welfare. For changes in consumption of the dirty good (equations 14 and 21), the marginal social value is the difference between the marginal tax rate on that good and its marginal social damage, i.e. $\mathcal{T}'_x(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}$. For changes in income, the marginal social value is simply the income tax rate $\mathcal{T}'_z(z)$. In total, the impact on social welfare is the sum of three effects:

$$\Delta\mathcal{W}(z) = \Delta\mathcal{W}^z(z) + \Delta\mathcal{W}^{x|z}(z) + \Delta\mathcal{W}^{z\to x}(z) \tag{25}$$

where the components are defined as follows.

$$\Delta\mathcal{W}^z(z) = \mathcal{T}'_z(z)\frac{z\varepsilon_z(z)}{1-\mathcal{T}'_z(z)}x'_{\text{het}}(z) \tag{26}$$

$$\Delta\mathcal{W}^{x|z}(z) = -\left(\mathcal{T}'_x(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})(z)}{\lambda}\right)\frac{\varepsilon_{x|z}(z)\,\hat{x}(z)}{1+\mathcal{T}'_x(\hat{x}(z))} \tag{27}$$

$$\Delta\mathcal{W}^{z\to x}(z) = \left(\mathcal{T}'_x(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right)x'_{\text{inc}}(z)\frac{z\varepsilon_z(z)}{1-\mathcal{T}'_z(z)}x'_{\text{het}}(z) \tag{28}$$

The first of these effects, $\Delta\mathcal{W}^z$, captures the impact on government revenue from taxable income changes. The second, $\Delta\mathcal{W}^{x|z}$, is the uninternalized component of the externality caused by changes in consumption of the externality-generating good, which are directly caused by the tax reform. Finally, $\Delta\mathcal{W}^{z\to x}$ is the effect of further changes in consumption of x that are induced indirectly by changes in taxable income.

2.10 A Neccessary Condition for Pareto Efficiency

Setting the total impact on welfare ($\Delta\mathcal{W}$) equal to zero yields an optimal tax condition at each income level z, which simply requires that there be no net gain to distribution neutral tax reforms at any given level of income.

$$\frac{\mathcal{T}_x'\left(\hat{x}\left(z\right)\right)-\frac{\mathcal{D}'(\bar{x})}{\lambda}}{1+\mathcal{T}_x'\left(\hat{x}\left(z\right)\right)}=\eta_{x,z}^{\text{Taste}}\left(z\right)\frac{\varepsilon_z\left(z\right)}{\varepsilon_{x|z}\left(z\right)}\left(\frac{\mathcal{T}_z'\left(z\right)+\left[\mathcal{T}_x'\left(\hat{x}\left(z\right)\right)-\frac{\mathcal{D}'(\bar{x})}{\lambda}\right]x_{\text{inc}}'(z)}{1-\mathcal{T}_z'\left(z\right)}\right) \tag{29}$$

In our proof in Appendix B, we derive equation 29 more formally, considering the broader set of distribution neutral reforms. Specifically, the result follows from applying the fundamental lemma of the calculus of variations.

Our final step is to rearrange equation 29 to isolate the wedge between the marginal tax rate on x and its marginal damage, i.e. $\mathcal{T}_x'(\hat{x}(z))-\frac{\mathcal{D}'(\bar{x})}{\lambda}$. This yields Proposition 1.

Proposition 1. *If the income and commodity taxes are optimal and are allowed to be non-linear, they must satisfy equation 30:*

$$\mathcal{T}_x'\left(\hat{x}\left(z\right)\right)-\frac{\mathcal{D}'(\bar{x})}{\lambda}=\frac{RR(z)\mathcal{T}_z'\left(z\right)}{1-RR(z)x_{inc}'(z)} \tag{30}$$

where:

$$RR(z)\equiv\eta_{x,z}^{Taste}\left(z\right)\frac{\varepsilon_z\left(z\right)}{\varepsilon_{x|z}\left(z\right)}\frac{1+\mathcal{T}_x'\left(\hat{x}\left(z\right)\right)}{1-\mathcal{T}_z'\left(z\right)} \tag{31}$$

measures relative responsiveness to changes in the carbon versus the income tax rates. If this condition is not satisfied, there exists a simultaneous reform to both taxes that yields a Pareto gain.

From Equation 30, we can see that $\eta_{x,z}^{\text{Taste}}\left(z\right)$ is a sufficient statistic that determines the sign and scales the size of any deviation from setting the marginal tax on x equal to the marginal damage, $\mathcal{D}'(\overline{x})$. If there is no taste heterogeneity, i.e., $\eta_{x,z}^{\text{Taste}}\left(z\right)=0$, the optimal corrective tax is the standard Pigouvian one: $\mathcal{T}_x'\left(\hat{x}\left(z\right)\right)=\mathcal{D}'(\bar{x})/\lambda$. In this case, there is no "tagging" role for the carbon tax, because tastes do not change along the income distribution. Only if $\eta_{x,z}^{\text{Taste}}\left(z\right)\neq 0$ does the optimal carbon tax differ from the Pigouvian benchmark. We can also see that such deviations may vary across different levels of income, which could motivate means-tested rebates or subsidies for externality-producing goods.

Unlike the standard Pigouvian tax, the optimal tax depends on the magnitude of behavioral responses. The stronger the demand response (i.e. the larger is $\varepsilon_{x|z}\left(z\right)$), the closer the optimal tax will be to the Pigouvian benchmark. This happens because the planner's ability to successfully exploit tagging and use deviations from the Pigouvian level of the tax to

redistribute depends on how responsive consumption of the tag is to taxation. Conversely, a higher value of the elasticity of income favors a larger deviation from Pigouvian taxation: If the income tax is highly distortionary, using other instruments such as the carbon tax can be an efficient alternative. For a similar reason, we see that the deviation from Pigouvian taxation is proportional to the level of the marginal tax rate on income: A low income tax rate implies a small marginal distortion from the income tax in the first place.

The optimal tax could even have the opposite sign of the marginal damage, at least in theory. In this case, a good causing a negative externality is optimally subsidized, or a good causing a positive externality is taxed. This occurs if $\eta_{x,z}^{\text{Taste}}(z)$ is negative and the magnitude of the optimal deviation is large. That could happen because consumption of x is a very effective tag (i.e. $\eta_{x,z}^{\text{Taste}}(z)$ is itself large), because the response to the tax is small ($\varepsilon_{x|z}(z)$ is small), the Pareto parameter of taxable income ($1 - H_z(z) / [zh_z(z)]$) at z is small, and/or the social marginal utility is small at incomes above z.

The denominator of equation 30 amplifies positive deviations from the Pigouvian level, and dampens negative deviations. Suppose that the taste elasticity is positive, motivating a carbon tax greater than the Pigouvian level. Raising the carbon tax and lowering the income tax produces efficiency gains. Incomes increase. If $x'_{\text{inc}}(z) > 0$, consumption of the carbon-intensive good increases as well. There is an uninternalized social *benefit* to that to the extent that $\mathcal{T}'_x(\hat{x}(z)) > \frac{\mathcal{D}'(\bar{x})}{\lambda}$, amplifying the original gains further. Conversely, suppose that the taste elasticity is negative. Then we realize efficiency gains as we lower the carbon tax below the Pigouvian level, the resulting rise in incomes and carbon-intensive consumption produces an uninternalized social *cost* because $\mathcal{T}'_x(\hat{x}(z)) < \frac{\mathcal{D}'(\bar{x})}{\lambda}$. This makes the efficiency gains from lowering the carbon tax less attractive than otherwise.

A different way of seeing this is that using the carbon tax for redistribution is more beneficial if the income tax is more distortionary. The net wedge from the income tax is:

$$\mathcal{T}'_z(z) + \left[\mathcal{T}'_x(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x'_{\text{inc}}(z). \tag{32}$$

It is clear from this expression that raising $\mathcal{T}'_x$ increases that wedge, making the income tax more distortionary, and thus relying on the carbon tax for redistribution more beneficial. Conversely, progressively lowering $\mathcal{T}'_x$ reduces the wedge from income taxation. Mechanically, this means that positive deviations from the Pigouvian level of the carbon tax are amplified, and the negative deviations are dampened.

2.11 Level of the Optimal Taxes

A very similar approach can be used to derive conditions characterizing the optimal levels of the externality and income taxes. The fact that a reform to each tax schedule is not by itself distribution neutral means that they involve trade-offs between individuals, and therefore marginal social welfare weights. We follow Diamond (1975) and Ferey et al. (2024) in defining augmented welfare weights:

$$g(w) = \gamma(w)\frac{u_c(c(w), x(w), z(w) \mid w)}{\lambda} + \left[\mathcal{T}_x'\left(x\left(w\right)\right) - \frac{\mathcal{D}'\left(\bar{x}\right)}{\lambda}\right]\left.\frac{\partial \hat{x}(w)}{\partial I}\right|_{z=z(w)} + \left\{\mathcal{T}_z'(z(w)) + \left[\mathcal{T}_{x(w)}'(x\left(w\right)) - \frac{\mathcal{D}'\left(\bar{x}\right)}{\lambda}\right] x_{\text{inc}}'(z(w))\right\}\frac{\partial z(w)}{\partial I}, \tag{33}$$

where $\frac{\partial x(w)}{\partial I}$ and $\frac{\partial z(w)}{\partial I}$ are the effects of additional disposable income on consumption of x and income; and λ is the multiplier on the government's budget constraint. These augmented weights incorporate both the direct social value of increasing the consumption of an individual of type w, and indirect impacts on social welfare from income effects.

The resulting optimal tax conditions for the two schedules are stated in Proposition 2. We defer the derivation of these conditions to Appendix B.

Proposition 2. *The optimal tax on the externality-producing good must satisfy Equation 34:*

$$\frac{\mathcal{T}_x'\left(\hat{x}\left(z\right)\right) - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + \mathcal{T}_x'\left(\hat{x}\left(z\right)\right)} = \eta_{x,z}^{\textit{Taste}}\left(z\right)\frac{1}{\varepsilon_{x|z}\left(z\right)}\frac{1 - H_z\left(z\right)}{zh_z\left(z\right)}\left(1 - \bar{g}_+\left(z\right)\right), \tag{34}$$

and the optimal tax on income must satisfy Equation 35:

$$\frac{\mathcal{T}_z'\left(z\right) + \left[\mathcal{T}_x'\left(x\left(\hat{w}\left(z\right)\right)\right) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x_{inc}'(z)}{1 - \mathcal{T}_z'\left(z\right)} = \frac{1}{\varepsilon_z\left(z\right)}\frac{1 - H_z\left(z\right)}{zh_z\left(z\right)}\left(1 - \bar{g}_+\left(z\right)\right), \tag{35}$$

where $\bar{g}_+\left(z\right)$ is the average marginal social welfare weight of agents with income above z.

Equation 34 closely resembles the standard "ABC" expression for the optimal non-linear income tax for a redistributive social planner (Mirrlees, 1971). Part of the intuition is therefore familiar. A higher marginal tax rate at $\hat{x}\left(z\right)$ is optimal if: redistribution to agents with incomes below z is highly valued (i.e. if the average welfare weight above z, $\bar{g}_+\left(z\right)$, is relatively small); people do not respond much to the carbon tax (i.e. $\varepsilon_{x|z}\left(z\right)$ is small); and/or a large amount of revenue can be raised relative to the amount of income distorted by a local increase in the marginal tax rate ($1 - H_z(z)$ large relative to $zh_z(z)$).

There are two key differences from the standard optimal tax condition. The first is that there is a Pigouvian correction, which additively adjusts the marginal tax to take into account the marginal damage on others of consuming x, $\mathcal{D}'(\bar{x})$. The second is the presence of the taste elasticity, $\eta_{x,z}^{\text{Taste}}(z)$, which captures the same intuition as above.

Equation 35 is almost identical to the usual "ABC" formula. A higher marginal tax rate is optimal if redistribution is valued ($\bar{g}_+(z)$ is large), people do not respond much to the income tax ($\varepsilon_z(z)$ small), or a large amount of revenue can be raised relative to the amount of income distorted by a local increase in the marginal tax rate ($1 - H_z(z)$ large relative to $zh_z(z)$). The only difference from the standard intuition is the presence of $\left[\mathcal{T}_x'\left(x\left(\hat{w}\left(z\right)\right)\right) - \mathcal{D}'\left(\bar{x}\right)/\lambda\right] x_{\text{inc}}'(z)$ on the left hand side. As before, this term captures the intuition that marginal expenditure on the taxed good, x is socially beneficial if the carbon tax is set higher than the Pigouvian level; or socially costly otherwise.

2.12 Restricting to Linear Taxation

Some version of a non-linear tax on carbon may well be feasible in practice. Indeed, there are many real-life examples of non-linear pricing and taxation schemes for electricity, road usage, and vehicle registration fees. However, this may not be true in all settings or for all externality-generating goods. We therefore also consider the optimal linear tax on the externality-generating good, and the extent to which this deviates from the Pigouvian prescription of setting it to the level of marginal damage.

If the tax on the externality-generating good must be linear ($\mathcal{T}_x'\left(\hat{x}\left(z\right)\right) = t_x$), the Pareto efficiency condition for the linear tax rate is similar to the condition for the non-linear corrective tax (Equation 30). However, it averages over all levels of income z and consumption of x. Specifically, the optimal linear tax is characterized by Proposition 2.

Proposition 3. *If the income and commodity taxes are optimal and the corrective tax is restricted to being linear, they must satisfy equation 36:*

$$t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda} = \frac{\mathbb{E}_z\left[RR\left(z\right)\mathcal{T}_z'\left(z\right)\right]}{1 - \mathbb{E}_z\left[RR\left(z\right)x_{inc}'(z)\right]} \tag{36}$$

where $\mathbb{E}_z$ is the expectation over the income distribution, and:

$$RR\left(z\right) \equiv \eta_{x,z}^{Taste}\left(z\right)\frac{\varepsilon_z\left(z\right)\hat{x}\left(z\right)}{E_z\left[\varepsilon_{x|z}\left(z\right)\hat{x}\left(z\right)\right]}\left(\frac{1+t_x}{1-\mathcal{T}_z'\left(z\right)}\right). \tag{37}$$

and $x_{inc}'(z) = \frac{\partial x(\hat{w}(z);z)}{\partial z}$ is the causal effect of income on consumption of x. If this condition is not satisfied, there exists a simultaneous reform to both taxes that yields a Pareto gain.

The derivation is straightforward. For a linear dirty goods tax, the reforms we can consider are restricted to $\tau_x(\hat{x}(z)) \equiv \hat{x}(z)$. Plugging this into the equation for the welfare impact (equation 25) and setting it to zero gives us equation 36.

3 Multidimensional Heterogeneity

We now extend our analysis to allow for multidimensional heterogeneity. In Section 2, we assumed that heterogeneity was unidimensional and that all agents with the same taxable income z would make identical consumption choices. In practice, households with the same income may differ in their preferences for carbon-intensive goods due to factors such as geographic location, household composition, or idiosyncratic tastes. Here, we allow for such within-income heterogeneity.

Formally, we assume each agent has some type $(w, \theta) \in \mathbb{R}_{++} \times \Theta$ and that type (w, θ) agents make choices to maximize some utility function $u(c, x, z; w, \theta)$. To simplify the discussion, we largely restrict our attention here to the case of a linear tax.

3.1 Regularity Assumptions

We largely maintain the assumptions for the univariate case in Section 2.4, which ensure that choices of $z(w, \theta)$ and $x(w, \theta)$ are unique and respond smoothly to reforms. However, further assumptions are required in the multivariate case. Consistent with Jacquet and Lehmann (2021), we assume that the mapping $w \mapsto z(w, \theta)$ is strictly monotone, which ensures that the conditional inverse $\hat{w}(z; \theta)$ is well-defined. For all z, we assume that the conditional mean $\overline{x}(z) = \mathbb{E}(x \mid z)$ exists and is continuously differentiable.

3.2 Vertically Neutral Reforms

In the multidimensional case, it is not generally possible to construct a Pareto gain via a joint perturbation to the income and corrective taxes, because some individuals at any given level of income may be disproportionately intense consumers of the externality-generating good. This type of within-income heterogeneity has been studied empirically by Pizer and Sexton (2019), Cronin et al. (2019), and Rausch et al. (2011). Sallee (2022) discusses the potential for true Pareto gains in light of this heterogeneity.

It is still possible to construct a reform in the same vein as Section 2, which ensures that there is no mechanical redistribution between people with different levels of income. This allows us to derive a similar necessary condition for the tax system to be optimal. Formally, we consider a joint, infinitesimal reform that is distribution neutral on average at each

income level z: what we call a *vertically neutral* tax reform. Any pair of reform functions τ_x and τ_z is vertically neutral if, for each income level z, we have:

$$\tau_z(z) = -\mathbb{E}[\tau_x(x(w,\theta))|z(w,\theta) = z]. \tag{38}$$

In the linear tax case, this is simply $\tau_z(z) = -\bar{x}(z)$, where $\bar{x}(z) \equiv \mathbb{E}[x(w,\theta)|z(w,\theta) = z]$. The implied reform to marginal income tax rates is then:

$$\begin{aligned}\tau_z'(z) &= -\bar{x}'(z), \\ &= -\frac{\mathrm{d}\mathbb{E}[x(w,\theta)|z(w,\theta) = z]}{\mathrm{d}z}.\end{aligned}$$

3.3 Effects of Vertically Neutral Reforms

The effects of a vertically neutral reform with multidimensional heterogeneity largely parallel the unidimensional case. However, there are new effects that stem from hetereogenity in the impact of a reform on different agents with the same income level.

Let $\hat{x}(z,\theta) \equiv x(\hat{w}(z;\theta),\theta)$ be the dirty good consumption of a type θ agent that has chosen taxable income z, with $\hat{w}(z;\theta)$ being the inverse of $z(w,\theta)$, holding θ constant. Alternatively, we can express the consumption level of this agent as:

$$\hat{x}(z,\theta) = x(\hat{w}(z;\theta),\theta;z) \tag{39}$$

where $x(w,\theta;z)$ is the amount of x that type (w,θ) would consume if they had taxable income z (whether or not this is the level of income they actual choose).

As in Section 2, a vertically neutral reform induces three types of compensated behavioral responses. First, a reform which increases the linear commodity tax rate generates the following first-order compensated response of dirty good consumption:

$$\left.\frac{\partial\hat{x}(z,\theta)}{\partial\kappa}\right|_{z,\,\kappa=0} = -\frac{\varepsilon_{x|z}(z)\hat{x}(z,\theta)}{1+t_x}, \tag{40}$$

for agents of type θ with taxable income z.

Next, we turn to the effect of the vertically neutral reform on the first stage of the agent's problem, in which they choose which level of income to earn. A type θ agent with income z sees their effective marginal income tax rate change by $-\kappa x'_{\text{het}}(z;\theta)$, where:

$$x'_{\text{het}}(z;\theta) \equiv \bar{x}'(z) - x'_{\text{inc}}(z;\theta). \tag{41}$$

The interpretation of $x'_{\text{het}}(z;\theta)$ is very similar to the unidimensional case. Fixing some

income level z, it is the difference between the cross-sectional relationship between income and the average level of dirty good consumption, and the causal effect of additional income on consumption of x for agents of type θ.

This change in the incentive to earn income leads to a potentially type-specific compensated taxable income response of:

$$\left.\frac{\partial z\left(\hat{w}\left(z;\theta\right),\theta\right)}{\partial \kappa}\right|_{\kappa=0}^{c} = \frac{z\varepsilon_z(z;\theta)}{1-\mathcal{T}_z'(z)} x'_{\text{het}}(z;\theta) \tag{42}$$

where $\varepsilon_z(z;\theta)$ is the elasticity of taxable income for an agent of type (w,θ). Also as before, this change in income induces a change in dirty good consumption:

$$\left.\frac{\partial \hat{x}(z,\theta)}{\partial z}\right|_{\kappa=0} \left.\frac{\partial z\left(\hat{w}\left(z;\theta\right),\theta\right)}{\partial \kappa}\right|_{\kappa=0}^{c} = x'_{\text{inc}}(z;\theta)\frac{z\varepsilon_z(z;\theta)}{1-\mathcal{T}_z'(z)} x'_{\text{het}}(z;\theta). \tag{43}$$

Unlike a distribution neutral reform in the unidimensional case, there is redistribution between agents from this reform, because it is only *vertically* neutral. Unlike the distribution neutral case, some agents are left unable to afford their original bundle, while others can now afford strictly more than before.

To be more precise, the impact of our vertically neutral reform on the virtual income of a type θ agent with taxable income z is $\left(\hat{x}(z,\theta) - \bar{x}\left(z\right)\right)\kappa$. The impact on social welfare is equal to this change in income, scaled by the marginal social welfare weight of the affected agent, $g(z,\theta)$. These welfare weights are defined analogously to the unidimensional case:

$$\begin{aligned} g\left(z,\theta\right) \equiv \gamma(\hat{w}(z,\theta);\theta)\frac{u_c(\hat{c}(z,\theta),\hat{x}(z,\theta),z;\hat{w}(z,\theta),\theta)}{\lambda} + \left[\mathcal{T}_x'\left(\hat{x}(z,\theta)\right) - \frac{\mathcal{D}'\left(\bar{x}\right)}{\lambda}\right] \left.\frac{\partial \hat{x}(z,\theta)}{\partial I}\right|_z \\ + \left\{\mathcal{T}_z'(z) + \left[\mathcal{T}'_{x(w)}(\hat{x}(z,\theta)) - \frac{\mathcal{D}'\left(\bar{x}\right)}{\lambda}\right] x'_{\text{inc}}(z,\theta)\right\} \frac{\partial z(\hat{w}(z,\theta),\theta)}{\partial I}, \end{aligned}$$

Integrating across all levels of income and all agent types within each level of income, we get the following expression:

$$\begin{aligned} \left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = & \underbrace{\int \mathbb{E}\left[\left(x(z;\theta) - \bar{x}(z)\right)g(z;\theta) \mid z\right] h_z(z)\, dz}_{\text{Welfare effect from mechanical redistribution}} \\ & + \underbrace{\int \mathbb{E}\left[\Delta\boldsymbol{\mathcal{W}}^{x|z}(z;\theta) + \Delta\boldsymbol{\mathcal{W}}^{z}(z;\theta) + \Delta\boldsymbol{\mathcal{W}}^{z\to x}(z;\theta) \mid z\right] h_z(z)\, dz}_{\text{Welfare effects from compensated behavioral responses}}, \end{aligned} \tag{44}$$

where the three behavioral impacts conceptually mirror the unidimensional case.

As before, $\Delta\mathcal{W}^z$ captures the impact on government revenue from taxable income choices. The second term, $\Delta\mathcal{W}^{x|z}$, is the impact of changes in consumption of x that are directly caused by the tax reform. Finally, $\Delta\mathcal{W}^{z\to x}$ is the effect of further changes in consumption of x that are induced indirectly by changes in income.

$$\Delta\mathcal{W}^z(z;\theta) = \mathcal{T}_z'(z)\frac{\varepsilon_z(z;\theta)z}{1-\mathcal{T}_z'(z)}x'_{\text{het}}(z;\theta) \tag{45}$$

$$\Delta\mathcal{W}^{x|z}(z;\theta) = -\left(t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right)\frac{\varepsilon_{x|z}(z;\theta)x(z;\theta)}{1+t_x} \tag{46}$$

$$\Delta\mathcal{W}^{z\to x}(z;\theta) = \left(t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right)x'_{\text{inc}}(z;\theta)\frac{\varepsilon_z(z;\theta)z}{1-\mathcal{T}_z'(z)}x'_{\text{het}}(z;\theta) \tag{47}$$

3.4 Heterogeneity in Behavioral Responses

The presence of mechanical redistribution is not the only difference between equation 44 and its unidimensional counterpart (equation 25). Variation in the behavioral responses of different agents with the same level of income is also important.

To understand this, consider the income tax revenue that results from a marginal increase in the dirty goods tax rate by $\mathrm{d}t_x$, along with the compensating change to the income tax. This is $\mathbb{E}\left[\Delta\mathcal{W}^z(z;\theta) \mid z\right]$, which is the expectation of the impact at z across agents with different values of θ. Recall that this behavioral response arises from an agent anticipating the impact of a change in the tax rate on x on their incentive to earn additional income. For an agent of type θ and income z, the net impact of the vertically neutral reform is to change the agent's effective marginal income tax rate by $x'_{\text{het}}(z;\theta)\mathrm{d}t_x$. Clearly, this varies across agents with different types to the extent that there is heterogeneity in the causal income effect, $x'_{\text{inc}}(z;\theta)$. In addition, there may be heterogeneity in the elasticity of taxable income, $\varepsilon_z(z;\theta)$, across types with the same income.

Heterogeneity in these behavioral responses means that we cannot simply replace each behavioral parameter with its mean value. Instead, we need to take into account the potential for $\varepsilon_z(z;\theta)$ and $x'_{\text{inc}}(z;\theta)$ to be correlated at a given level of income. This leads to the following equation for $\mathbb{E}\left[\Delta\mathcal{W}^z(z;\theta) \mid z\right]$:

$$\mathbb{E}\left[\Delta\mathcal{W}^z(z;\theta) \mid z\right] = \mathcal{T}_z'(z)\cdot\frac{\bar{\varepsilon}_z(z)z}{1-\mathcal{T}_z'(z)}\cdot\bar{x}'_{\text{het}}(z) + \frac{\mathcal{T}_z'(z)z}{1-\mathcal{T}_z'(z)}\mathbb{C}\left[\varepsilon_z(z;\theta), x'_{\text{het}}(z;\theta)|z\right] \tag{48}$$

where $\bar{x}'_{\text{het}}(z)$ is the average value of $x'_{\text{het}}(z;\theta)$ for agents at income level z and $\mathbb{C}$ indicates a covariance. Equation 48 highlights that ignoring heterogeneity and using $\bar{x}'_{\text{het}}(z)$ to measure this effect may be valid, but only if $\varepsilon_z(z;\theta)$ and $x'_{\text{inc}}(z;\theta)$ are uncorrelated.

Turning to the expectation of the second effect, $\mathbb{E}\left[\Delta\boldsymbol{\mathcal{W}}^{z\to x}(z;\theta)\mid z\right]$, we can again rewrite this as a function of the average values of each statistic and their covariances:

$$\begin{aligned}\mathbb{E}\left[\Delta\mathcal{W}^{z\to x}(z;\theta)\mid z\right] = &\left(\frac{t_x-\frac{\mathcal{D}'(\bar{x})}{\lambda}}{1-\mathcal{T}_z'(z)}\right) z\bar{\varepsilon}_z(z)\bar{x}'_{\text{inc}}(z)\bar{x}'_{\text{het}}(z) \\ &+\left(\frac{t_x-\frac{\mathcal{D}'(\bar{x})}{\lambda}}{1-\mathcal{T}_z'(z)}\right) z\bar{\varepsilon}_z(z)\mathbb{C}\left[x'_{\text{inc}}(z;\theta),x'_{\text{het}}(z;\theta)|z\right] \\ &+\left(\frac{t_x-\frac{\mathcal{D}'(\bar{x})}{\lambda}}{1-\mathcal{T}_z'(z)}\right) z\,\mathbb{C}\left[\varepsilon_z(z;\theta),x'_{\text{inc}}(z;\theta)x'_{\text{het}}(z;\theta)|z\right].\end{aligned} \tag{49}$$

The first term in this expression is simply the product of mean behavioral responses, which would also be present in a model without multidimensional heterogeneity.

The second and third terms involve covariances. We can understand these as arising from the two steps that lead from the tax reform to indirect changes in consumption of x. First, the income response for an individual of type (w,θ) is proportional to $\varepsilon_z(z;\theta)\times x'_{\text{het}}(z;\theta)$. Second, fixing this income response, the income effect on consumption of x is proportional to $x'_{\text{inc}}(z;\theta)$. Not only is there potential for these two components of the response to be correlated, but the definition of $x'_{\text{het}}(z;\theta)$ in terms of $x'_{\text{inc}}(z;\theta)$ means that there is a *mechanical* relationship between them. In fact, this mechanical connection means that the first covariance collapses to the negative variance of causal income effects at z:

$$\mathbb{C}\left[x'_{\text{inc}}(z;\theta),x'_{\text{het}}(z;\theta)\mid z\right] = -\mathbb{V}\left[x'_{\text{inc}}(z;\theta)\mid z\right]. \tag{50}$$

This is because agents with a larger $x'_{\text{het}}(z;\theta)$ experience a larger change in their effective marginal income tax rates, and also have a smaller consumption response to any given increase in their incomes.

The second covariance makes a further adjustment to account for the possibility that the product of the causal income and heterogeneity effects $x'_{\text{inc}}(z;\theta)x'_{\text{het}}(z;\theta)$ may be correlated with the elasticity of taxable income, $\varepsilon_z(z;\theta)$. This highlights a useful and tractable special case: If we assume that $\varepsilon_z(z;\theta)$ is constant within income levels, then the third term is zero. However, the first covariance remains. This highlights that heterogeneity matters even without correlation between the various statistics. Only with the further restriction that there is no within-income heterogeneity at all, so that $\text{Var}(x'_{\text{inc}}(z;\theta)\mid z)=0$, does equation 50 become isomorphic to the unidimensional case. In this sense, multidimensional heterogeneity leads to robust attenuation of the indirect consumption effect.

3.5 Optimal Linear Taxation: The Multidimensional Case

Our final step is to put together the impacts on social welfare from a vertically neutral reform, and present an optimal tax condition, harnessing the fact that the welfare effect of any vertically neutral reform (equation 44) must equal zero at the optimum.

By construction, the reforms we have considered do not redistribute vertically. For simplicity, we also assume throughout the analysis below that welfare weights are uncorrelated with consumption of the externality-generating good conditional on income, at least on average. That is, we assume that:

$$\mathbb{E}\left[\mathbb{C}\left(g\left(w,\theta\right),x\left(w,\theta\right)|z\left(w,\theta\right)=z\right)\right]=0. \tag{51}$$

In practice, the sign of this covariance depends on whether more intense consumers of carbon-intensive goods have higher or lower welfare weights conditional on income.

Assuming that the effect on welfare of redistribution within income levels is negligible on average, rearranging equation 44 yields the following characterization of the optimal linear tax on the externality-generating good.

Proposition 4. *Assume that welfare weights are uncorrelated with consumption of the externality-generating good conditional on income:*

$$\mathbb{E}\left[\mathbb{C}(g(z;\theta),x(z;\theta)\mid z)\right]=0. \tag{52}$$

Then the optimal linear tax on the externality-generating good satisfies:

$$t_x-\frac{\mathcal{D}'(\bar{x})}{\lambda}=\frac{\mathbb{E}_z\left[\mathbb{E}_\theta\left[RR(z;\theta)\mathcal{T}_z'(z)\mid z\right]\right]}{1-\mathbb{E}_z\left[\mathbb{E}_\theta\left[RR(z;\theta)x'_{inc}(z;\theta)\mid z\right]\right]}, \tag{53}$$

where:

$$RR(z;\theta)\equiv\eta_{x,z}^{Taste}(z;\theta)\left(\frac{\varepsilon_z(z;\theta)\bar{x}(z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]}\right)\left(\frac{1+t_x}{1-\mathcal{T}_z'(z)}\right), \tag{54}$$

and the type-specific taste elasticity is:

$$\eta_{x,z}^{Taste}(z;\theta)\equiv\frac{zx'_{het}(z;\theta)}{\bar{x}(z)}=\frac{z(\bar{x}'(z)-x'_{inc}(z;\theta))}{\bar{x}(z)}. \tag{55}$$

Equation (53) closely parallels the optimal linear tax condition from the unidimensional case (Proposition 3), but with expectations taken over both income levels and θ types within each income level. The taste elasticity $\eta_{x,z}^{\text{Taste}}(z;\theta)$ now measures how a type θ individual's causal income effect diverges from the overall cross-sectional relationship between income and consumption.

3.6 Decomposing the Multidimensional Optimal Tax Condition

To guide our empirical implementation, we can decompose the expectations in equation 53 into average effects and within-income covariance terms. This follows the same logic as when we decomposed the individual behavioral responses in Section 3.4.

First, the numerator of the right-hand side of equation 53 can be written as $\mathcal{T}'_z(z)/[1-\mathcal{T}'_z(z)] \times \mathbf{N}$ where:

$$\mathbf{N} = \mathbb{E}_z\left[\bar{\eta}^{\text{Taste}}_{x,z}(z)\frac{\bar{x}(z)\bar{\varepsilon}_z(z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]}\right] - \mathbb{E}_z\left[\frac{z\,\mathbb{C}(\varepsilon_z(z;\theta), x'_{\text{inc}}(z;\theta)\mid z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]}\right].$$

where $\bar{\eta}^{\text{Taste}}_{x,z}(z)$ and $\bar{\varepsilon}_z(z)$ are averages across θ types at income z.

The first term of this equation involves average values of each sufficient statistic at each income level. To use this first term while ignoring the second would be equivalent to naively applying the unidimensional formula with the average values of each sufficient statistic. The second term is a correction involving the within-income covariance between the elasticity of taxable income and the causal income effect on x.

Section 3.4 explains where this covariance term comes from. We now provide complementary intuition. Consider a population where, conditional on z, agents with higher ε_z also have higher x'_{inc} (and thus lower $\eta^{\text{Taste}}_{x,z}$). Such agents have two key characteristics. First, they respond strongly to income tax changes. Second, the tagging value of carbon is low for such agents. Thus, if the planner could condition taxes on θ, they would set a smaller deviation from Pigouvian taxation for these agents. In reality, the planner can only condition on z. However, the optimal uniform policy must account for the fact that the most responsive agents are precisely those for whom tagging is least valuable.

Second, the denominator of the right-hand side of equation 53 can be written as $1-[1+t_x]/[1-\mathcal{T}'_z(z)] \times \mathbf{D}$ where:

$$\mathbf{D} = \mathbb{E}_z\left[\bar{\eta}^{\text{Taste}}_{x,z}(z)\frac{\bar{x}(z)\mathbb{E}[x'_{\text{inc}}(z;\theta)\varepsilon_z(z;\theta)|z]}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]}\right] - \mathbb{E}_z\left[\frac{z\,\mathbb{C}(x'_{\text{inc}}(z;\theta), x'_{\text{inc}}(z;\theta)\varepsilon_z(z;\theta)|z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]}\right].$$

As Section 3.4 explains, this covariance comes from correlation between the tagging value of the externality-generating good, and indirectly induced consumption of that good. Suppose that among agents at a given level of income, two things go together: (1) a high tagging value of x, as reflected by a low value of x'_{inc}; and (2) a small indirect response of x because x'_{inc} is small, ε_z is small, or both. Then this covariance term will be positive, which increases the denominator and dampens any optimal deviation from the Pigouvian level

of the optimal tax. Intuitively, the planner must take into account that agents for whom tagging is most valuable are precisely those for whom any efficiency gains from a reform lead indirectly to more counterproductive externality-generating consumption.

3.7 A Tractable Case: Constant Elasticity of Taxable Income

In our empirical implementation, we assume that the elasticity of taxable income does not vary with θ at each income level, so that $\varepsilon_z(z;\theta) = \bar{\varepsilon}_z(z)$ for all θ. Proposition 1 presents the optimal tax condition for this empirically implementable case. In addition to the average value of each statistic, the condition features the variance of income effects, x'_{inc}. Holding all other empirical objects constant, this extra term always attenuates any deviation of the optimal tax on x from the Pigouvian level. Only in the special case in which there is no within-income heterogeneity, so that $\text{Var}(x'_{\text{inc}}(z;\theta) \mid z) = 0$, does equation (56) simplify to the unidimensional optimal linear tax condition in Proposition 3.

Corollary 1. If $\varepsilon_z(z;\theta) = \bar{\varepsilon}_z(z)$ for all θ at each income level z, the optimal linear tax satisfies:

$$t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda} = \frac{\mathbb{E}_z\left[\overline{\text{RR}}(z)\mathcal{T}'_z(z)\right]}{1 - \mathbb{E}_z\left[\overline{\text{RR}}(z)\bar{x}'_{\text{inc}}(z)\right] + \mathbb{E}_z\left[\widetilde{\text{RR}}(z)\mathbb{V}(x'_{\text{inc}}(z;\theta) \mid z)\right]}, \tag{56}$$

where:

$$\overline{\text{RR}}(z) \equiv \bar{\eta}^{\text{Taste}}_{x,z}(z)\left(\frac{\bar{\varepsilon}_z(z)\bar{x}(z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]}\right)\left(\frac{1+t_x}{1-\mathcal{T}'_z(z)}\right), \tag{57}$$

$$\widetilde{\text{RR}}(z) \equiv \frac{\bar{\varepsilon}_z(z)z}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]}\frac{1+t_x}{1-\mathcal{T}'_z(z)}, \tag{58}$$

and $\bar{x}'_{\text{inc}}(z) \equiv \mathbb{E}[x'_{\text{inc}}(z;\theta) \mid z]$ is the average causal income effect at z.

Corollary 1 identifies the sufficient statistics that are needed to characterize optimal policy with multidimensional heterogeneity, subject to our restriction that the elasticity of taxable income is constant. First, we require the average value of several statistics at each income level: the cross-sectional slope, $\bar{x}'(z)$; the causal income effect, $\bar{x}'_{\text{inc}}(z)$; the elasticity of taxable income, $\bar{\varepsilon}_z(z)$; and the weighted average demand elasticity, $\varepsilon_{x|z}(z;\theta)$. Second, we need to measure the within-income variance in income effects, $\text{Var}(x'_{\text{inc}}(z;\theta) \mid z)$.

4 Empirical Application

Equations 29 and 36 allow us to quantify the optimal carbon tax for the United States, taking into account distributional concerns. For a given marginal social cost of carbon, we show how the Pareto-efficient carbon tax may deviate from the standard Pigouvian

correction under both nonlinear and linear commodity tax systems, presenting estimates for a range of different assumptions regarding the compensated elasticities of taxable income and dirty good consumption. We initially consider the simpler unidimensional case, and then we incorporate multidimensional heterogeneity.

We start with our empirical quantification of taste heterogeneity, $\eta_{x,z}^{\text{Taste}}(z)$. This is the focus of Section 4.1. In Section 4.2, we then describe how we estimate and calibrate the other key objects which jointly determine the optimal carbon tax.

4.1 Quantifying Taste Heterogeneity

To estimate the extent of taste heterogeneity, we use the decomposition of the cross-sectional relationship provided by equation 18. This requires an estimate of $\hat{x}'(z)$, which we obtain from cross-sectional variation in the consumer expenditure survey; and an estimate of x'_{inc} at each of many different points in the income distribution, which we obtain by running our own survey.

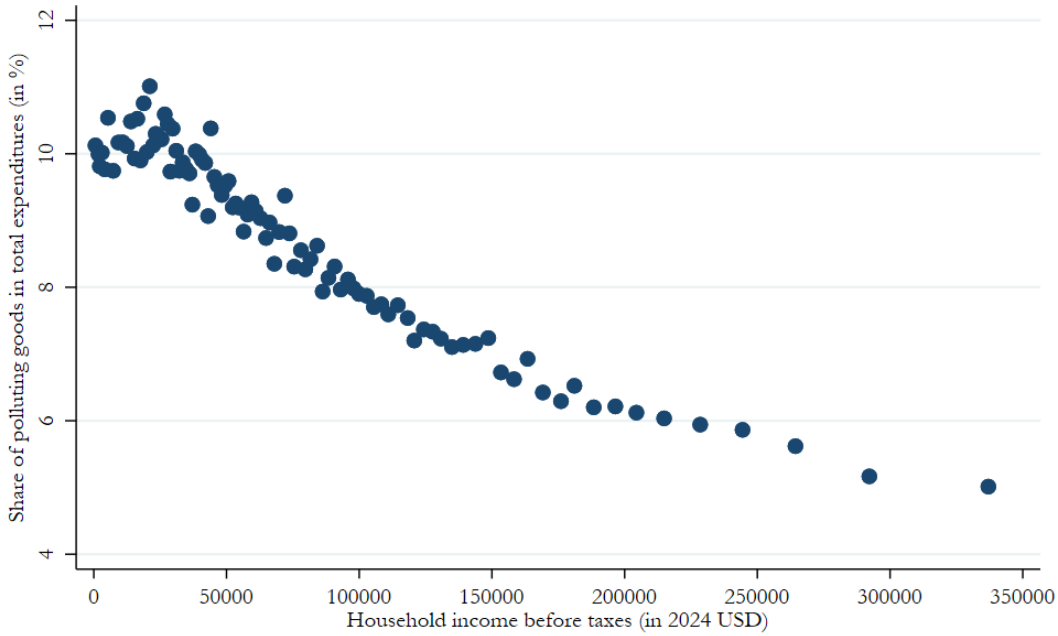
4.1.1 Measuring Cross-Sectional Variation Using The CEX

We first measure the cross-sectional relationship between taxable income and consumption of carbon-intensive goods, $\hat{x}'(z)$. We do this using the Consumer Expenditure Survey (CEX), which provides quarterly expenditure data from 2019 to 2023. Specifically, we use information from the public use microdata (PUMD) files of the Consumer Expenditure Surveys provided by the United States Bureau of Labour Statistics. We rely on the Interview survey, a rotating panel survey with approximately 6,000 observations per quarter containing detailed data on expenditures, income, and demographic characteristics of consumers in the United States. Each observation is at the household level.[4] Our sample is composed of 113,108 quarterly observations across 46,918 households. Because of the rotating nature of the panel, households may be sampled from once up to 4 consecutive times (quarters) across the sample period (2019-2023).

We compute expenditure on carbon-intensive goods as a percentage of total consumption for each quarterly household observation. We focus on two carbon-intensive expenditure categories: (i) an aggregate dirty good, and (ii) gasoline and other motor oils. Aggregate dirty good consumption is defined as the sum of electricity expenditure, gasoline and motor oil expenditure, all heating fuels expenditure, and natural gas expenditure. Appendix A.1 shows the share of consumption accounted for by these and other categories.

[4]See detailed definition of a "Consumer Unit" at: https://www.bls.gov/cex/csxfaqs.htm#qc1.

(a) Aggregate Dirty Good



(b) Gasoline

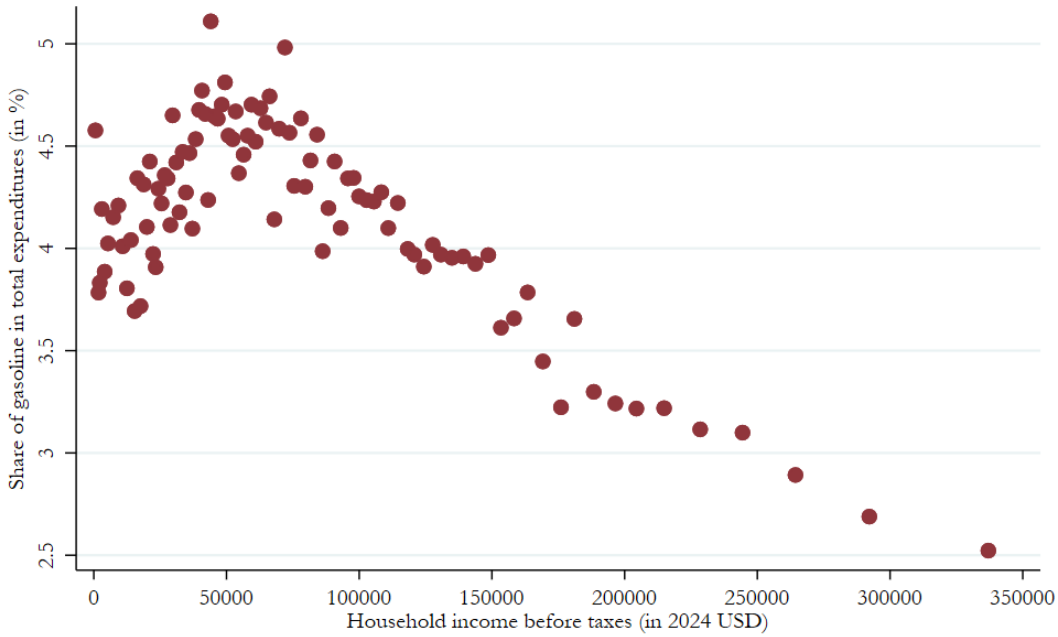


Figure 1. Expenditure shares of carbon-intensive goods across the income distribution

Note: These binscatter plots show (i) the average share (in %) of expenditure on polluting goods as a fraction of total consumption, and (ii) the corresponding average household taxable income, for each income percentile using data from the 2018-2023 Consumer Expenditure Survey quarterly waves. The "dirty good" expenditure category includes gasoline, electricity, natural gas, and heating fuels. Household taxable incomes were converted to 2024 USD using the average annual Consumer Price Index in the US.

To smooth the cross-sectional relationship between household taxable income and consumption of carbon-intensive goods, we assign each household to a year-specific income percentile (because taxable income is measured annually). We then average measures of consumption and income across quarterly observations of households within each income percentile bin, using CEX-provided survey weights to ensure the sample is representative of the US population. This binning procedure abstracts away from any within-income heterogeneity in dirty good consumption. In Section 4.3.3, we relax this restriction.

Panel (a) of Figure 1 shows the average *dirty good* expenditure as a share (in percent) of total expenditures at each point in the income distribution. Panel (b) shows the same measure for *gasoline* only.[5] Poorer households spend a much larger share of their consumption budget on carbon-intensive goods than do higher-income households. For the aggregate dirty good displayed in panel (a), the share ranges from a high of 11% at the bottom of the income distribution to a low of approximately 5% at the very top. Focusing on gasoline only, displayed in panel (b), the expenditure share ranges from 4 to 5% at the bottom of the income distribution to 2.5% at the top.

While consumption *shares* of carbon-intensive goods are generally *decreasing* with income (see Figure 1), it is important to note that *levels* are *increasing* with income. Figure 2 shows the cross-sectional relationship between quarterly consumption expenditure (in dollars) of carbon-intensive goods and household taxable income. The relationship is generally positive but flattens out at higher levels of household taxable income.

4.1.2 Measuring Income Effects Using A Hypothetical Choice Survey

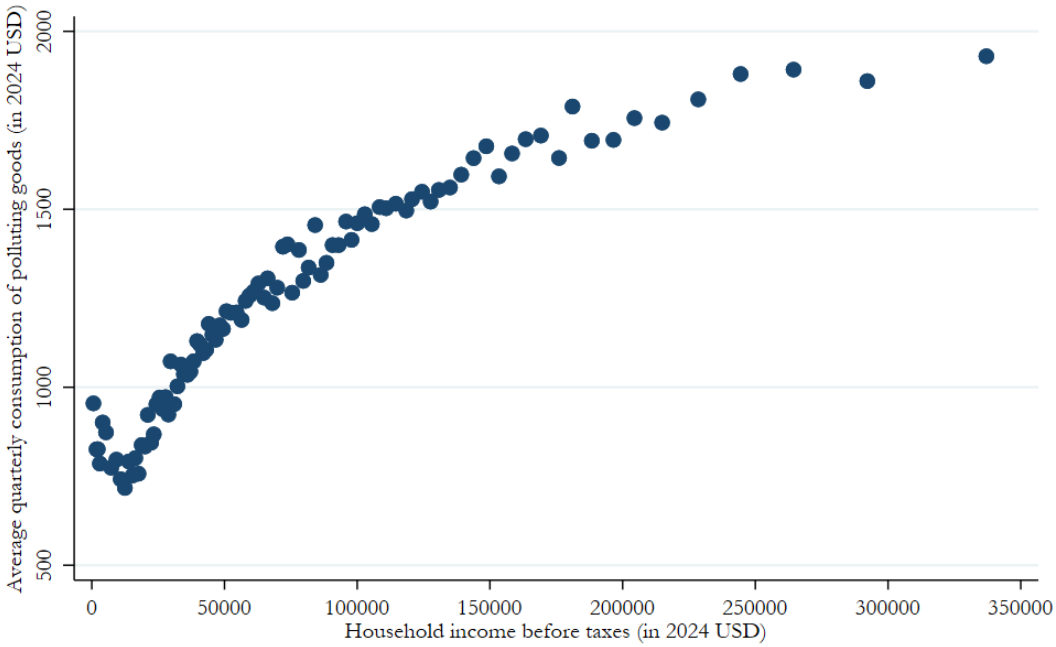
Next, we measure the causal response of additional income, x'_{inc}. This requires a survey of our own, which measures how each household's consumption would respond to a small amount of additional income. We designed and ran the survey using *Qualtrics*, and recruited participants on *Prolific Academic*.[6] The sample is comprised of 1,448 US residents aged 25 to 54, and was designed to be representative of the US population along the targeted dimensions of age, gender, and race (see Appendix Table A.4).[7]

[5] Using total expenditure rather than annual income as the denominator is preferable when studying the distributional burden of excise taxes. As Poterba (1989) explains, this is because total expenditures are typically less volatile than income and thus serve as a better proxy for lifetime income. For our analyses, using total expenditures as the denominator also has the additional benefit of abstracting away from differences in the saving rate across incomes. The main takeaway from Figure 1 remains the same if we compute shares of income rather than expenditure.

[6] The survey and our analysis of the results were pre-registered on the Open Science Framework (OSF) platform in January, 2025: `https://osf.io/swqze`.

[7] The survey was conducted in two waves, the first in February 2025, and the second in July 2025.

(a) Aggregate Dirty Good



(b) Gasoline

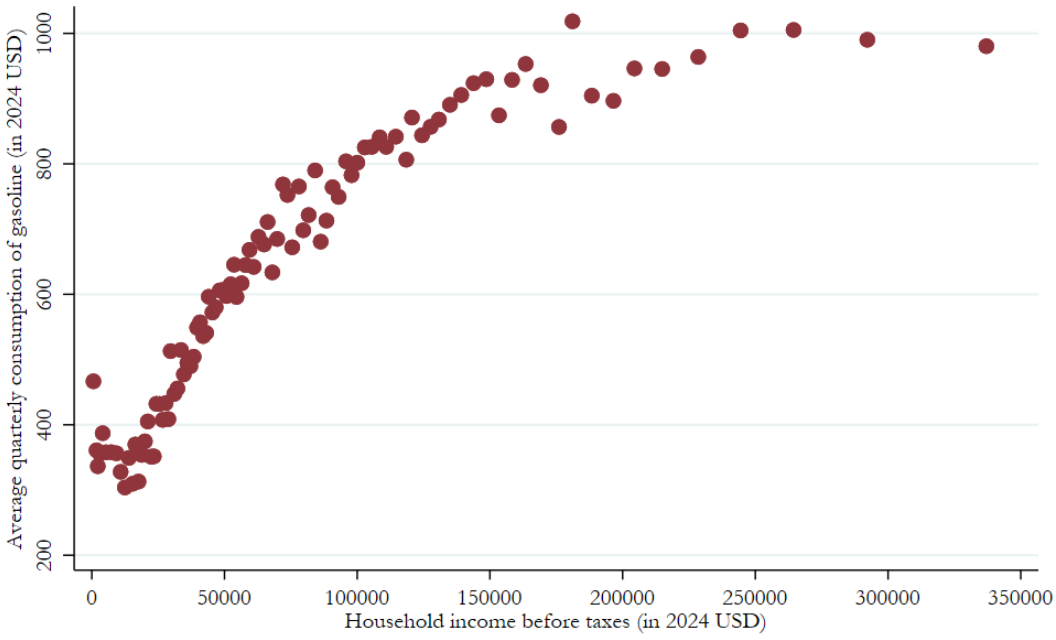


Figure 2. Quarterly consumption expenditure of carbon-intensive goods across the income distribution

Note: These binscatter plots show (i) the average quarterly expenditures on polluting goods, and (ii) the corresponding average household taxable income, for each income percentile using data from the 2018-2023 Consumer Expenditure Survey quarterly waves. The "dirty good" expenditure category includes gasoline, electricity, natural gas, and heating fuels. Household consumption expenditures and taxable incomes were converted to 2024 USD using the average annual Consumer Price Index in the US.

The primary purpose of the survey was to elicit how each household would respond to an unexpected increase in household (after-tax) income. Respondents were asked to allocate a $1,000 permanent increase in household income across an exhaustive list of expenditure categories (mirroring those from the CEX), and a residual savings bucket.[8] This elicitation procedure provides us with each respondent's marginal propensity to consume (MPC) out of a permanent increase in household income, and its breakdown between carbon-intensive goods (electricity, natural gas, heating fuels, and gasoline) and other goods. We also collected information about respondents' *status quo* household income (both before- and after-tax), and the breakdown of these expenditures.

Response Quality. We took several steps to ensure the quality of respondents, including screening out respondents based on (i) past studies' rejection rates, and (ii) the number of hours spent on the platform. As stated in our pre-registration plan, we also filter out participants who fail attention checks, speedsters, extreme outliers, and respondents who provide inconsistent or nonsensical information on before- vs. after-tax income or marginal propensity to consume (see Appendix Table A.5 for a breakdown of sample exclusions).

In addition, we have a powerful way to validate the representativeness of our survey sample. Figure A.1 plots the relationship between our respondents' reported consumption expenditures and their incomes. This relationship can be compared to that in the Consumer Expenditure Survey (Figure 1). The figures align remarkably closely, suggesting that our responses are of high quality and our respondents quite representative.

Marginal Consumption Patterns In The Survey. Figure 3 shows a binscatter plot of the relationship between the share of carbon-intensive goods in total marginal consumption, and household taxable income. We also include corresponding global polynomial fits (of order 2). Panel (a) shows the aggregate results for all of the carbon-intensive goods, while panel (b) shows the results for gasoline. In both cases, we see a declining marginal propensity to consume these goods as income rises.

Estimating Causal Income Effects. Our estimates for x'_{inc} build on our survey respondents' marginal consumption choices in response to a hypothetical increase in their incomes (see Figure 3). However, we need to make a few additional assumptions to estimate $x'_{\text{inc}}(z)$ based on the consumption response to a small permanent increase in after-tax income.

First, note that in our theoretical framework, the response of x to a pure income shock (the quantity we observe in our custom survey) is given by:

$$\frac{dx\left(\hat{w}\left(z\right);z\right)}{dI} = \eta_{x|z}(z) + \frac{\partial x\left(\hat{w}\left(z\right);z\right)}{\partial z}\,\eta_z(z), \tag{59}$$

[8] See Figure A.2 for a screenshot of the online survey page.

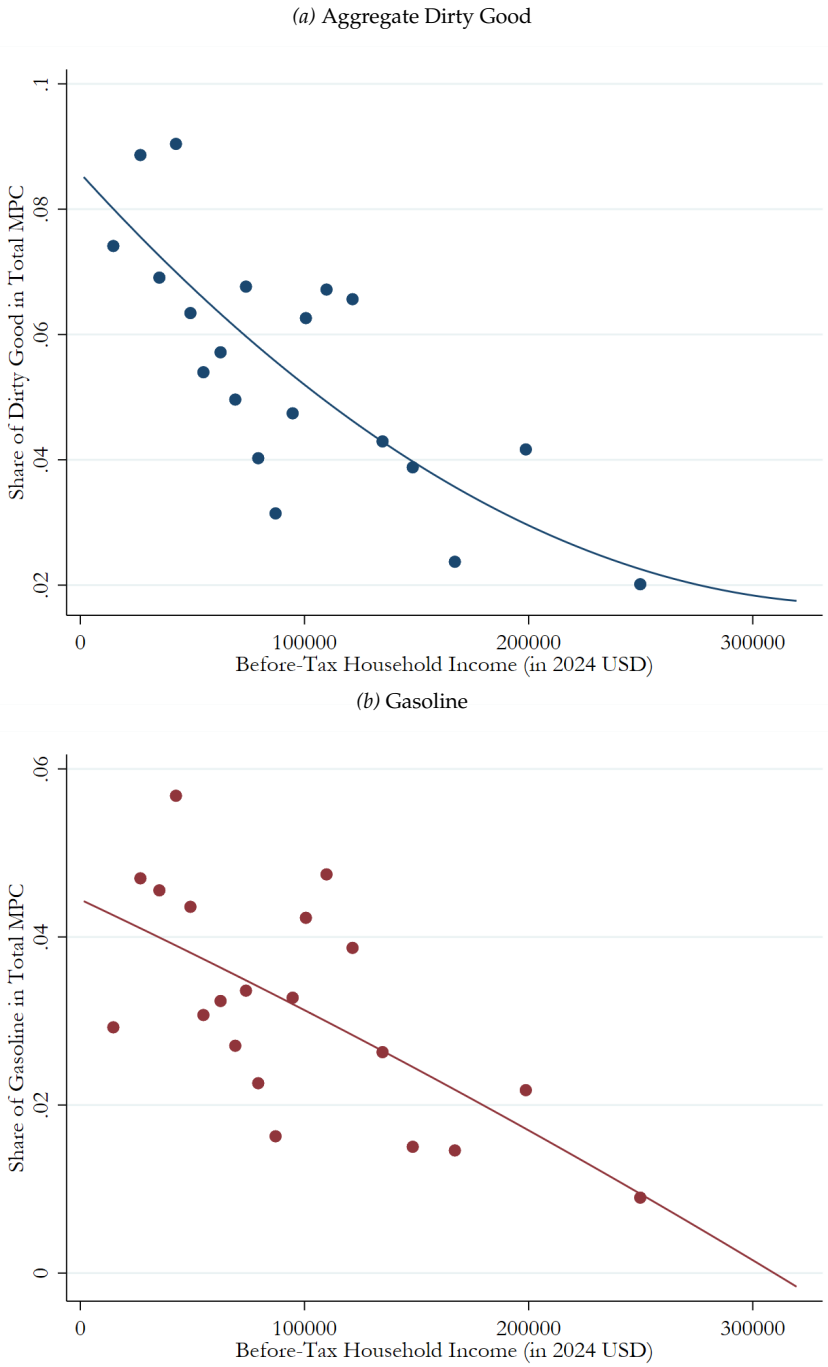


Figure 3. Share of carbon-intensive goods in total MPC across the income distribution

Note: These binscatter plots show (i) the average share of expenditure on polluting goods as a fraction of total marginal propensity to consume (MPC) out of a small and unexpected $1,000 increase in after-tax household income, and (ii) the corresponding average household taxable income, averaged across 20 bins of household taxable income (selected with *binsreg* (Cattaneo et al., 2024)), and (iii) a global polynomial of degree 2 estimated to fit the relationship between these two variables, using data from a custom representative survey of the US population ($N = 1,448$). The "dirty good" expenditure category includes gasoline, electricity, natural gas, and heating fuels.

where η_z and $\eta_{x|z}$ are income effects due to a reform that lowers the tax burden slightly, while leaving both marginal tax rates unchanged. Imposing weak separability between income and consumption choices allows us to make more progress here.[9] In this case, the causal effect of z on x only occurs via the change in disposable income I, i.e. $\partial x/\partial z = \partial I/\partial z \times \partial x/\partial I|_{z=z(w)} = (1 - T_z') \times \partial x/\partial I|_{z=z(w)}$, and we can rewrite the income effect on consumption of the dirty good holding taxable income fixed as:

$$\eta_{x|z}(z) = \frac{1}{1 - T_z'(z)} \frac{\partial x\left(\hat{w}\left(z\right); z\right)}{\partial z}. \tag{60}$$

This means the total impact of a pure income shock on consumption of x can be written as

$$\frac{dx\left(\hat{w}\left(z\right); z\right)}{dI} = \left(\frac{1}{1 - T_z'(z)} + \eta_z(z)\right) \frac{\partial x\left(\hat{w}\left(z\right); z\right)}{\partial z}. \tag{61}$$

If we further assume that the income effect on taxable income is negligible ($\eta_z \approx 0$), then

$$x'_{\text{inc}}(z) \equiv \frac{\partial x\left(\hat{w}\left(z\right); z\right)}{\partial z} = \left(1 - T_z'(z)\right) \frac{dx\left(\hat{w}\left(z\right); z\right)}{dI}. \tag{62}$$

4.1.3 Abstracting Away From Savings

Our model is a static one, and thus abstracts away from the savings decision (i.e., from consumption across multiple time periods). However, in the data, households do save a fraction of their disposable income. Therefore, we make the following two adjustments to abstract away from the savings decision. First, when constructing $\hat{x}'(z)$ with the CEX data, we use *shares* (as in Figure 1) rather than *levels* at each income percentile, and multiply these shares by the level of after-tax income. This results in a "savings-adjusted" level of consumption of the dirty good x.[10] Second, we use *shares* of total marginal expenditure (as depicted in Figure 3), i.e. the additional amount spent on carbon-intensive goods as a fraction of the total additional amount spent, rather than the dollar amount spent on carbon-intensive goods divided by \$1,000.

[9] There is limited evidence that gasoline consumption is complementary with leisure (West and Williams III, 2007), although this finding is not universal (Beznoska, 2014). If this is true for overall carbon intensity, our estimates understate the optimal tax on carbon-intensive goods. See Kaplow (2012) for further discussion.

[10] This approach to handling the savings decision effectively assumes that consumption shares remain constant over time, which is true with a utility function of the form: $U_i = u^i\left(f_i(c_1, x_1) + \beta f_i(c_2, x_2), z\right)$ where $f_i(c_t, x_t)$ is homothetic within each period $t \in \{1, 2\}$ and identical across periods. Note also that in order to have a direct dependence of z on within-period consumption preferences, we need to allow the homothetic preference parameters to depend on z (on top of their type dependence).

4.1.4 Combining Cross-Sectional and Causal Variation To Estimate Taste Heterogeneity

Our final step is to smooth the data over the income distribution, and calculate $x'_{\text{het}}(z)$ as the difference between the fitted $\hat{x}'(z)$ and the fitted $x'_{\text{inc}}(z)$. We first convert the discrete distribution (at each percentile of taxable income) of (i) taxable income z, (ii) savings-adjusted consumption of the dirty good x, and (iii) shares of total MPCs using the global polynomials from Figure 3 rescaled by the net-of-tax (income) rate to obtain $x'_{\text{inc}}(z)$ into a continuous representation by constructing a grid of equally log-spaced points. We then fit smoothing splines to the log-transformed variables, and convert them back to the original levels. Finally, we compute the derivative of the fitted cross-sectional relationship $\hat{x}(z)$ to recover $\hat{x}'(z)$ to which we can subtract the smoothed $x'_{\text{inc}}(z)$ to obtain $x'_{\text{het}}(z)$.

The results are depicted in Figure 4. For brevity, we restrict the remainder of our empirical analysis to the aggregate dirty good. However, we present results for each components Appendix A, which are very similar in their implications.

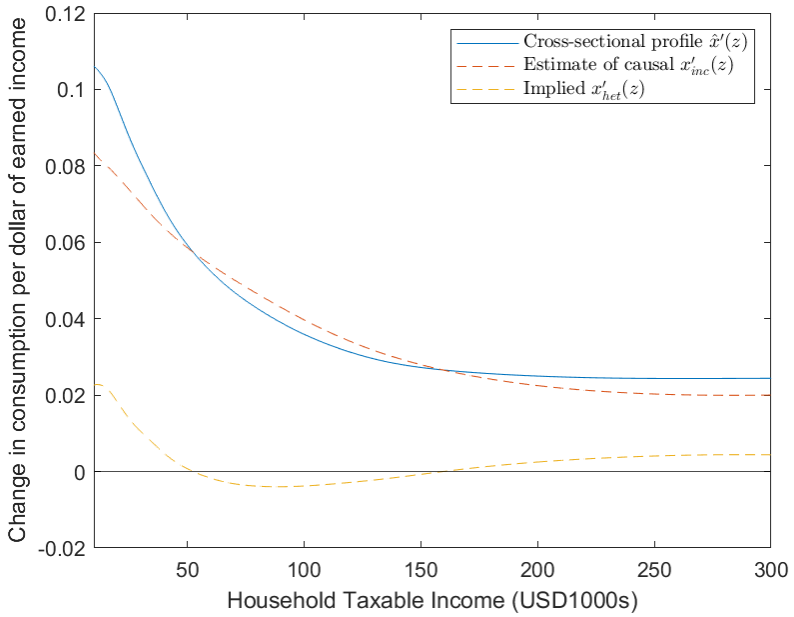


Figure 4. Decomposition of the cross-sectional profile $\hat{x}'(z)$ - Aggregate Dirty Good

Note: This figure shows how we decompose the cross-sectional relationship between income and carbon-intensive good consumption. The blue line is a smoothed estimate of the cross-sectional slope, $\hat{x}'(z)$ at each level of income. The dashed red line shows estimates of the causal effect of income on expenditure on carbon-intensive goods. The yellow dashed line subtracts the income effect from the cross-section to obtain the part of the relationship between x and z that comes from preference heterogeneity.

The blue curve in Figure 4 depicts the cross-sectional profile $\hat{x}'(z)$. It is positive, decreasing with income, but quickly flattens out for household taxable incomes between \$100,000 and \$300,000. This illustrates the increasing and concave relationship between *levels* of

consumption of carbon-intensive goods and taxable income. The dashed orange curve shows our fitted estimate of $x'_{\text{inc}}(z)$, it is decreasing throughout the income distribution. The dashed yellow curve shows the implied (residual) $x'_{\text{het}}(z)$ which determines the sign of $\eta^{\text{Taste}}_{x,z}(z)$, and in turn the shape of our optimal and Pareto-efficient nonlinear tax schedules. Our estimates of $\hat{x}'(z)$ and $x'_{\text{inc}}(z)$ are remarkably close overall. This leads to estimates of $x'_{\text{het}}(z)$ that are close to zero, which implies that differences in taste contribute very little to variation in carbon intensity over the income distribution.

It is important to note that there was no reason *a priori* to expect our estimate for the causal effect of income to closely match the cross-sectional profile across the income distribution. The first is computed as a slope using expenditures (i.e., levels) *across* households from the large CEX surveys, while the second is an individual-level response (i.e., a change) from our custom hypothetical survey. Their similarity is thus striking.

For direct application in our quantitative analysis, Figure 5 translates this $x'_{\text{het}}(z)$ into the taste elasticity that appears in Equation 34. We see that $\eta^{\text{Taste}}_{x,z}(z)$ is positive for households with low incomes (below \$52,000), negative for households with taxable income between \$52,000 and \$160,000, and then being positive and increasing at the top of the income distribution. This schedule, together with the compensated elasticities of taxable income and dirty good consumption, the marginal damage, and the marginal income tax schedule, determines the shape of the Pareto efficient nonlinear tax schedule on consumption of the carbon-intensive good (see equation 29).

4.2 Estimating and Calibrating Other Quantities

Elasticity of consumption of carbon-intensive goods. We follow an empirical approach similar to that of Li et al. (2014). We construct a panel using the CEX, with variation over time and between states in gasoline excise taxes and prices. Then we study the consumption response of gasoline to tax changes from 1996 to 2021. We estimate the uncompensated elasticity to be 0.58 (SE: 0.29) in our preferred specification with month$\times$year fixed effects and household-level controls (see Appendix A.3 for details). We thus use three values for the compensated elasticity in our tax simulations for sensitivity analysis: 0.5 for our preferred *benchmark* scenario, 0.25 for the *low* elasticity of dirty good consumption scenario, and 0.75 for the *high* elasticity of dirty good consumption scenario.

Elasticity of taxable income. We calibrate the elasticity of taxable income $\varepsilon_z(z)$ for our *benchmark* scenario to be constant at 0.33, which is the preferred estimate (on the intensive margin) from the meta-analysis of Chetty (2012). We also consider a value of 0.7 for our *high* taxable income elasticity scenario.

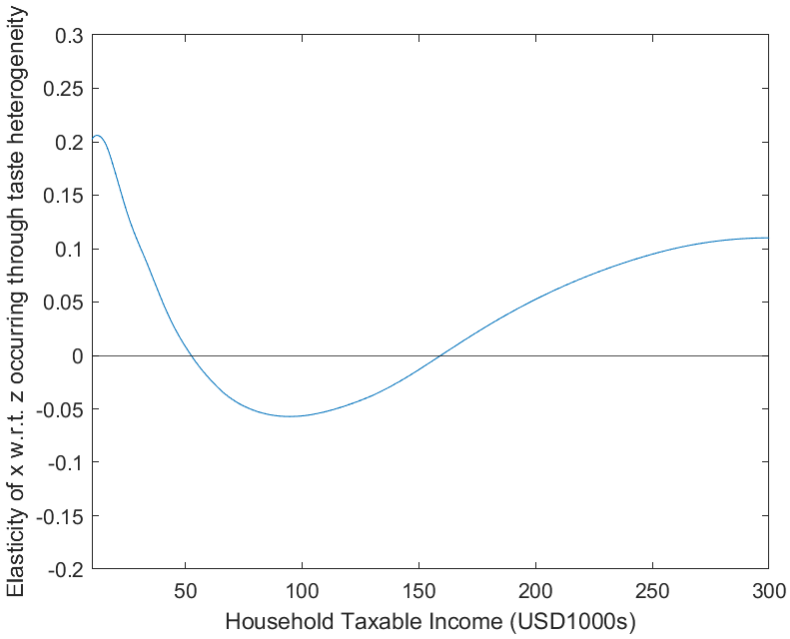


Figure 5. Elasticity of consumption of x with respect to income via taste heterogeneity only

Note: This figure plots the "taste elasticity", $\eta^{\text{Taste}}_{x,z}$. It is based on our estimates of $x'_{\text{het}}(z)$, which are shown by the dashed yellow line in Figure 4, but converted into an elasticity.

Marginal Damage From Carbon. We follow Hahn et al. (2024) and assume a Social Cost of Carbon (SCC) of $200 per ton. This is within the range of estimates provided by a recent report of the US Environmental Protection Agency (EPA, 2023). We assume that consumption of carbon-intensive goods generates an average of 2kg of CO_2 per dollar of final expenditure.[11] This results in a marginal damage of 40 cents per dollar of final expenditure, which implies a baseline Pigouvian tax of 40% in our model[12]. We note that our Pareto-efficient tax schedules are calibrated relative to this marginal damage estimate, which allows for flexible adjustments (of our results) based on alternative SCC values.

Marginal income tax rate schedule. Using the same approach as for the consumption shares described above, we compute the average before- and after-tax household income for each income percentile using data from the 2018 to 2023 CEX quarterly waves. We convert these values to 2024 USD using the average annual Consumer Price Index (CPI). We then smooth the relationship between after-tax and before-tax income, fit a smoothing spline and compute its derivative to recover the marginal tax rates on taxable income. The

[11]The aggregate dirty good has average CO_2 emissions of 2.24kg per dollar of final expenditure, calculated as the expenditure-weighted average of its three components for which emissions data are available: gasoline, electricity, and natural gas (see Appendix Table A.2 for a breakdown by component).

[12]Coincidentally, this approximately corresponds to the gasoline tax in France (TICPE) in 2024 (DILA, 2024); while in our data (see Appendix Section A.3 for details on data sources) the average gasoline tax across US states was 18.5% in 2021.

blue curve in Figure A.3 shows the resulting marginal income tax rates across the income distribution used in our simulations.

4.3 Empirically Optimal Tax Rates

Figure 6 plots the marginal tax rates on the carbon-intensive good that are implied by our Pareto efficiency conditions. Panels a and c present the nonlinear case (equation 29). Panels b and d present the case where the tax is restricted to be linear (equation 36). Panels a and b assume an elasticity of taxable income $\varepsilon_z = 0.33$, which is most consistent with Chetty's 2012 meta-analysis. Panels c and d assume $\varepsilon_z = 0.7$. In each panel, we present estimates for a range of values for the compensated net-of-tax rate elasticity of spending on the carbon-intensive good.

4.3.1 Optimal Non-linear Taxation of Carbon

We focus first on the nonlinear case, assuming that non-linear taxation of these particular commodities is feasible. Panel (a) of Figure 6 presents our results using our baseline value of ε_z. There is a mild U-shaped pattern of the Pareto-efficient carbon tax over the income distribution, which is centered approximately on the level of marginal damage from these goods. Optimal marginal tax rates start near the level of marginal damage for very low-income households, fall below it for low- and middle-income households, and rise above the Pigouvian level for the highest-income households. This pattern is somewhat amplified when consumption of carbon-intensive goods is assumed to be very inelastic (i.e., $\varepsilon_{x|z}$ is very low).

When the elasticity of taxable income is higher ($\varepsilon_z = 0.7$), panel (c) shows that there are larger deviations from Pigouvian taxation, although they are still centered around the level of marginal damage. This is especially the case when $\varepsilon_{x|z}$ is low. This combination of parameters implies that income taxation is highly distortionary relative to adjusting the carbon tax to redistribute, providing that some part of the covariance of income and carbon intensity is explained by differences in preferences ($\eta^{\text{Taste}}_{x,z}(z) \neq 0$). The most important implication of these more extreme parameter values is that the optimal carbon tax is meangfully higher than the Pigouvian level for the highest-income individuals. For middle-income households, the optimal carbon tax is lower than the Pigouvian level, but it is never comes close to being Pareto efficient to subsidize carbon consumption.

4.3.2 Optimal Linear Taxation of Carbon

Nonlinear taxation of these commodities may prove logistically or politically challenging. In this case, the carbon tax must be restricted to be linear. Panels (b) and (d) show that the

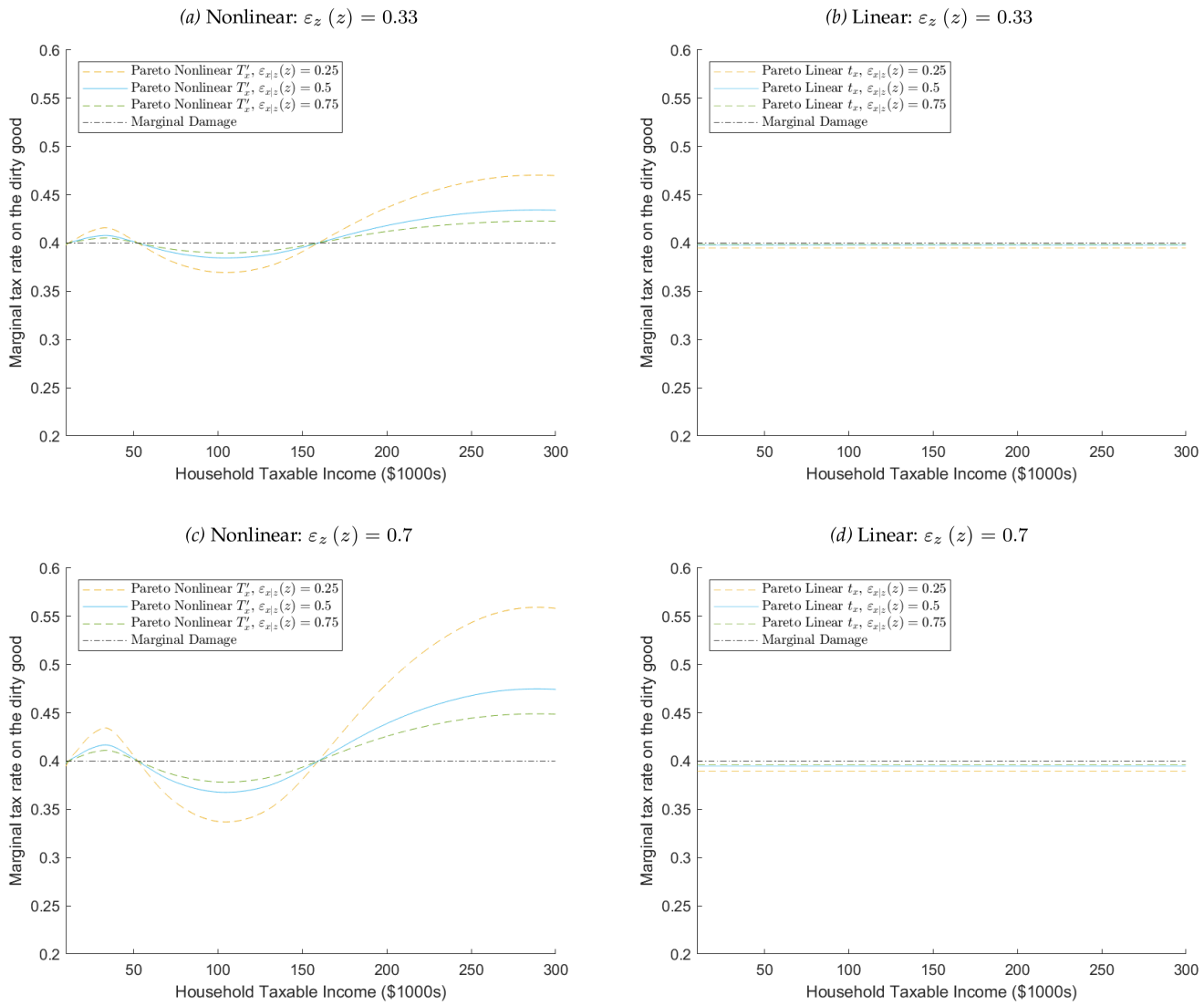


Figure 6. Dirty Good Tax Schedule: Marginal Rates implied by Pareto Efficient formulas

Note: This figure plots the Pareto efficient carbon tax for different values of the elasticities of taxable income and consumption of the carbon-intensive good. Three cases are shown in each panel, along with the benchmark of the assumed value of marginal damage. Panels (a) and (c) show nonlinear marginal tax schedules, while panels (b) and (d) restrict the carbon tax to be linear.

Pareto-efficient linear tax is extremely close to the Pigouvian level for all parameter values. In our benchmark scenario with $\varepsilon_z = 0.33$, and $\varepsilon_{x|z} = 0.5$, the Pareto-efficient linear tax is equal to 99.4% of the value of marginal damage. In this sense, distributional concerns have very little bearing at all on the optimal carbon tax if it is required to be linear.

This stark result comes from the fact that the Pareto-efficient linear tax is a weighted average of the quantities determining the marginal tax rates at each income level in the nonlinear case (see equation 36). Intuitively, departures from the Pigouvian level of the carbon tax will depend on a weighted average tagging benefit of carbon. As Figure 5 shows, $\eta_{x,z}^{\text{Taste}}(z)$ is negative for the portion of the income distribution with the largest share of households, but positive at high levels of income. Integrating over the distribution, the Pareto-efficient linear tax is therefore below, but very close to, the Pigouvian level.

4.3.3 Optimal Linear Taxation of Carbon with Multidimensional Heterogeneity

To empirically quantify the optimal carbon tax with multidimensional heterogeneity, we exploit our individual-level survey measure of the causal effect of income on consumption of the carbon-intensive good. We first group survey respondents whose household taxable income ranges from \$600 to \$325,000, into equally-sized deciles of taxable income; and then compute the conditional variance of $\partial x/\partial z$ in each income decile. Figure 7 shows these conditional variances together with bootstrapped confidence intervals.

The conditional variance of the causal consumption response to a small increase in taxable income is largest for households in the 2nd decile, earning between \$36,000 and \$50,000. It is otherwise slightly decreasing across the income distribution, from 0.013 to 0.004.

We assign these conditional variances to all income levels within each decile and smooth out their relationship with taxable income across the income distribution (as described above). This allows us to simulate the optimal linear externality tax using the formula displayed in equation 56. We re-use most of the empirical quantities from the unidimensional case, now highlighting the conditional average interpretation, e.g. $\eta_{x,z}^{\text{Taste}}(z) = \mathbb{E}[\eta_{x,z}^{\text{Taste}}(z,\theta) \mid z]$ and $\bar{x}(z) = \mathbb{E}[x(z,\theta) \mid z]$. We continue to assume constant values for the elasticities of taxable income and consumption of the carbon-intensive good.

Figure 8 plots the optimal tax rate for the multidimensional case (orange dotted line). For comparison, we also include the the linear tax rate implied by the Pareto efficient linear tax formula in the unidimensional case (green dashed line). Panel (a) assumes an elasticity of taxable income $\varepsilon_z = 0.33$, and an elasticity of consumption of the carbon-intensive good $\varepsilon_{x|z} = 0.5$, while panel (b) assumes $\varepsilon_z = 0.7$, and $\varepsilon_{x|z} = 0.25$. As expected based

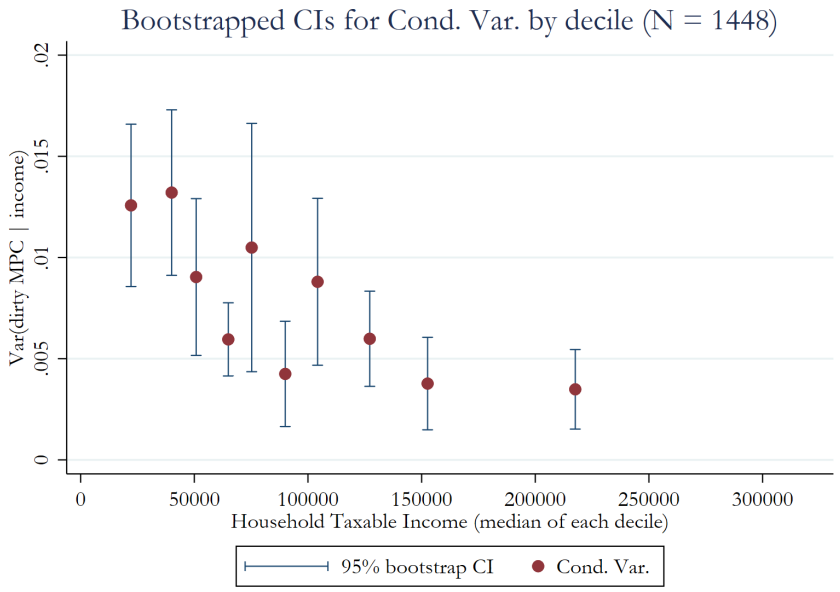


Figure 7. Conditional Variance of $\partial x/\partial z$ at Each Income Decile

Note: This figure shows the conditional variance of the causal effect of income on consumption of the dirty good, $\text{Var}(\frac{\partial x(w,\theta)}{\partial z} \mid z)$, measured in our hypothetical survey ($N = 1,448$), and estimated by grouping survey respondents across equal-binned deciles of household taxable income. It is based on each respondent's share of MPC allocated to the aggregate dirty good.

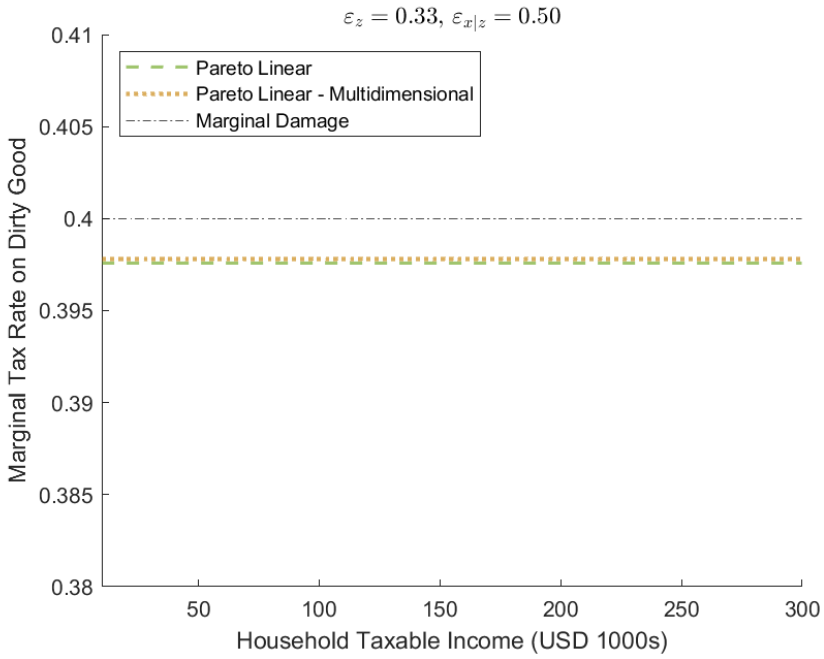
on our theoretical results, within-income variation of the causal effect of income leads to attenuation and pushes the optimal tax closer to the Pigouvian benchmark.

Comparing across the two panels, we see that the degree of attenuation is larger in the lower panel, where the deviation from Pigouvian taxation was relatively larger as well. This is because the amount of attenuation is proportional to the ratio of elasticities $\varepsilon_z/\varepsilon_{x|z}$ (see equation 56). Recall that this same ratio scales the intensity of the tagging benefit in the first place, which implies that the degree of attenuation from within-income heterogeneity in $\partial x/\partial z$ is greatest when the tagging benefit is largest. As a result, there is only minimal attenuation in Figure 8(a) where the optimal linear rate in unidimensional case was almost exactly at the Pigouvian level already. There is more attenuation in panel (b), where the higher ratio of elasticities led to a larger deviation.

5 Conclusion

The use of taxes or subsidies to internalize externalities is widely accepted by economists as an important role of government. When it comes specifically to climate change, economists have forcefully advocated for a carbon tax on these grounds (Akerlof et al., 2019). But this idea has gained limited traction politically.

(a) Baseline calibration



(b) Higher ε_z, lower $\varepsilon_{x|z}$

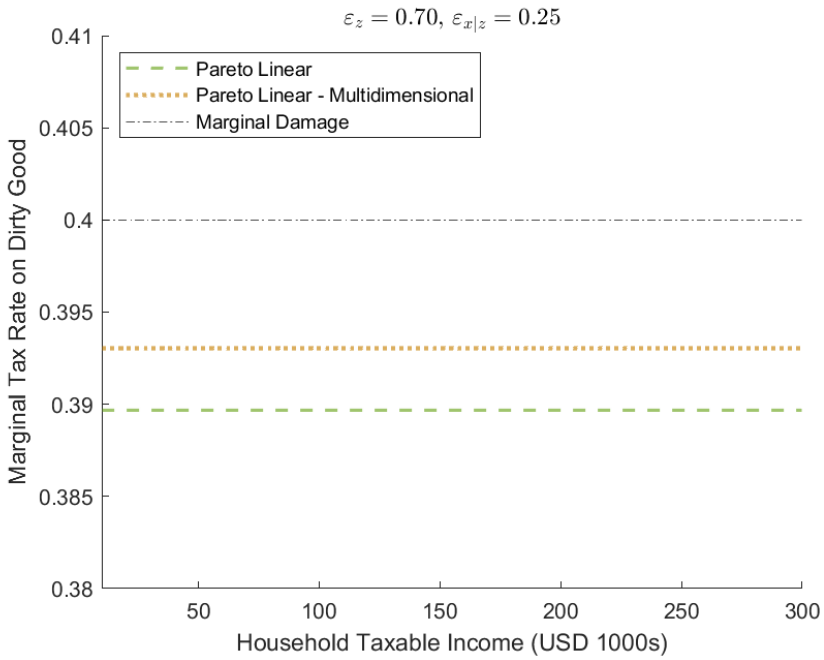


Figure 8. Dirty Good Marginal Linear Tax Rate: Unidimensional vs. Multidimensional Heterogeneity

Note: This figure plots the Pareto efficient (linear) carbon tax in the unidimensional case vs. the quasi Pareto efficient linear tax rate when accounting for multidimensional heterogeneity, for different values of the elasticities of taxable income and consumption of the carbon-intensive good.

We formally study one specific critique of carbon taxes, and of corrective taxes more generally: They can be regressive. Our theoretical results show that the fact that poorer individuals are more carbon-intensive in their consumption does not necessarily change the optimal carbon tax. Instead, the key distinction is whether this reflects income effects or preference heterogeneity. We show how to quantify preference heterogeneity empirically, and take our sufficient statistics tax formulas to the data.

The implications of our empirical results for the optimal carbon tax depend on whether nonlinear carbon taxation is feasible. If only linear carbon taxation is possible, then the optimal carbon tax is remarkably close to the classic Pigouvian prescription of setting the tax equal to the marginal damage from emitted carbon. This conclusion for the optimal linear tax is robust to a very wide range of different parameter values.

The conclusion is more nuanced if nonlinear carbon taxation is possible. Our benchmark scenario implies that marginal tax rates are modestly below the Pigouvian level for middle-income households, but above that level for higher-income households. This pattern is amplified if the elasticity of taxable income is higher, or the behavioral response to carbon taxation is smaller: In such cases, adjusting the carbon tax becomes a more attractive tool to redistribute income, relative to income taxation.

We expect that our conceptual and empirical frameworks could be applied to a wide range of settings that involve both distributional concerns and externalities. This could include other environmental externalities, public health problems, or even criminal behavior. Our results for this particular application suggest that preference-driven differences in carbon intensity have a limited impact overall. Our estimate of the optimal carbon tax is therefore close to the Pigouvian benchmark, especially if the tax is constrained to be linear. However, this result may differ for other externalities. In the future, it will also be important to integrate our detailed work on the consumption side of the economy with richer analysis of the production side along the lines of Bierbrauer (2024) and others.

References

ACEA (2020). Co2-based motor vehicle taxes in the european union. Report, ACEA (European Automobile Manufacturers' Association).

Ahlvik, L., Liski, M., and Mäkimattila, M. (2024). Pigouvian income taxation. CESifo Working Paper 11174, CESifo.

Akerlof, G., Aumann, R., Baily, M., Bernanke, B., Boskin, M., Deaton, A., Diamond, P., Engle, R., Fama, E., Feldstein, M., Furman, J., Goolsbee, A., Greenspan, A., Hansen, L., Hart, O., Holmström, B., Hubbard, G., Kahneman, D., Krueger, A., Kydland, F., Lazear, E., Lucas, R., Mankiw, N. G., Maskin, E., McFadden, D., Merton, R., Myerson, R., Phelps, E., Romer, C., Rosen, H., Roth, A., Sargent, T., Scholes, M., Sen, A., Sharpe, W., Shiller, R., Shultz, G., Sims, C., Solow, R., Spence, M., Summers, L., Thaler, R., Tyson, L., Volcker, P., and Yellen, J. (2019). Economists' statement on carbon dividends. *The Wall Street Journal*.

Akerlof, G. A. (1978). The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *The American Economic Review*, 68(1):8–19.

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

Andersson, J. J. (2019). Carbon taxes and co2 emissions: Sweden as a case study. *American Economic Journal: Economic Policy*, 11(4):1–30.

Aronsson, T. and Blomquist, S. (2003). Optimal taxation, global externalities and labor mobility. *Journal of Public Economics*, 87:2749–2764.

Aronsson, T. and Johansson-Stenman, O. (2008). When the joneses' consumption hurts: Optimal public good provision and nonlinear income taxation. *Journal of Public Economics*, 92:986–997.

Atkinson, A. B. and Stiglitz, J. E. (1976). The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, 6(1–2):55–75.

Barrage, L. (2020). Optimal dynamic carbon taxes in a climate–economy model with distortionary fiscal policy. *The Review of Economic Studies*, 87(1):1–39. Advance access published 15 October 2019.

Belfiori, E., Carroll, D., and Hur, S. (2024). Unequal climate policy in an unequal world. Globalization Institute Working Paper 427, Federal Reserve Bank of Dallas.

Beznoska, M. (2014). Estimating a consumer demand system of energy, mobility and leisure: A microdata approach for germany. *Working Paper*.

Bierbrauer, F. (2024). Is a uniform price on carbon desirable? a public finance perspective. CESifo Working Paper 11201, CESifo.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2024). On binscatter. *American Economic Review*, 114(5):1488–1514.

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

Craig, A. C. (2023). Optimal income taxation with spillovers from employer learning. *American Economic Journal: Economic Policy*, 15(2):82–125.

Craig, A. C. and Hines, James R., J. (2020). Taxes as pandemic controls. *National Tax Journal*, 73(4):969–986.

Cremer, H. and Gahvari, F. (2001). Second-best taxation of emissions and polluting goods. *Journal of Public Economics*, 80(2):169–197.

Cremer, H., Gahvari, F., and Ladoux, N. (1998). Externalities and optimal taxation. *Journal of Public Economics*, 70(3):343–364.

Cremer, H., Gahvari, F., and Ladoux, N. (2003). Environmental taxes with heterogeneous consumers: An application to energy consumption in france. *Journal of Public Economics*, 87(12):2791–2815.

Cremer, H., Gahvari, F., and Ladoux, N. (2010a). Environmental tax design with endogenous earning abilities (with applications to france). *Journal of Environmental Economics and Management*, 59(1):82–93.

Cremer, H., Gahvari, F., and Lozachmeur, J.-M. (2010b). Tagging and income taxation: Theory and an application. *American Economic Journal: Economic Policy*, 2(1):31–50.

Cronin, J. A., Fullerton, D., and Sexton, S. E. (2019). Vertical and horizontal redistribution from a carbon tax and rebate. *Journal of the Association of Environmental and Resource Economists*, 6(S1):S169–S208.

Dechezleprêtre, A., Fabre, A., Kruse, T., Planterose, B., Sanchez Chico, A., and Stantcheva, S. (2025). Fighting climate change: International attitudes toward climate policies. *American Economic Review*, 115(4):1258–1300.

Diamond, P. (1975). A many-person ramsey tax rule. *Journal of Public Economics*, 4(4):335–42.

Diamond, P. and Spinnewijn, J. (2011). Capital income taxes with heterogeneous discount rates. *American Economic Journal: Economic Policy*, 3(4):52–76.

Dietz, S. and Atkinson, G. (2010). The equity-efficiency trade-off in environmental policy: Evidence from stated preferences. *Land Economics*, 86(3):423–443.

DILA (2024). Que sont les droits d'accises sur les énergies et la TICPE ? Direction de l'information légale et administrative. Mis à jour le 31 juillet 2024.

Doligalski, P., Dworczak, P., Akbarpour, M., and Kominers, S. D. (2025). Optimal redistribution via income taxation and market design. Working Paper.

Dorsey, J. and Wolfson, D. C. (2024). Unequal uptake: Assessing distributional disparities in the residential solar market. *NBER Working Paper n. 32814*.

Douenne, T., Hummel, A. J., and Pedroni, M. (2025). Optimal fiscal policy in a climate-economy model with heterogeneous households. *Working Paper*.

EPA (2023). Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances. Technical report, U.S. Environmental Protection Agency.

Ferey, A., Lockwood, B., and Taubinsky, D. (2021). Sufficient statistics for nonlinear tax systems with general across-income heterogeneity. *NBER Working Paper n. 29582*.

Ferey, A., Lockwood, B. B., and Taubinsky, D. (2024). Sufficient statistics for nonlinear tax systems with general across-income heterogeneity. *American Economic Review*, 114(10).

Fricke, L., Fuest, C., and Sachs, D. (2025). Carbon emissions and redistribution: The design of carbon tax rebates. Working paper.

Fried, S., Novan, K., and Peterman, W. B. (2018). The distributional effects of a carbon tax on current and future generations. Working paper.

Fried, S., Novan, K., and Peterman, W. B. (2024). Understanding the inequality and welfare impacts of carbon tax policies. Working paper.

Fuster, A. and Zafar, B. (2023). *Survey experiments on economic expectations*, chapter 4. Academic Press.

Gauthier, S. and Henriet, F. (2018). Commodity taxes and taste heterogeneity. *European Economic Review*, 101:284–296.

Gerritsen, A. (2024). Optimal nonlinear taxation: a simpler approach. *International Tax and Public Finance*, 31(2):486–510.

Golosov, M., Tsyvinski, A., and Werquin, N. (2014). A variational approach to the analysis of tax systems. Technical report, National Bureau of Economic Research.

Goulder, L. H., Hafstead, M. A. C., Kim, G., and Long, X. (2019). Impacts of a carbon tax across US household income groups: What are the equity–efficiency trade-offs? *Journal of Public Economics*, 175:44–64.

Grainger, C. A. and Kolstad, C. D. (2010). Who pays a price on carbon? *Environmental and Resource Economics*, 46(3):359–376.

Hahn, R. W., Hendren, N., Metcalfe, R. D., and Sprung-Keyser, B. (2024). A welfare analysis of policies impacting climate change. *NBER Working Paper n. 32728*.

Hassett, K. A., Mathur, A., and Metcalf, G. E. (2009). The incidence of a u.s. carbon tax: A lifetime and regional analysis. *Energy Journal*, 30(2):155–178.

Intergovernmental Panel on Climate Change (2023). *Climate Change 2023: Synthesis Report*. IPCC.

Jacobs, B. and de Mooij, R. A. (2015). Pigou meets mirrlees: On the irrelevance of tax distortions for the second-best pigouvian tax. *Journal of Environmental Economics and Management*, 71:90–108.

Jacobs, B. and van der Ploeg, F. (2019). Redistribution and pollution taxes with non-linear engel curves. *Journal of Environmental Economics and Management*, 95:198–226.

Jacquet, L. and Lehmann, E. (2021). Optimal income taxation with composition effects. *Journal of the European Economic Association*, 19(2):1299–1341.

Kaplow, L. (2006). Optimal income transfers. *International Tax and Public Finance*, 14(3):295–325.

Kaplow, L. (2012). Optimal control of externalities in the presence of income taxation. *International Economic Review*, 53(2):487–509.

Kopczuk, W. (2003). A note on optimal taxation in the presence of externalities. *Economics Letters*, 80:81–86.

Li, S., Linn, J., and Muehlegger, E. (2014). Gasoline taxes and consumer behavior. *American Economic Journal: Economic Policy*, 6(4):302–342.

List, J., Sinha, P., and Taylor, M. H. (2006). Using choice experiments to value non-market goods and services: Evidence from field experiments. *B.E. Journal of Economic Analysis & Policy*, 6(2).

Lockwood, B. B., Nathanson, C. G., and Weyl, E. G. (2017). Taxation and the allocation of talent. *Journal of Political Economy*, 125(5):1–49.

Lyubich, E. (2024). The role of people vs. places in individual carbon emissions. *Working paper*.

Mankiw, N. G. and Weinzierl, M. (2010). The optimal taxation of height: A case study of utilitarian income redistribution. *American Economic Journal: Economic Policy*, 2(1):155–176.

Mathur, A. and Morris, A. C. (2014). Distributional effects of a carbon tax in broader us fiscal reform. *Energy Policy*, 66:326–334.

Metcalf, G. E. (2009). A distributional analysis of green tax reforms. *National Tax Journal*, 52(4):655–681.

Micheletto, L. (2008). Redistribution and optimal mixed taxation in the presence of consumption externalities. *Journal of Public Economics*, 92(10-11):2262–2274.

Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38(2):175–208.

Parsons, D. O. (1996). Imperfect 'tagging' in social insurance programs. *Journal of Public Economics*, 62(1–2):183–207.

Pigou, A. C. (1920). *The Economics of Welfare*. Macmillan and Co., London.

Piketty, T. and Saez, E. (2013). A theory of optimal inheritance taxation. *Econometrica*, 81(5):1851–1886.

Pirttilä, J. and Tuomala, M. (1997). Income tax, commodity tax and environmental policy. *International Tax and Public Finance*, 4:379–393.

Pizer, W. A. and Sexton, S. (2019). The distributional impacts of energy taxes. *Review of Environmental Economics and Policy*.

Poterba, J. M. (1989). Lifetime incidence and the distributional burden of excise taxes. *The American Economic Review*, 79(2):325–330.

Rausch, S., Metcalf, G. E., and Reilly, J. M. (2011). Distributional impacts of carbon pricing: A general equilibrium approach with micro-data for households. *Energy Economics*, 33:S20–S33.

Rothschild, C. and Scheuer, F. (2016). Optimal taxation with rent-seeking. *American Economic Journal: Economic Policy*, 83(1):1225–1262.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The review of economic studies*, 68(1):205–229.

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

Sallee, J. M. (2022). Pigou creates losers: On the implausibility of achieving pareto improvements from efficiency-enhancing policies. *Working Paper*.

Sandmo, A. (1975). Optimal taxation in the presence of externalities. *The Swedish Journal of Economics*, pages 86–98.

Walters, A. A. (1961). The theory and measurement of private and social cost of highway congestion. *Econometrica*, 29(4):676–699.

Weitzman, M. L. (1974). Prices vs. quantities. *Review of Economic Studies*, 41(4):477–491.

West, S. and Williams III, R. C. (2007). Optimal taxation and cross-price effects on labor supply: Estimates of the optimal gas tax. *Journal of Public Economics*, 91(3-4):593–616.

Williams III, R. C., Gordon, H., Burtraw, D., Carbone, J. C., and Morgenstern, R. D. (2015). The initial incidence of a carbon tax across income groups. *National Tax Journal*, 68(1):195–214.

Zhang, S. and Qin, X. (2015). Lessons learned from china's residential tiered electricity pricing reform. Research report, International Institute for Sustainable Development.

A Empirical Appendix

A.1 Average Expenditure Shares and CO_2 Emissions across Categories

Table A.1. Average Expenditure Shares across Categories

	Avg.
Food and Beverages	0.183
Housing (excluding utilities)	0.272
Electricity, natural gas, and home heating fuels	0.044
Water, sewage, and other public services	0.013
Healthcare	0.095
Gasoline and other motor fuels	0.039
Flights / Airfare expenditures	0.005
Other Expenditures	0.347
Observations	113,108

Note: This table shows shares allocated to each consumption category among consumers in the Consumer Expenditure Survey (2018–2023).

Table A.2. Average CO_2 Emissions (in kg) per Dollar of Final Expenditure in the United States

	Unit	**Price Per Unit (in $)**	**CO_2 Emissions per Unit (in Kg)**	**CO_2 Emissions (Kg per $)**
Gasoline	Gallon (3.78 Litres)	3.635	8.78	2.415
Electricity	KWh	0.2275	0.387	1.699
Natural Gas	1000 Cubic Feet	15.39	53.035	3.446

Note: We use 2023 data from U.S. Energy Information Administration (EIA) for information on prices per unit, as well as 2023 data from the U.S. Environmental Protection Agency (EPA), and the Code of Federal Regulations for information on CO2 emissions per unit.

A.2 Share of Carbon-Intensive Goods in Status Quo Expenditures

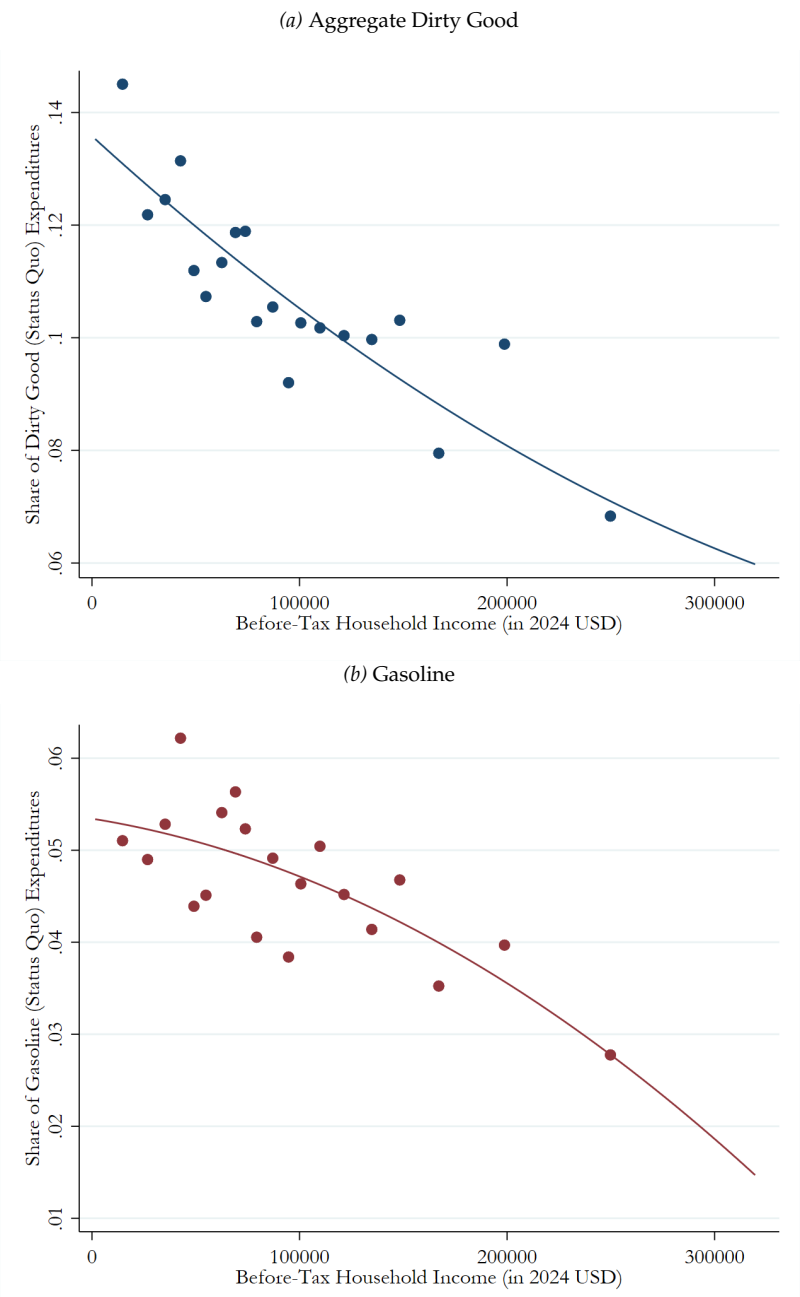


Figure A.1. Share of carbon-intensive goods in total (monthly) expenditures in custom survey

Note: These binscatter plots show (i) the average share of expenditure on polluting goods as a fraction of total expenditures, and (ii) the corresponding average household taxable income, averaged across 20 bins of household taxable income (selected with the command binsreg (Cattaneo et al., 2024)), and (iii) a global polynomial of degree 2 estimated to fit the relationship between these two variables, using data from a custom representative survey of the US population ($N = 1,448$). The "dirty good" expenditure category includes gasoline, electricity, natural gas, and heating fuels.

A.3 Estimating the Uncompensated Elasticity of Gasoline Consumption

Following Li et al. (2014), we estimate a log-log two-way fixed effects regression to recover the elasticity of gasoline consumption with respect to the net-of-tax rate. We use granular data at the household level rather than aggregated data at the state level and estimate the following regression equation:

$$\log(q_{ist}) = \alpha \log(p_{st}) + \beta \log(1 + \frac{\tau_{st}}{p_{st}}) + \Theta X_{ist} + \delta_s + \phi_t + e_{ist} \tag{63}$$

where q_{ist} is the household-level consumption of gasoline (in quantities), p_{st} is the tax-exclusive price of gasoline, τ_{st} is the gasoline excise tax (in dollars); X_{ist} is a vector of household-level controls; and δ_s and ϕ_t are state, and time (yearly or monthly) fixed effects.

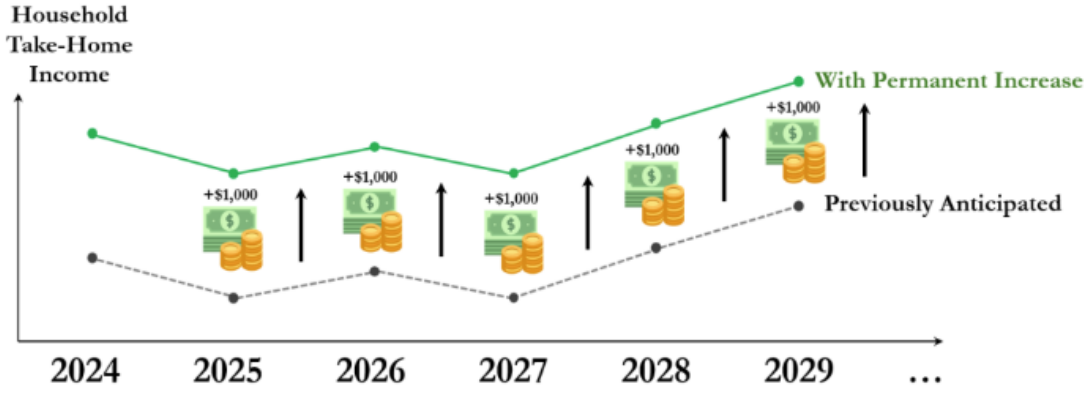
The coefficient $-\beta$ in the first row can be interpreted as the uncompensated elasticity of gasoline consumption with respect to the net-of-tax rate.

Table A.3. Elasticity of gasoline consumption with respect to tax rate: 1996-2021

	(1)	(2)	(3)	(4)
$\log(1 + \text{tax ratio})$	-0.325	-0.386	-0.539*	-0.584**
	(0.312)	(0.313)	(0.288)	(0.288)
$\log(\text{gas price})$	-0.453***	-0.468***	-0.419***	-0.429***
	(0.143)	(0.141)	(0.136)	(0.133)
State FE	Y	Y	Y	Y
Year FE	Y		Y	
Month × Year FE		Y		Y
Household Level Controls	N	N	Y	Y
R^2	0.089	0.109	0.198	0.218
Observations	596,822	596,822	542,574	542,574
Clusters (State × Year)	1,035	1,035	991	991

Note: This table shows the coefficient estimates from the regression of equation (63). All regressions control for family size and use annual weights to make the sample representative of the United States each calendar year (see section 6.3.1 of https://www.bls.gov/cex/pumd-getting-started-guide.htm#section3). Additional household-level control variables include age (of the reference person), urban vs. rural status, taxable income, education level, and family type. Quarterly data from 1996 to 2021 on household gasoline consumption are from the interview module of the Consumer Expenditure Surveys (CEX) Public use microdata files. We can use Month × Year fixed effects because of the rotating nature of the survey. Gasoline tax data are obtained from the U.S. Department of Transportation, Federal Highway Administration, and gasoline prices are from the U.S. Energy Information Administration (State Energy Data System, SEDS) for the same period. Standard errors are clustered at the State × Year level. *** p<0.01, ** p<0.05, * p<0.1.

A.4 Screenshot of Main Survey Elicitation Instrument



Imagine once again that your after-tax household's income is $1,000 higher **each year** than what you originally expected, forever. On **average**, how much of this extra $1,000 would your household spend (rather than save) on the following categories, **each year**?

For each category, indicate a number corresponding to the additional dollar amount spent on each category. These categories, including savings, are exhaustive, and the total amount (at the bottom) will automatically sum up to $1,000. The amount saved each year will automatically be calculated and displayed below.

Food (both at and away from home) and beverages (including alcoholic beverages)	0
Housing expenditure excluding utilities (includes rent, owned dwelling expenditures, mortgage payments, home insurance, furnishing and appliances, and domestic services)	0
Electricity, natural gas (for heating and cooking), fuel oil and other home heating fuels	0
Water, sewage, and other public services (e.g. trash and garbage collection)	0
Healthcare (health insurance, medical services, drugs, and medical supplies)	0
Gasoline, diesel and/or other motor fuels	0
Flights / Airfare expenditures	0
All other categories (telecommunications services, entertainment, other types of transportation, clothing, personal insurance and pensions, and education expenditures)	0
Savings	1000
Total	1,000

Additional amount **saved each year**: $1,000

Figure A.2. Survey Screen for the Causal Consumption Response elicitation block

A.5 Status Quo Income Marginal Tax Rates

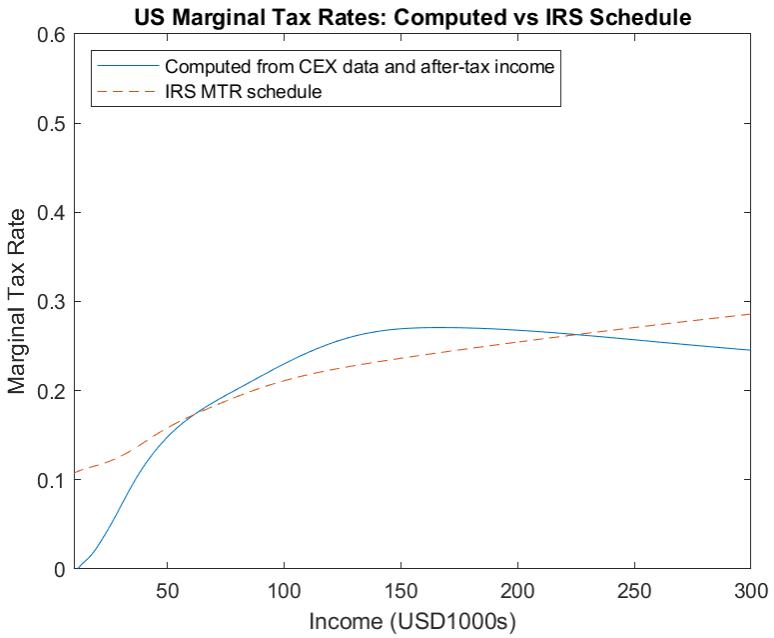


Figure A.3. Status quo income marginal tax rates across the income distribution

Note: This figure shows the status quo income marginal tax rates across the household income distribution. The blue curve shows estimated marginal tax rates used in our analysis, and computed from CEX data (2018–2023, converted to 2024 USD) by fitting a smoothing spline to the relationship between after-tax and before-tax income binned by income percentiles, and computing 1 minus its derivative. The dashed orange line shows the 2024 IRS statutory tax rate schedule (adjusted to account for the share of married couples) for illustrative comparison.

A.6 Pareto Efficient Marginal Tax Schedules - Gasoline Only

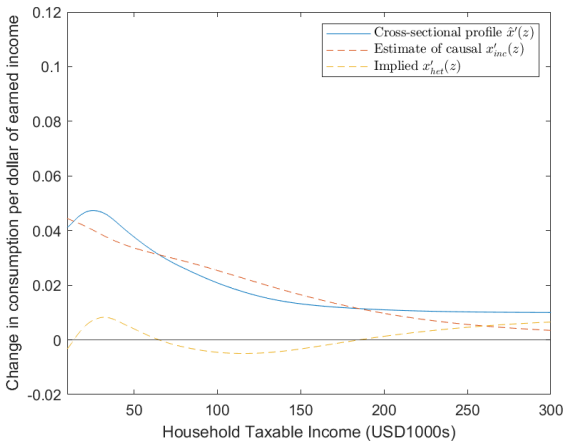


Figure A.4. Gasoline: Decomposition of the cross-sectional profile $\hat{x}'(z)$

Note: This figure shows how we decompose the cross-sectional relationship between income and gasoline consumption. The blue line is a smoothed estimate of the cross-sectional slope (from the 2018-2023 CEX surveys), $\hat{x}'(z)$ at each level of income. The dashed red line shows estimates of the causal effect of income on expenditure on carbon-intensive goods (from our custom hypothetical survey). The yellow dashed line subtracts the income effect from the cross-section to obtain the part of the relationship between x and z that comes from preference heterogeneity.

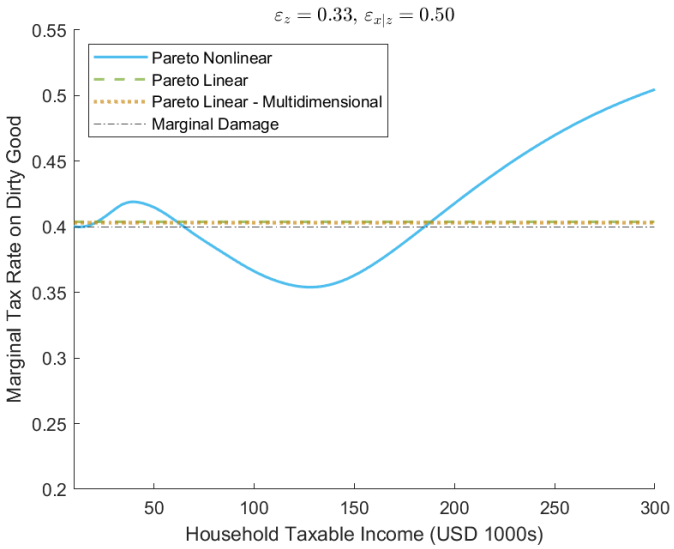


Figure A.5. Gasoline Marginal Tax Rates: Unidimensional vs. Multidimensional Heterogeneity

Note: This figure plots the Pareto efficient carbon tax for **gasoline only** in the unidimensional case (Linear and Nonlinear) vs. the quasi Pareto efficient linear tax rate when accounting for multidimensional heterogeneity, when the elasticity of taxable income is $\varepsilon_z = 0.33$, and the elasticity of consumption of the carbon-intensive good $\varepsilon_{x|z} = 0.5$.

A.7 Pareto Efficient Marginal Tax Schedules - Electricity, Natural Gas, and Heating Fuels Only

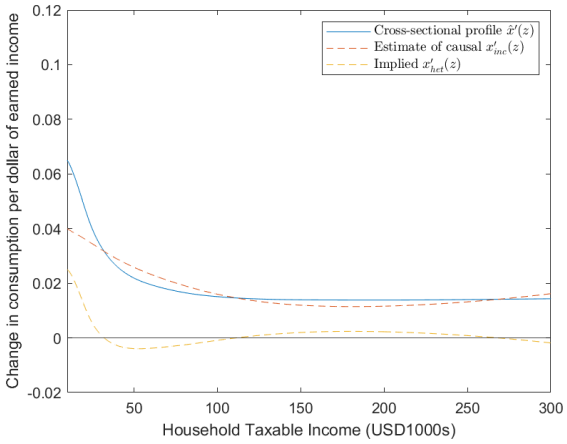


Figure A.6. Electricity, Natural Gas, and Heating Fuels: Decomposition of the cross-sectional profile $\hat{x}'(z)$

Note: This figure shows how we decompose the cross-sectional relationship between income and electricity, natural gas, and heating fuels consumption. The blue line is a smoothed estimate of the cross-sectional slope (from the 2018-2023 CEX surveys), $\hat{x}'(z)$ at each level of income. The dashed red line shows estimates of the causal effect of income on expenditure on carbon-intensive goods (from our custom hypothetical survey). The yellow dashed line subtracts the income effect from the cross-section to obtain the part of the relationship between x and z that comes from preference heterogeneity.

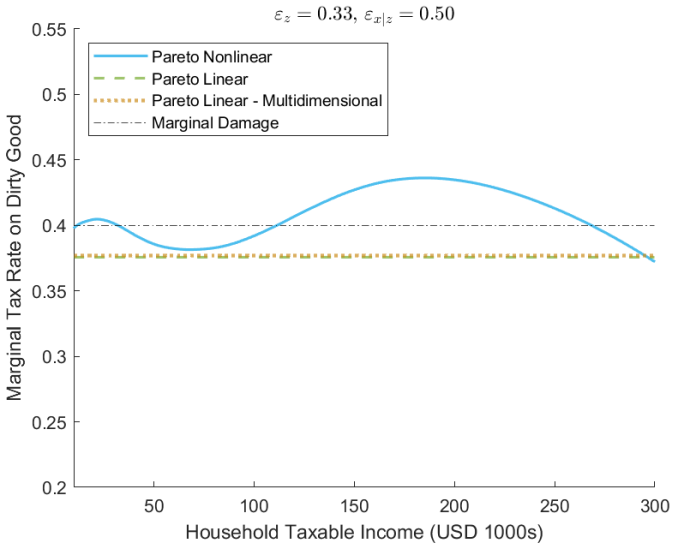


Figure A.7. Electricity, Natural Gas, and Heating Fuels Marginal Tax Rates: Unidimensional vs. Multidimensional Heterogeneity

Note: This figure plots the Pareto efficient carbon tax for **electricity, natural gas, and heating fuels consumption only** in the unidimensional case (Linear and Nonlinear) vs. the quasi Pareto efficient linear tax rate when accounting for multidimensional heterogeneity, when the elasticity of taxable income is $\varepsilon_z = 0.33$, and the elasticity of consumption of the carbon-intensive good $\varepsilon_{x|z} = 0.5$.

A.8 Online Survey Demographics Breakdown & Sample Definition

Ethnicity	Freq.	%	(C)
White	1097	75.8%	73.9%
Black	199	13.7%	14.3%
Asian	87	6.0%	7.5%
Mixed	43	3.0%	2.5%
Other	22	1.5%	1.7%

Age	Freq.	%	(C)
25-29	263	18.2%	16.9%
30-34	276	19.1%	18.0%
35-39	294	20.3%	17.3%
40-44	196	13.5%	16.8%
45-49	223	15.4%	15.2%
50-54	196	13.5%	15.9%

Sex	Freq.	%	(C)
Female	748	51.7%	49.7%
Male	700	48.3%	50.3%

Table A.4. Demographics Comparison: Online Survey vs 2023 Census Population Estimates

Note: This table shows the breakdown for our custom online survey sample by ethnicity, age, and sex. The rightmost (C) column in each subtable corresponds to the actual demographic distribution of the US population aged 25-54 drawn from 2023 Population Estimates (as of July 1) of the United States Census Bureau constructed with data from the 2020 Census and measures of population change. (See `https://www.census.gov/data/tables/time-series/demo/popest/2020s-national-detail.html`.)

Table A.5. Online Survey Response Quality: Sample Exclusions

	N	%
Total # of Respondents	2,219	100.00
Suspected Bots	18	0.81
Speedsters	421	18.97
Failed Attention Check	44	1.98
Inconsistent or Outlier Income Response	55	2.48
Zero Expenditure, Zero or Inconsistent MPC	233	10.50
Final Sample	1,448	65.25

Note: This table reports the sequential exclusion of Prolific respondents from the combined Waves 1 and 2 of our online survey. Exclusions are applied in the order shown, as per our pre-registration, with each category excluding observations from the remaining sample after previous exclusions. Percentages are calculated relative to total observations.

B Proofs of Theoretical Results

Proof of Proposition 1. Consider the effects of an infinitesimal reform to the baseline tax schedules $\mathcal{T}_z(\cdot)$ and $\mathcal{T}_x(\cdot)$, as described in Section 2.4.

Behavioral Responses

Section 2.6 introduces notation for the various behavioral responses, but we define them more formally here. First, the compensated net-of-tax rate elasticity of taxable income is:

$$\varepsilon_z(w) \equiv -\frac{1-\mathcal{T}_z'(z(w))}{\tau_z'(z(w))\,z(w)}\left.\frac{\partial z(w)}{\partial \kappa}\right|_{\kappa=\tau_z(z(w))+\tau_x(x(w))=\tau_x'(x(w))=0} \tag{64}$$

In words, this is the response of taxable income to a change to the marginal income tax rate without any change in the overall tax burden or the marginal tax rate on the externality-generating good.

Second, holding z constant, the compensated net-of-tax rate elasticity of spending on the externality-generating good is:

$$\varepsilon_{x|z}(w) \equiv -\frac{1+\mathcal{T}_x'(x(w))}{\tau_x'(x(w))\,x(w)}\left.\frac{\partial x(w;z(w))}{\partial \kappa}\right|_{z(w),\kappa=\tau_z(z(w))+\tau_x(x(w))=\tau_z'(z(w))=0}. \tag{65}$$

Our assumptions ensure that all of these responses are smooth.

A Reform To Test Pareto Efficiency

To characterize the set of Pareto efficient tax systems, consider a distribution neutral tax reform as presented in Section 2.5. The income tax is changed in a way that precisely offsets the impact of the change to the tax on the dirty good.

$$\tau_z(z) = -\tau_x(\hat{x}(z)). \tag{66}$$

This implies marginal income tax rate changes of

$$\tau_z'(z) = -\hat{x}'(z)\,\tau_x'(\hat{x}(z)), \tag{67}$$

Welfare Effects of a Tax Reform

This distribution neutral reform has no mechanical effect on the utility level of any agent at any level of income. We argued in Section 2.9 that this leaves us with three effects that

are local to the level of income at which agents face a change in their marginal tax rate.

$$\Delta \mathcal{W}^{z}(z) = \mathcal{T}_z'(z)\frac{z\varepsilon_z(z)}{1-\mathcal{T}_z'(z)}x'_{\text{het}}(z)$$
$$\Delta \mathcal{W}^{x|z}(z) = -\left(\mathcal{T}_x'(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})(z)}{\lambda}\right)\frac{\varepsilon_{x|z}(z)\,\hat{x}(z)}{1+\mathcal{T}_x'(\hat{x}(z))}$$
$$\Delta \mathcal{W}^{z\to x}(z) = \left(\mathcal{T}_x'(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right)x'_{\text{inc}}(z)\frac{z\varepsilon_z(z)}{1-\mathcal{T}_z'(z)}x'_{\text{het}}(z)$$

where λ is the multiplier on the planner's budget constraint, and $\mathcal{D}'(\bar{x})/\lambda$ is the marginal externality cost associated with consumption of x.

Integrating over all types in the economy and imposing the value of $\tau_x'(\hat{x}(z'))$ for our specific reform, the social welfare impact is:

$$\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = \int (\Delta \mathcal{W}^{z}(z) + \Delta \mathcal{W}^{x|z}(z) + \Delta \mathcal{W}^{z\to x}(z))\tau_x(\hat{x}(z))\,h_z(z)\,\mathrm{d}z, \tag{68}$$

where $h_z(z)$ is the distribution of taxable income.

A Necessary Condition for Optimality

If the tax system in set optimally, it must be the case that any such infinitesimal tax reform has no first-order effect on welfare. That is to say, for any $\tau_x(\cdot)$ we must have $\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = 0$.

By the fundamental lemma of the calculus of variations,[13] we obtain equation 29, which we restate here for convenience

$$\frac{\mathcal{T}_x'(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1+\mathcal{T}_x'(\hat{x}(z))} = \eta^{\text{Taste}}_{x,z}(z)\frac{\varepsilon_z(z)}{\varepsilon_{x|z}(z)}\left(\frac{\mathcal{T}_z'(z) + \left[\mathcal{T}_x'(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right]x'_{\text{inc}}(z)}{1-\mathcal{T}_z'(z)}\right)$$

It is straightforward to then re-arrange this to obtain 30, also restated:

$$\mathcal{T}_x'(\hat{x}(z)) - \frac{\mathcal{D}'(\bar{x})}{\lambda} = \frac{\text{RR}(z)\mathcal{T}_z'(z)}{1-\text{RR}(z)x'_{\text{inc}}(z)},$$

[13] If for all continuous functions $\rho : \mathbb{R} \to \mathbb{R}$ we have $\int_0^\infty f(z)\rho(z)dz = 0$, and if f is continuous, then $f(z) = 0$ for all z. Integrating over all types in the economy, the social welfare impact of a distribution-neutral reform is:

$$\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = \int (\Delta \mathcal{W}^{z}(z) + \Delta \mathcal{W}^{x|z}(z) + \Delta \mathcal{W}^{z\to x}(z))\tau_x'(\hat{x}(z'))h_z(z)\,\mathrm{d}z, \tag{69}$$

where $h_z(z)$ is the density of the distribution of income at z. Letting $f(z) \equiv \int(\Delta \mathcal{W}^{z}(z) + \Delta \mathcal{W}^{x|z}(z) + \Delta \mathcal{W}^{z\to x}(z))h_z(z)$ and $\rho(z) \equiv \tau_x'(\hat{x}(z))$. Equation 29 follows immediately.

where:

$$\text{RR}(z) \equiv \eta_{x,z}^{\text{Taste}}(z) \frac{\varepsilon_z(z)}{\varepsilon_{x|z}(z)} \frac{1+\mathcal{T}_x'(\hat{x}(z))}{1-\mathcal{T}_z'(z)}.$$

Pareto Inefficiency

Because this reform holds each agent's total tax liability constant, it has no first-order impact on agents' private utility. Rather, if equation 25 is not equal to zero, the planner can raise revenue by moving in one direction or the other ($\kappa > 0$ or $\kappa < 0$) with no impact on individual utility. The resulting surplus in government revenue can then be used to provide a lump sum transfer, leading to a Pareto improvement. □

Proof of Proposition 2. Consider the effects of a more general infinitesimal reform to the baseline tax schedules $\mathcal{T}_z(\cdot)$ and $\mathcal{T}_x(\cdot)$, as described in Section 2.4. We start introducing some additional notation, as well as characterizing the expanded set of behavioral responses to the reform, which our assumptions again ensure are smooth.

Behavioral Responses

In addition to the behavioral responses defined in our proof of Proposition 1, we will need notation for several additional responses. First, the income effects on taxable income and consumption of x are as follows:

$$\eta_z(w) \equiv -\frac{\frac{\partial z(w)}{\partial \kappa}\Big|_{\kappa=\tau_z'(z(w))=\tau_x'(x(w))=0}}{\tau_z(z(w))+\tau_x(x(w))} \quad \eta_{x|z}(w) \equiv -\frac{\frac{\partial x(w;z(w))}{\partial \kappa}\Big|_{z(w),\kappa=\tau_z'(z(w))=\tau_x'(x(w))=0}}{\tau_z(z(w))+\tau_x(x(w))}.$$

Using this notation, type w's overall responses to a reform are

$$\begin{aligned}\frac{\mathrm{d}z(w)}{\mathrm{d}\kappa}\Big|_{\kappa=0} = &-\frac{\varepsilon_z(w)\,z(w)\,\tau_z'(z(w))}{1-\mathcal{T}_z'(z(w))} + \chi(z(w))\,\tau_x'(x(w;z)) \\ &- \eta_z(w)\left[\tau_z(z(w))+\tau_x(x(w;z))\right]\end{aligned} \tag{70}$$

$$\begin{aligned}\frac{\mathrm{d}x(w;z(w))}{\mathrm{d}\kappa}\Big|_{\kappa=0} = &-\frac{\varepsilon_{x|z}(w)\,x(w)}{1+\mathcal{T}_x'(x(w))}\tau_x'(x(w)) - \eta_{x|z}(w)\left[\tau_z(z(w))+\tau_x(x(w))\right] \\ &+ x_{\text{inc}}'(z(w))\,\frac{\mathrm{d}z(w)}{\mathrm{d}\kappa}\Big|_{\kappa=0}\end{aligned} \tag{71}$$

where $\varepsilon_z(w)$ and $\varepsilon_{x|z}$ are as defined above, and:

$$\chi(w) \equiv \frac{1}{\tau_x'(x(w))}\,\frac{\partial z(w)}{\partial \kappa}\Big|_{\kappa=\tau_z(z(w))+\tau_x(x(w))=\tau_z'(z(w))=0}$$

is the cross-base response of taxable income to the commodity tax rate.

Inserting equation 70 into equation 71, we get:

$$\begin{aligned}\left.\frac{\mathrm{d}x\left(w;z\left(w\right)\right)}{\mathrm{d}\kappa}\right|_{\kappa=0} = &\left[-\frac{\varepsilon_{x|z}\left(w\right)x\left(w\right)}{1+\mathcal{T}_x'\left(x\left(w\right)\right)}+x'_{\text{inc}}(z(w))\chi\left(z\left(w\right)\right)\right]\tau_x'\left(x\left(w\right)\right)\\ &-x'_{\text{inc}}(z(w))\frac{\varepsilon_z\left(w\right)z\left(w\right)}{1-\mathcal{T}_z'\left(z\left(w\right)\right)}\tau_z'\left(z\left(w\right)\right)\\ &-\left[\eta_{x|z}\left(w\right)+x'_{\text{inc}}(z(w))\eta_z\left(w\right)\right]\left[\tau_z\left(z\left(w\right)\right)+\tau_x\left(x\left(w\right)\right)\right].\end{aligned}$$

By Slutsky symmetry, we have the following expression for the partial effect of taxable income on the choice of the externality-generating good:

$$x'_{\text{inc}}(z(w)) = \frac{x\left(w\right)}{z\left(w\right)}\left[\frac{-\frac{1-\mathcal{T}_z'(z(w))}{x(w)}\chi\left(w\right)}{\varepsilon_z\left(w\right)}\right] = -\frac{1-\mathcal{T}_z'\left(z\left(w\right)\right)}{z\left(w\right)}\frac{\chi\left(w\right)}{\varepsilon_z\left(w\right)}.$$

Put another way, we can write the cross-base response in terms of this partial effect:

$$\chi\left(w\right) = -\frac{z\left(w\right)\varepsilon_z\left(w\right)}{1-\mathcal{T}_z'\left(z\left(w\right)\right)}x'_{\text{inc}}(z(w)).$$

Thus, the income response (equation 70) can be rewritten as

$$\begin{aligned}\left.\frac{\mathrm{d}z\left(w\right)}{\mathrm{d}\kappa}\right|_{\kappa=0} = &-\frac{z\left(w\right)\varepsilon_z\left(w\right)}{1-\mathcal{T}_z'\left(z\left(w\right)\right)}\tau_z'\left(z\left(w\right)\right)-\frac{z\left(w\right)\varepsilon_z\left(w\right)}{1-\mathcal{T}_z'\left(z\left(w\right)\right)}x'_{\text{inc}}(z(w))\tau_x'\left(x\left(w\right)\right)\\ &-\eta_z\left(w\right)\left[\tau_z\left(z\left(w\right)\right)+\tau_x\left(x\left(w\right)\right)\right],\end{aligned} \tag{72}$$

and the response of x (equation 71) can written as:

$$\begin{aligned}\left.\frac{\mathrm{d}x\left(w;z\left(w\right)\right)}{\mathrm{d}\kappa}\right|_{\kappa=0} = &-\left[\frac{\varepsilon_{x|z}\left(w\right)x\left(w;z\left(w\right)\right)}{1+\mathcal{T}_x'\left(x\left(w\right)\right)}+\left(x'_{\text{inc}}(z(w))\right)^2\frac{z(w)\varepsilon_z(w)}{1-\mathcal{T}_z'\left(z\left(w\right)\right)}\right]\tau_x'(x(w))\\ &-x'_{\text{inc}}(z(w))\frac{z\left(w\right)\varepsilon_z\left(w\right)}{1-\mathcal{T}_z'\left(z\left(w\right)\right)}\tau_z'\left(z(w)\right)\\ &-\left[\eta_{x|z}\left(w\right)+x'_{\text{inc}}(z(w))\eta_z\left(w\right)\right]\left[\tau_z\left(z(w)\right)+\tau_x(x(w))\right].\end{aligned} \tag{73}$$

Welfare Effects of a Tax Reform

For any given pair of continuously differentiable reform functions τ_z and τ_x, an infinitesi-

mal reform away from $\mathcal{T}_z$ and $\mathcal{T}_x$ has the following effect on social welfare:

$$\frac{1}{\lambda}\frac{\partial \mathcal{W}}{\partial \kappa}\bigg|_{\kappa=0} + \frac{\partial \mathcal{R}}{\partial \kappa}\bigg|_{\kappa=0} \tag{74}$$

$$= \int (1 - g(w)) \left[\tau_z(z(w)) + \tau_x(x(w))\right] f(w)\, \mathrm{d}w \tag{A}$$

$$- \int \left(\frac{\mathcal{T}_z'(z(w)) + \left[\mathcal{T}_x'(x(w)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x_{\text{inc}}'(z(w))}{1 - \mathcal{T}_z'(z(w))} \right) \varepsilon_z(w)\, z(w)\, \tau_z'(z(w))\, f(w)\, \mathrm{d}w \tag{B}$$

$$- \int \left\{ \left(\frac{\mathcal{T}_z'(z(w)) + \left[\mathcal{T}_x'(x(w)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x_{\text{inc}}'(z(w))}{1 - \mathcal{T}_z'(z(w))} \right) x_{\text{inc}}'(z(w)) \varepsilon_z(w)\, z(w) \ldots \right.$$

$$\left. \ldots + \frac{\mathcal{T}_x'(x(w)) - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + \mathcal{T}_x'(x(w))} \varepsilon_{x|z}(w)\, x(w) \right\} \tau_x'(x(w))\, f(w)\, \mathrm{d}w \tag{C}$$

where λ is the multiplier on the planner's budget constraint, and $\mathcal{D}'(\bar{x})/\lambda$ is the marginal externality cost associated with consumption of x.

The first line (A) captures the social welfare benefit of any redistribution from the reform. This includes differences in income effects and externality impacts, by the definition of $g(w)$. The second line (B) captures the fiscal externality via the income tax. The third line (C) captures the fiscal externality via the tax on x.

Optimality Condition for T_z

If the tax system in set optimally, it must be the case that any such infinitesimal tax reform has no first-order effect on welfare. That is to say, for any $(\tau_z(\cdot), \tau_x(\cdot))$ we must have $\frac{1}{\lambda}\partial \mathcal{W}/\partial \kappa|_{\kappa=0} + \partial \mathcal{R}/\partial \kappa|_{\kappa=0} = 0$. To characterize the optimal income tax, we can therefore consider a reform only to the income tax by supposing that $\tau_x(x) = 0$ for all x, and then set equation 74 equal to zero.

With a change in variables, we can write the resulting optimality condition in terms of the income distribution. For all $\tau_z(\cdot)$:

$$\int (1 - g(\hat{w}(z)))\, \tau_z(z)\, h_z(z)\, \mathrm{d}z$$

$$- \int \left[\frac{\mathcal{T}_z'(z) + \left[\mathcal{T}_x'(x(\hat{w}(z))) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x_{\text{inc}}'(z(w))}{1 - \mathcal{T}_z'(z)} \right] \varepsilon_z(\hat{w}(z))\, z \tau_z'(z)\, h_z(z)\, \mathrm{d}z = 0.$$

By the fundamental lemma of the calculus of variations as we did for Proposition 1, we can then also write the following condition characterizing optimal marginal rates on income in

terms of the distribution of income along with various other sufficient statistics:[14]

$$\frac{\mathcal{T}_z'(z) + \left[\mathcal{T}_x'(x(\hat{w}(z))) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x'_{\text{inc}}(z(w))}{1 - \mathcal{T}_z'(z)} = \frac{\int_z^\infty (1 - g(\hat{w}(s))) h_z(s) \, \mathrm{d}s}{\varepsilon_z(\hat{w}(z)) z h_z(z)}. \tag{76}$$

Defining the average welfare weight for agents with income greater than z as:

$$1 - \bar{g}_+(z) = \frac{\int_z^\infty (1 - g(\hat{w}(s))) h_z(s) \, \mathrm{d}s}{1 - H_z(z)}, \tag{77}$$

we obtain the optimality condition stated in Proposition 2.

Optimality Condition for T_x

Next, to characterize the optimal tax on the externality-generating good, we consider a reform only to the tax on x by supposing that $\tau_z(z) = 0$ for all z, and set equation 74 to zero. Letting $H_x(x)$ be the distribution of x, with density $h_x(x)$, we can write the resulting optimality condition in terms of the distribution of consumption of x. For all $\tau_x(\cdot)$:

$$\int (1 - g(\tilde{w}(x))) \tau_x(x) h_x(x) \, \mathrm{d}x - \int \left[\frac{\mathcal{T}_x'(x) - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + \mathcal{T}_x'(s)} \varepsilon_{x|z}(\tilde{w}(x)) x \ldots \right.$$
$$\left. + \left(\frac{\mathcal{T}_z'(\tilde{z}(x)) + \left[\mathcal{T}_x'(x) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x'_{\text{inc}}(z(w))}{1 - \mathcal{T}_z'(z)} \right) x'_{\text{inc}}(z(w)) \varepsilon_z(\tilde{w}(x)) \tilde{z}(x) \right] \tau_x'(x) h_x(x) \mathrm{d}x = 0,$$

where $\tilde{w}(x)$ and $\tilde{z}(x)$ are the values of w and z which correspond to each value of x.

Similar to the case of the income tax, the fundamental lemma of the calculus of variations then lets us write:

$$\frac{\mathcal{T}_x'(x) - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + \mathcal{T}_x'(x)} \varepsilon_{x|z}(\tilde{w}(x)) x h_x(x) \tag{78}$$
$$+ x'_{\text{inc}}(z(w)) \left(\frac{\mathcal{T}_z'(\tilde{z}(x)) + \left[\mathcal{T}_x'(x) - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x'_{\text{inc}}(z(w))}{1 - \mathcal{T}_z'(z)} \right) \varepsilon_z(w(x)) \tilde{z}(x) h_x(x)$$

[14] Here we use a slightly different variant of fundamental lemma of the calculus of variations, which states that, for all continuously differentiable functions $h : \mathbb{R}_+ \to \mathbb{R}$ we have

$$\int_0^\infty f(x) h(x) dx + \int_0^\infty g(x) h'(x) dx = 0, \tag{75}$$

and if f and g are continuous, then $g'(x) = f(x) \quad \forall x \geq 0$, g is continuously differentiable, and $\lim_{x\to\infty} g(x) = \lim_{x\to 0} g(x) = 0$. Our application of it here follows Jacquet and Lehmann (2021).

$$= \int_x^\infty \left(1 - g\left(\tilde{w}\left(y\right)\right)\right) h_x\left(y\right) dy.$$

We can then write equation 78 in terms of the distribution of taxable income. To do so, we use the fact that:

$$\begin{aligned} h_z\left(z\right) &= \frac{\mathrm{d}x\left(\hat{w}\left(z\right);z\right)}{\mathrm{d}z} h_x\left(x\left(\hat{w}\left(z\right);z\right)\right) \\ &= \left[x'_{\text{inc}}(z(w)) + x'_{\text{het}}(z(w))\right] h_x\left(x\left(\hat{w}\left(z\right);z\right)\right). \end{aligned}$$

Using this result, multiplying both sides of equation 78 by $x'_{\text{inc}}(z(w)) + x'_{\text{het}}(z(w))$, and inserting the optimality condition for the income tax (equation 76) we can rewrite equation 78 as:

$$\begin{aligned} &\frac{\mathcal{T}'_x\left(\hat{x}\left(z\right)\right) - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + \mathcal{T}'_x\left(\hat{x}\left(z\right)\right)} \varepsilon_{x|z}\left(\hat{w}\left(z\right)\right) \hat{x}\left(z\right) h_z\left(z\right) + x'_{\text{inc}}(z(w)) \int_z^\infty \left(1 - g\left(\hat{w}\left(s\right)\right)\right) h_z\left(s\right) \mathrm{d}s \\ &\qquad = \left(x'_{\text{inc}}(z(w)) + x'_{\text{het}}(z(w))\right) \int_z^\infty \left(1 - g\left(\hat{w}\left(s\right)\right)\right) h_z\left(s\right) \mathrm{d}s, \end{aligned}$$

which further simplifies to

$$\frac{\mathcal{T}'_x\left(\hat{x}\left(z\right)\right) - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + \mathcal{T}'_x\left(\hat{x}\left(z\right)\right)} = x'_{\text{het}}(z(w)) \frac{\int_z^\infty \left(1 - g\left(\hat{w}\left(s\right)\right)\right) h_z\left(s\right) \mathrm{d}s}{\varepsilon_{x|z}\left(\hat{w}\left(z\right)\right) \hat{x}\left(z\right) h_z\left(z\right)}. \tag{79}$$

Using the definitions of $\eta^{\text{Taste}}_{x,z}\left(z\right)$ and $\bar{g}_+\left(z\right)$, we obtain the optimality condition stated in the lemma. □

Proof of Proposition 3. The proof here proceeds in a very similar manner to fully non-linear case. Each agent now maximizes her utility subject to her slightly modified budget constraint, producing indirect utility $v(w)$.

$$\begin{aligned} v(w) = \max_{z,x,c} u(c, x, z \mid w) \\ \text{s.t.} \quad c + x + t_x x = z - T_z(z) \end{aligned} \tag{80}$$

Here, t_x is a linear tax on x.

A welfarist social planner chooses t_x and a twice-differentiable $\mathcal{T}_z$ to maximize social

welfare while raising enough revenue to cover an exogenous revenue requirement, $\mathcal{R}$.

$$\max \boldsymbol{\mathcal{W}} = \int \gamma(w) v(w) dF - \mathcal{D}(\overline{x}) \tag{81}$$

$$\text{s.t.} \quad \mathcal{R} = \int \left(T_z(z(w)) + t_x x\right) dF(w) \tag{82}$$

$$\text{and} \quad \overline{x} = \int x(w) dF$$

Mathematically, everything remains identical to the proof of Proposition 1 up until the expression for the general impact on welfare of a tax reform (equation 74).

Imposing that the initial tax on x is linear yields equation 83.

$$\frac{1}{\lambda} \left.\frac{\partial \mathcal{W}}{\partial \kappa}\right|_{\kappa=0} + \left.\frac{\partial \mathcal{R}}{\partial \kappa}\right|_{\kappa=0} \tag{83}$$

$$= \int \left(1 - g\left(w\right)\right) \left[\tau_z\left(z\left(w\right)\right) + \tau_x\left(x\left(w\right)\right)\right] f\left(w\right) \mathrm{d}w \tag{A}$$

$$- \int \left(\frac{\mathcal{T}_z'\left(z\left(w\right)\right) + \left[t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x'_{\text{inc}}(z(w))}{1 - \mathcal{T}_z'\left(z\left(w\right)\right)} \right) \varepsilon_z\left(w\right) z\left(w\right) \tau_z'\left(z\left(w\right)\right) f\left(w\right) \mathrm{d}w \tag{B}$$

$$- \int \left\{ \left(\frac{\mathcal{T}_z'\left(z\left(w\right)\right) + \left[t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x'_{\text{inc}}(z(w))}{1 - \mathcal{T}_z'\left(z\left(w\right)\right)} \right) x'_{\text{inc}}(z(w)) \varepsilon_z\left(w\right) z\left(w\right) \ldots \right.$$

$$\left. \cdots + \frac{t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + t_x} \varepsilon_{x|z}\left(w\right) x\left(w\right) \right\} \tau_x'\left(x\left(w\right)\right) f\left(w\right) \mathrm{d}w \tag{C}$$

From here, we proceed similarly to the proof of Proposition 1.

However, the reform we consider must now be linear as well. Specifically, consider a joint reform where: $\tau_x\left(x\left(\hat{w}\left(z\right); z\right)\right) = \rho \times x(z)$ for some constant ρ; and $\tau_z\left(z\right) = -\rho \times x(z)$. For such a reform, we have $\tau_x'\left(x\left(\hat{w}\left(z\right); z\right)\right) = \rho$ and:

$$\tau_z'\left(z\right) = -\rho \left[x'_{\text{inc}}(z(w)) + x'_{\text{het}}(z(w))\right]. \tag{84}$$

Combining that with equation 83, the impact on welfare is proportional to:

$$\int \left(\frac{\mathcal{T}_z'\left(z\left(w\right)\right) + \left[t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x'_{\text{inc}}(z(w))}{1 - \mathcal{T}_z'\left(z\left(w\right)\right)} \right) \varepsilon_z\left(w\right) z\left(w\right) \left[x'_{\text{inc}}(z(w)) + x'_{\text{het}}(z(w))\right] f\left(w\right) \mathrm{d}w$$

$$-\int \left\{ \left(\frac{\mathcal{T}_z'\left(z\left(w\right)\right) + \left[t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x_{\text{inc}}'(z(w))}{1 - \mathcal{T}_z'\left(z\left(w\right)\right)} \right) x_{\text{inc}}'(z(w)) \varepsilon_z\left(w\right) z\left(w\right) \dots \right.$$
$$\left. \dots + \frac{t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + t_x} \varepsilon_{x|z}\left(w\right) x\left(w\right) \right\} f\left(w\right) \mathrm{d}w. \tag{85}$$

In this case, this simplifies to:

$$\int \left(\frac{\mathcal{T}_z'\left(z\left(w\right)\right) + \left[t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}\right] x_{\text{inc}}'(z(w))}{1 - \mathcal{T}_z'\left(z\left(w\right)\right)} \right) \varepsilon_z\left(w\right) z\left(w\right) x_{\text{het}}'(z(w)) \cdots$$
$$\dots - \left\{ \frac{t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda}}{1 + t_x} \varepsilon_{x|z}\left(w\right) x\left(w\right) \right\} f\left(w\right) \mathrm{d}w. \tag{86}$$

To obtain the equation in the proposition, simply set this to zero to yield an optimality condition. Then replace the integrals with expectations over the income distribution, and impose the definition of $\text{RR}(z)$.

By the same logic as before, there is no mechanical redistribution from this reform. It only impacts social welfare by changing government revenue or the externality. Thus, if this reform has a positive impact on social welfare, it is a Pareto improvement. □

Proof of Proposition 4. Setup and Behavioral Responses

Consider agents characterized by type (w, θ) where $w \in W \subseteq \mathbb{R}_{++}$ captures productivity and $\theta \in \Theta$ captures other preference heterogeneity. Each agent solves:

$$\begin{aligned} \max_{z,x,c} \quad & u(c, x, z; w, \theta) \\ \text{s.t.} \quad & c + x + \mathcal{T}_x(x) = z - \mathcal{T}_z(z) \end{aligned}$$

Define indirect utility $v(w, \theta)$ and choices $z(w, \theta)$, $x(w, \theta)$. The first-order conditions yield:

$$\begin{aligned} 1 + \mathcal{T}_x'(x) &= \text{MRS}_x(c, x, z; w, \theta), \\ 1 - \mathcal{T}_z'(z) &= \text{MRS}_z(c, x, z; w, \theta). \end{aligned}$$

For a reform parameterized by κ with $T_z(z; \kappa) = \mathcal{T}_z(z) + \kappa\tau_z(z)$ and $T_x(x; \kappa) = \mathcal{T}_x(x) + \kappa\tau_x(x)$, implicit differentiation of the first-order conditions yields the total behavioral

responses. Decomposing into substitution and income effects:

$$\left.\frac{dz(w,\theta)}{d\kappa}\right|_{\kappa=0} = -\frac{\varepsilon_z(w,\theta)z(w,\theta)\tau_z'(z(w,\theta))}{1-\mathcal{T}_z'(z(w,\theta))} + \chi(w,\theta)\tau_x'(x(w,\theta)) - \eta_z(w,\theta)[\tau_z(z(w,\theta)) + \tau_x(x(w,\theta))], \tag{87}$$

where $\varepsilon_z(w,\theta)$ is the compensated elasticity of taxable income, $\chi(w,\theta)$ is the cross-base response, and $\eta_z(w,\theta)$ is the income effect.

Slutsky Symmetry

By the symmetry of the Slutsky matrix, the cross-price effect of the commodity tax on labor supply equals the cross-price effect of the wage on commodity demand. This yields:

$$\chi(w,\theta) = -\frac{z(w,\theta)\varepsilon_z(w,\theta)}{1-\mathcal{T}_z'(z(w,\theta))}\frac{\partial x(w,\theta;z(w,\theta))}{\partial z}. \tag{88}$$

Substituting (88) into (87) and similarly deriving $dx/d\kappa$:

$$\left.\frac{dx(w,\theta)}{d\kappa}\right|_{\kappa=0} = -\frac{\varepsilon_{x|z}(w,\theta)x(w,\theta)}{1+\mathcal{T}_x'(x(w,\theta))}\tau_x'(x(w,\theta)) - \left[\eta_{x|z}(w,\theta) + \frac{\partial x}{\partial z}\eta_z(w,\theta)\right][\tau_z(z)+\tau_x(x)] + \frac{\partial x}{\partial z}\left.\frac{dz}{d\kappa}\right|_{\kappa=0}.$$

Welfare Effects of a Reform

Social welfare is $\mathcal{W} = \int_\Theta \int_W \gamma(w,\theta)v(w,\theta)f_w(w|\theta)dw\,\mu(d\theta)$. By the envelope theorem:

$$\left.\frac{dv(w,\theta)}{d\kappa}\right|_{\kappa=0} = -\frac{\partial v}{\partial c}[\tau_z(z(w,\theta)) + \tau_x(x(w,\theta))].$$

Total revenue is $\mathcal{R} = \int_\Theta \int_W [T_z(z;\kappa) + T_x(x;\kappa)]f_w(w|\theta)dw\,\mu(d\theta)$. The Lagrangian $\mathcal{L} = \mathcal{W}/\lambda + \mathcal{R}$ responds to the reform as:

$$\left.\frac{d\mathcal{L}}{d\kappa}\right|_{\kappa=0} = \underbrace{\int_\Theta \int_W (1-g(w,\theta))[\tau_z(z)+\tau_x(x)]f_w dw\,\mu(d\theta)}_{\text{Mechanical effect net of welfare loss}} + \underbrace{\int_\Theta \int_W \mathcal{T}_z'(z)\left.\frac{dz}{d\kappa}\right|_{\kappa=0} f_w dw\,\mu(d\theta)}_{\text{Behavioral effect on income tax revenue}}$$

$$+ \underbrace{\int_\Theta \int_W \mathcal{T}_x'(x) \left.\frac{dx}{d\kappa}\right|_{\kappa=0} f_w dw\, \mu(d\theta)}_{\text{Behavioral effect on commodity tax revenue}},$$

where the social marginal utility of income is:

$$g(w,\theta) \equiv \frac{\gamma(w,\theta)}{\lambda}\frac{\partial v}{\partial c} + \mathcal{T}_z'(z)\eta_z(w,\theta) + \mathcal{T}_x'(x)\left[\eta_{x|z}(w,\theta) + \frac{\partial x}{\partial z}\eta_z(w,\theta)\right].$$

Change of Variables to (z,x)

Assume that, conditional on θ, $z(w,\theta)$ is continuous and monotonic in w. This ensures a continuous marginal distribution of z with density $h_z(z)$. However, the conditional distribution of x given z may have mass points, so we work with the CDF $H_x(x|z) = P(x(w,\theta) \leq x \mid z(w,\theta) = z)$ and corresponding measure $\nu_x(A|z) = \int_{x\in A} dH_x(x|z)$.

Define $\bar{\ell}(z,x) \equiv \mathbb{E}[\ell(w,\theta) \mid z(w,\theta) = z, x(w,\theta) = x]$ for $\ell \in \{g, \varepsilon_z, \varepsilon_{x|z}\}$. The welfare effect becomes:

$$\begin{aligned}
\left.\frac{d\mathcal{L}}{d\kappa}\right|_{\kappa=0} = &\int_{\mathbb{R}_{++}}\int_{\mathbb{R}_{++}} (1-\bar{g}(z,x))[\tau_z(z)+\tau_x(x)]\, dH_x(x|z)\, h_z(z)\, dz \\
&- \int_{\mathbb{R}_{++}}\int_{\mathbb{R}_{++}} \frac{\mathcal{T}_z'(z)\bar{\varepsilon}_z(z,x) + \mathcal{T}_x'(x)\mathbb{E}[x'_{\text{inc}}\varepsilon_z|z,x]}{1-\mathcal{T}_z'(z)} z\tau_z'(z)\, dH_x(x|z)\, h_z(z)\, dz \\
&- \int_{\mathbb{R}_{++}}\int_{\mathbb{R}_{++}} \left\{ \frac{\mathcal{T}_z'(z)\mathbb{E}[x'_{\text{inc}}\varepsilon_z|z,x] + \mathcal{T}_x'(x)\mathbb{E}[(x'_{\text{inc}})^2\varepsilon_z|z,x]}{1-\mathcal{T}_z'(z)} z \right. \\
&\qquad \left. + \frac{\mathcal{T}_x'(x)\bar{\varepsilon}_{x|z}(z,x)x}{1+\mathcal{T}_x'(x)} \right\} \tau_x'(x)\, dH_x(x|z)\, h_z(z)\, dz.
\end{aligned}$$

Vertically Neutral Reform

Consider a linear tax reform $\tau_x(x) = x$ with compensating income tax changes $\tau_z(z) = -\mathbb{E}[\tau_x(X)|Z=z] = -\bar{x}(z)$, which holds average tax payments constant at each income level. This yields $\tau_z'(z) = -\bar{x}'(z)$.

Preference Heterogeneity Sufficient Statistic

Define the preference heterogeneity sufficient statistic:

$$x'_{het}(z;\theta) \equiv \bar{x}'(z) - x'_{\text{inc}}(z;\theta).$$

This is the residual of the cross-sectional relationship between x and z after removing the causal income effect for type θ agents.

Optimal Tax Condition

For a linear tax $\mathcal{T}_x(x) = t_x x$, substituting the vertically neutral reform into the welfare effect and setting $\frac{d\mathcal{L}}{d\kappa}\big|_{\kappa=0} = 0$ yields:

$$\begin{aligned}
0 = &- \int_{\mathbb{R}_{++}} \mathbb{C}[g(z;\theta), x(z;\theta)|z] h_z(z)\, dz \\
&+ \int_{\mathbb{R}_{++}} \frac{\mathcal{T}_z'(z)}{1-\mathcal{T}_z'(z)} \mathbb{E}[x'_{het}(z;\theta)\varepsilon_z(z;\theta)|z] z h_z(z)\, dz \\
&+ t_x \int_{\mathbb{R}_{++}} \frac{1}{1-\mathcal{T}_z'(z)} \mathbb{E}[x'_{het}(z;\theta)x'_{inc}(z;\theta)\varepsilon_z(z;\theta)|z] z h_z(z)\, dz \\
&- \frac{t_x}{1+t_x} \mathbb{E}[x(z;\theta)\varepsilon_{x|z}(z;\theta)].
\end{aligned}$$

Under the assumption that $\mathbb{E}[\mathbb{C}(g(w,\theta), x(w,\theta)|z)] = 0$, the optimal tax satisfies equation (53):

$$t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda} = \frac{\mathbb{E}_{z,\theta}[\text{RR}(z;\theta)\mathcal{T}_z'(z)]}{1 - \mathbb{E}_{z,\theta}[\text{RR}(z;\theta)x'_{\text{inc}}(z;\theta)]},$$

where $\text{RR}(z;\theta) = \eta^{\text{Taste}}_{x,z}(z;\theta)\left(\frac{\varepsilon_z(z;\theta)\bar{x}(z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]}\right)\left(\frac{1+t_x}{1-\mathcal{T}_z'(z)}\right)$.

□

Proof of Corollary 1. Suppose $\varepsilon_z(z;\theta) = \bar{\varepsilon}_z(z)$ for all θ at each z. Then the covariance terms in the numerator and denominator of the general optimal tax condition simplify.

For the numerator, $\mathbb{C}[\varepsilon_z(z;\theta), x'_{inc}(z;\theta)|z] = 0$ by assumption, leaving only the average terms.

For the denominator, using $x'_{het} = \bar{x}' - x'_{inc}$:

$$\begin{aligned}
\mathbb{E}[x'_{\text{het}}x'_{\text{inc}}\varepsilon_z|z] &= \bar{\varepsilon}_z(z)\mathbb{E}[(\bar{x}'(z) - x'_{\text{inc}}(z;\theta))x'_{\text{inc}}(z;\theta)|z] \\
&= \bar{\varepsilon}_z(z)\left[\bar{x}'(z)\bar{x}'_{\text{inc}}(z) - \mathbb{E}[(x'_{\text{inc}})^2|z]\right] \\
&= \bar{\varepsilon}_z(z)\left[\bar{x}'(z)\bar{x}'_{\text{inc}}(z) - (\bar{x}'_{\text{inc}}(z))^2 - \mathbb{V}[x'_{\text{inc}}|z]\right] \\
&= \bar{\varepsilon}_z(z)\bar{x}'_{\text{het}}(z)\bar{x}'_{\text{inc}}(z) - \bar{\varepsilon}_z(z)\mathbb{V}[x'_{\text{inc}}|z].
\end{aligned}$$

Substituting into the general formula and collecting terms yields equation (56):

$$t_x - \frac{\mathcal{D}'(\bar{x})}{\lambda} = \frac{\mathbb{E}_z[\overline{\text{RR}}(z)\mathcal{T}_z'(z)]}{1 - \mathbb{E}_z[\overline{\text{RR}}(z)\bar{x}'_{\text{inc}}(z)] + \mathbb{E}_z[\widetilde{\text{RR}}(z)\mathbb{V}(x'_{\text{inc}}(z;\theta)|z)]},$$

where:

$$\overline{\text{RR}}(z) = \bar{\eta}^{\text{Taste}}_{x,z}(z) \frac{\bar{\varepsilon}_z(z)\bar{x}(z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]} \frac{1+t_x}{1-\mathcal{T}'_z(z)},$$

$$\widetilde{\text{RR}}(z) = \frac{\bar{\varepsilon}_z(z)z}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)x(z;\theta)]} \frac{1+t_x}{1-\mathcal{T}'_z(z)}.$$

The variance term $\mathbb{V}[x'_{\text{inc}}|z]$ appears in the denominator with a positive sign, which attenuates any deviation from Pigouvian taxation. Only when $\mathbb{V}[x'_{\text{inc}}|z] = 0$ for all z does the formula reduce to the unidimensional case. □