Australian National University

Crawford School of Public Policy

# TTPI Working Paper Series

**Tax and Transfer Policy Institute**

# Do Distributional Concerns Justify Lower Environmental Taxes?

## Working Paper 16/YEAR
## September 2026

**Ashley C. Craig**
Australian National University

**Thomas Lloyd**
EM Lyon Business School

**Dylan T. Moore**
University of Hawaii at Manoa

## Abstract
Should externality taxes be reduced if they are regressive? We introduce a model in which the government raises revenue from distortionary income and commodity taxes which also play an externality correction role. Whether the tax on the externality should be adjusted away from the Pigouvian level of marginal damage depends on why households differ in their consumption patterns. Differences in demand due to preference heterogeneity across the income distribution change the optimal tax, while differences due to income effects do not. We derive sufficient statistics for optimal policy, and use them to study the optimal tax on carbon-intensive consumption in the United States. Our empirical results suggest that the carbon tax should be set only slightly lower than the Pigouvian level. When we allow for heterogeneity in preferences at each income level as well as across the income distribution, our recommended carbon tax moves even closer to the Pigouvian benchmark.

Tax and Transfer Policy Institute

Crawford School of Public Policy

College of Law, Governance and Policy

+61 2 6125 9318

tax.policy@anu.edu.au

The Australian National University

Canberra ACT 0200 Australia

**www.anu.edu.au**

CRICOS Provider No. 00120C

**The Tax and Transfer Policy Institute** (TTPI) is an independent policy institute that was established in 2013 with seed funding from the federal government. It is supported by the Crawford School of Public Policy of the Australian National University.

TTPI contributes to public policy by improving understanding, building the evidence base, and promoting the study, discussion and debate of the economic and social impacts of the tax and transfer system.

**The Crawford School of Public Policy** is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

# Do Distributional Concerns Justify Lower Environmental Taxes?

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

[1]*Australian National University*
[2]*EM Lyon Business School*
[3]*University of Hawai'i at Mānoa*

September 2026

**Abstract**

Should externality taxes be reduced if they are regressive? We introduce a model in which the government raises revenue from distortionary income and commodity taxes which also play an externality correction role. Whether the tax on the externality should be adjusted away from the Pigouvian level of marginal damage depends on why households differ in their consumption patterns. Differences in demand due to preference heterogeneity across the income distribution change the optimal tax, while differences due to income effects do not. We derive sufficient statistics for optimal policy, and use them to study the optimal tax on carbon-intensive consumption in the United States. Our empirical results suggest that the carbon tax should be set only slightly lower than the Pigouvian level. When we allow for heterogeneity in preferences at each income level as well as across the income distribution, our recommended carbon tax moves even closer to 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, Idann Gidron, James Hines, Matthew Lilley, Ilias Matterne, Chris Murphy, Matthew Shapiro, Itai Sher, Damián Vergara, and Celine Wasem for comments and suggestions, as well as seminar participants at the University of Michigan, UMass Amherst, the University of Melbourne, the Australian National University (ANU), the 2025 IIPF Annual Congress, the 2025 National Tax Association Annual Meeting, the Australian Conference of Economists, and the 2026 Mannheim Taxation conference. 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). We acknowledge the use of AI (ChatGPT, Claude, Gemini and Refine.ink) for text editing and coding assistance, and proof verification.

# 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).[1] 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 trade-off between externality correction and distributional impacts.

The fact that carbon taxes are regressive may be an important reason why they are politically unpopular (Hassett et al., 2009; Grainger and Kolstad, 2010; Mathur and Morris, 2014; Metcalf, 1999; 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-intensive consumption 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 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. In the knife-edge case where any two households would allocate their spending in precisely the same way if they had the same income, the carbon tax is optimally set at the Pigouvian level of marginal damage (Kaplow, 2012). In this case, first-best externality correction remains optimal even in a second-best world. By contrast, if households differ in how they would choose to spend a given level of income, the optimal tax can change because there may then be a trade-off between the efficiency of redistribution and the extent of externality correction.

[1] Panelists in the Initiative on Global Markets (IGM) Economic Experts Panel have consistently endorsed carbon taxation as a preferred climate policy (IGM Forum, 2011, 2012, 2018, 2025).

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 her 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 tax on these carbon-intensive goods is slightly below 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. A carbon tax set at the Pigouvian level ensures that households spend their after-tax incomes efficiently. Starting from this level, raising the carbon tax distorts consumption choices because the externality is no longer exactly internalized. However, the reform also changes labor supply choices: Raising the carbon tax lowers labor supply, while the compensatory income tax reduction increases labor supply. If all households have identical preferences for carbon intensity, the two labor supply effects cancel out. This is because identical preferences imply that the combined reform to the income and carbon taxes 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 all 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. In the multidimensional case, adjustments to the income tax can no longer compensate every household for the introduction (or increase) of a carbon tax. The reason is simple: at a given level of income, households differ in how much of the carbon-intensive good they consume. Consequently, some are more exposed to dirty good tax changes than others, and a single income-based adjustment cannot make all of them whole.

We nonetheless derive an empirically tractable optimal tax condition for the multidimensional case. While the lessons from the unidimensional case carry over qualitatively, an 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 attenuation arises due to imperfect targeting. Because the planner cannot charge different carbon-tax rates to different types who earn the same income, the rate that is most efficient for one group is necessarily inefficiently high (or low) for another. This limits the potential efficiency gains that deviating from Pigouvian taxation can achieve.

Assuming that differences in preferences do justify deviating from the Pigouvian level of taxation, distributional concerns lead the government to trade off externality correction and redistribution. 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 becomes more attractive to use the carbon tax to achieve distributional aims. Using the carbon tax to redistribute is also appealing if carbon intensity is inelastic to changes in the tax on carbon. Conversely, 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. If nonlinear taxation is feasible, for example through nonlinear electricity and gas pricing, the optimal carbon tax may be progressive or regressive. There are many real-life examples of nonlinear 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 nonlinearity 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 case where preferences vary across but not within income levels. Our model provides a convenient way to identify differences in preferences between households with different status quo incomes. Following Ferey et al. (2024), we compare two statistics. First, we measure the cross-sectional relationship between income and consumption of carbon-intensive goods. This captures differences in consumption due to 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, leaving only the part that stems from preferences.

We implement this framework for the United States using survey evidence from 25,888 households in the Consumer Expenditure Surveys (CEX) and 1,551 respondents from a custom online survey. Our analysis focuses on the four expenditure categories that account for nearly all spending on goods with *direct* household emissions, aggregated into a single "dirty good": gasoline, natural gas, electricity, and heating fuels.[3] 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.

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.

[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.

[3]While an economy-wide carbon tax would also price emissions from other goods and services, these four categories account for 93.7% of expenditures on goods with direct household carbon emissions, and for 96.9% of the associated emissions. The remainder corresponds to air travel.

Next, we elicit respondents' causal consumption responses to a small permanent increase in household income. We find that the causal effect of additional income on carbon intensity slightly exceeds the cross-sectional slope we see in the Consumer Expenditure Survey. This implies that lower-income households have a relative preference for carbon-intensive goods. However, the gap is small, which indicates that cross-sectional variation in consumption of these carbon-intensive goods is largely explained by income effects, not preference heterogeneity. We note that this alignment between causal and cross-sectional responses is not mechanical or obvious. Indeed, the causal consumption responses are within-individual effects, which are compared to variation across individuals.

The interpretation of these survey responses hinges on the reliability of stated preferences. Growing evidence suggests that stated preferences perform well (List et al., 2006; Fuster and Zafar, 2023), yet reliability inevitably depends on context. In our setting, we have a way of supporting the reliability of our survey. Because we measure status quo allocations across different types of goods in both our survey and the Consumer Expenditure Survey, we can verify that 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 set the hypothetical income increase to be familiar and relevant for respondents (Fuster and Zafar, 2023).

Our empirical results allow us to estimate the optimal tax on carbon-intensive goods. Differences in preferences justify only small deviations from the Pigouvian prescription of setting the tax equal to marginal damage. We estimate that the Pigouvian level of the tax is about 49 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 3.3 cents below this. From a welfare perspective, there is almost no difference: Imposing a tax equal to marginal damage captures 99.61% of the gain from the optimal distributionally adjusted carbon tax (Appendix C). The small magnitude of this adjustment scales with the relative responsiveness of the two tax bases, namely the ratio of taxable income elasticity and the compensated demand elasticity for dirty good consumption. When we later account for heterogeneity within as well as across income levels, this further compresses the optimal tax toward its Pigouvian level.

If taxes on externality-generating goods can be nonlinear—for which there are some real-world precedents—there are slightly larger deviations from the Pigouvian prescription in some parts of the income distribution. For a nonlinear tax, we compare the cross-sectional and causal relationships between income and consumption at each income level. We find that the magnitude of the preference-heterogeneity component of variation in carbon intensity is larger at lower and middle incomes. At around $100,000, the estimated

taste-heterogeneity component is slightly negative. This results in an optimal tax of 44.7 cents, which is 4.6 cents lower than the Pigouvian level (-9.3%). Households with incomes closer to $250,000 would face an optimal carbon tax of 46.7 cents, which is only 2.5 cents lower than the Pigouvian level (-5.1%).

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 the optimal carbon tax stays relatively close to but below the Pigouvian level, and that the efficiency gains from using the carbon tax to redistribute are small. 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 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.

We build on an approach to optimal commodity taxation introduced by Allcott et al. (2019), which Ferey et al. (2024) later generalized and applied to savings taxation.[4] They highlight the core insight that causal income effects can be subtracted from cross-sectional variation in consumption to measure the degree to which preferences vary across income levels. We apply these ideas to shed light on how externality-correcting taxes should change with preference heterogeneity, including multidimensional heterogeneity; we then apply them to study carbon emissions by households, an important example 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). Other related recent work includes Aronsson and Johansson-Stenman (2026) and Djalalova and Rothschild (2025) who explore the implications of preference heterogeneity for optimal public goods expenditure.

Empirically, recent studies highlight significant heterogeneity in consumer preferences for carbon-intensive goods vs. cleaner alternatives. For example, Lyubich (2025) 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., 2010; Douenne et al., 2026; Barrage, 2020; Fried et al., 2024). Bierbrauer (2024) asks a closely related question in a multi-sector general equilibrium model, and finds that a redistributive case for sector-specific carbon prices arises only when tastes for green versus brown goods vary with income. We note also Douenne et al. (2025), who introduce incomplete markets and idiosyncratic labor-income risk and find that these new concerns justify only negligible deviations of the carbon tax from the social cost of carbon, with

[4]Both papers also discuss corrective motives, the first related to sin taxes, and the second to the underprovision of bequests. Our paper further contributes to this literature by providing an empirical application for household carbon consumption, as well as an alternative approach to characterizing the implications of multidimensional heterogeneity.

welfare gains from the adjustment of the same order of magnitude as ours.

We also do not consider multinational issues (see Aronsson and Blomquist, 2003) or explicitly model multiple generations (Fried et al., 2018). Nor do we consider corrective policy instruments beyond taxation. Nonetheless, our work complements Fricke et al. (2025), who show that a binding cap on carbon emissions lowers the efficiency cost of redistribution and makes the optimal income tax more redistributive.

# 2 Externality Correction With Taste Heterogeneity

Consider a population of individuals indexed by type $w \in W = \mathbb{R}_{++}$, where $w$ has differentiable cumulative distribution $F$. Individual $w$ chooses income $z \in \mathbb{R}_+$. Income has differentiable cumulative distribution $H_z(z)$ and density $h_z(z)$. Income is subject to nonlinear income taxation with tax schedule $T_z(z)$. She allocates her after-tax income, $z - T_z(z)$, between a numeraire consumption good, $c$, and an externality-generating good, $x$, which we refer to as the *dirty good*. A tax, $T_x(x)$, must be paid on consumption of the dirty good. We constrain $T_z(z)$ and $T_x(x)$ to be twice continuously differentiable.

We focus in this section on the case in which heterogeneity is unidimensional in the sense that $w \in W \subset \mathbb{R}$ 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 consumption goods in the sense that they would prefer different consumption bundles even if they were forced to have the same $z$ and $z - T(z)$. In Section 3, we further 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 dirty 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, increasing in $c$ and $x$, and decreasing in $z$. We further assume that $u(c, x, z \mid w)$ is strictly concave in $(c, x, z)$.

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

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

where:

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

is the amount of the dirty good a type $w$ agent would choose if her taxable income were $z$. The amount of the dirty good that a type $w$ agent actually chooses is $x\left(w\right) \equiv x\left(w; z\left(w\right)\right)$.

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

$$1 - T_z'\left(z\right) - \frac{\partial x\left(w; z\right)}{\partial z}\left(1 + T_x'\left(x\right)\right) = -\frac{u_z + \frac{\partial x(w;z)}{\partial z} u_x}{u_c} \tag{4}$$

$$\text{and} \qquad 1 + T_x'\left(x\right) = \frac{u_x}{u_c}. \tag{5}$$

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: $1 - T_z'\left(z\right) = -\frac{u_z}{u_c}$, but this masks important intuition as we highlight below.

Equation 4 highlights something important: an agent's anticipation of her own behavior at stage 2 enters her stage-1 decision. For each additional dollar of income she considers earning, she evaluates how it would change her disposable income and how much she would spend on good $x$, given her preferences and the tax rate. In this way, the agent can be thought of as choosing her level of taxable income subject to an *effective net of income tax rate* of $1 - T_z'\left(z\right) - \frac{\partial x(w;z)}{\partial z} T_x'\left(x\right)$. Thus, when a rational agent sees her purchasing power decline due to commodity taxes, this reduces her marginal benefit of earning income. The size of this reduction depends on what fraction of her 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, $T_z$ and $T_x$. It does so to jointly maximize social welfare, $\boldsymbol{\mathcal{W}}$, while raising enough revenue to cover an exogenous revenue requirement, $\bar{\mathcal{R}}$.

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

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

Here, $\gamma(w)$ are Pareto weights; and $\mathcal{D}(X)$ is a separable atmospheric externality, which is an increasing function of aggregate consumption of the dirty good.[5]

## 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 tax reforms 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 initial tax schedules $\mathcal{T}_x(x)$ and $\mathcal{T}_z(z)$, we ask how welfare and revenue change when we perturb them in directions $\tau_x(x)$ and $\tau_z(z)$ respectively. We assume that both $\tau_x(x)$ and $\tau_z(z)$ are twice continuously differentiable and compactly supported (or linear in Subsection 2.10). Letting $\kappa \in \mathbb{R}$ index the magnitude of the reform, the new tax schedules are:

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

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,

$$\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = \frac{1}{\lambda}\left.\frac{\partial \mathcal{W}}{\partial \kappa}\right|_{\kappa=0} + \left.\frac{\partial \mathcal{R}}{\partial \kappa}\right|_{\kappa=0}, \tag{9}$$

starting from the initial tax system. Here, $\lambda$ 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

[5] This is equivalent to having the harm from the dirty good entering individuals' utility in a separable way. To see the equivalence, consider $\tilde{u} \equiv u(c, x, z \mid w) - \widetilde{x}(X; w)$, then $\tilde{v}(w) = v(w) - \widetilde{x}(X; w)$ and $\mathcal{D}'(X) = \int \gamma(w)\widetilde{x}'(X; w)dF$. Our formulas are also locally unaffected if we define welfare using a social welfare function, $\mathcal{W} \equiv \int W(v(w))\, dF - \mathcal{D}(X)$, rather than Pareto weights.

improvements can be possible and thus, at the optimum, we must have:

$$\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = 0 \quad \text{for all } \tau_z(\cdot) \text{ and } \tau_x(\cdot). \tag{10}$$

## 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 of 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 an agent with taxable income $z$, $\tau_x(\hat{x}(z))$, we also change her income tax liability by a precisely offsetting amount, such that the agent can still afford her original bundle. That is to say, the income tax reform is defined by:

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

Notice, this implies marginal income tax rate changes of:

$$\tau_z'(z) = -\hat{x}'(z)\,\tau_x'(\hat{x}(z)), \tag{12}$$

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.

Thus for any distribution-neutral reform, if $\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} > 0$, then a reform in that direction is a Pareto improvement because positive gains in government revenue net of externality costs can be obtained without changes to individual agent utility. This allows for a small transfer to be made that makes all agents strictly better off.[6] Put another way, for any Pareto-efficient tax system, it must be the case that:

$$\left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = 0 \quad \text{for all } \tau_z(\cdot) \text{ and } \tau_x(\cdot) \text{ satisfying } \tau_z(z) = -\tau_x(\hat{x}(z)). \tag{13}$$

[6] We assume the costs of the externality are homogeneous across households, so that $\mathcal{D}'(X)/\lambda$ measures the cost of compensating any household for a marginal change in $X$. If externality costs are heterogeneous, a reform satisfying this condition need not be Pareto-improving in general.

## 2.6 Behavioral Responses

By construction, any distribution-neutral reform produces a change in relative prices without a change in agents' 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:

$$1 + \mathcal{T}_x' \left(\hat{x}\left(z\right)\right) + \kappa \tau_x' \left(\hat{x}\left(z\right)\right) = \frac{u_x}{u_c}. \tag{14}$$

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

$$\left.\frac{\partial \hat{x}(z)}{\partial \kappa}\right|_{z,\,\kappa=0} = -\frac{\varepsilon_{x|z}(z)\,\hat{x}(z)}{1+\mathcal{T}_x'(\hat{x}(z))}\tau_x'(\hat{x}(z)) \tag{15}$$

where:

$$\varepsilon_{x|z}(z) \equiv -\frac{1+\mathcal{T}_x'(\hat{x}(z))}{\hat{x}(z)}\left.\frac{\partial \hat{x}(z)}{\partial \mathcal{T}_x'}\right|_z^c \tag{16}$$

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:

$$1 - \mathcal{T}_z'\left(z\right) + \underbrace{\kappa \hat{x}'\left(z\right)\tau_x'\left(\hat{x}\left(z\right)\right)}_{\text{Reform to } T_z} - x_{\text{inc}}'(z)\left(1 + \mathcal{T}_x'\left(x\right) + \underbrace{\kappa\tau_x'\left(\hat{x}\left(z\right)\right)}_{\text{Reform to } T_x}\right) = -\frac{u_z + x_{\text{inc}}'(z)u_x}{u_c} \tag{17}$$

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 $\hat{w}\left(z\right)$ agent on her 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 her marginal return to earning income, given that some of her income will be spent on that good. After rearranging the left hand side of the equation, the agent's effective marginal return to earning income becomes:

$$1 - \mathcal{T}_z'\left(z\right) - x_{\text{inc}}'(z)\left(1 + \mathcal{T}_x'\left(x\right)\right) + \kappa\left(\hat{x}'\left(z\right) - x_{\text{inc}}'(z)\right)\tau_x'\left(\hat{x}\left(z\right)\right). \tag{18}$$

This 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 taste heterogeneity:

$$x'_{\text{het}}(z) \equiv \hat{x}'(z) - x'_{\text{inc}}(z). \tag{19}$$

As equation 18 shows, the change in the effective marginal income tax rate induced by the reform is proportional to this preference heterogeneity effect.

To build some intuition for this, consider what is implied by the fact that a distribution-neutral reform leaves every agent's utility constant (to first order) at her initial level of income. To achieve this for all agents, the compensation that accompanies an increase to the dirty-good tax must be based on the cross-sectional relationship between income and dirty-good consumption. Specifically, as income rises, the income-tax reduction which the reform provides will evolve according to how much the dirty-good tax liability changes: $\kappa\hat{x}'(z)\,\tau'_x(\hat{x}(z))$. By contrast, the manner in which such a reform impacts the incentive to earn income depends on the viewpoint of an agent who is considering marginally increasing her income. Such an agent anticipates increasing her consumption of the dirty good by the amount of the causal income effect, $x'_{\text{inc}}(z)$. Thus, the reform alters the change in dirty-good tax liability which the agent anticipates if she were to marginally increase her income in proportion to this causal effect: $\kappa x'_{\text{inc}}(z)\,\tau'_x(\hat{x}(z))$. The end result is that a distribution-neutral reform will, in general, either over- or under-compensate agents who deviate from their initially optimal choice. Specifically, their marginal incentive to earn income changes by $\kappa x'_{\text{het}}(z)\,\tau'_x(\hat{x}(z))$.

In the edge case where preferences are identical and weakly separable between consumption and labor, as in Atkinson and Stiglitz (1976), consumption bundles are entirely determined by income. Any agent contemplating earning some income level $z$ would choose exactly the consumption bundle chosen by the agents initially observed at that income. Cross-sectional variation in compensation via the reform thus coincides with the compensation needed to hold marginal earnings incentives fixed: i.e., $\hat{x}'(z) = x'_{\text{inc}}(z)$. Thus, $x'_{\text{het}}(z) \equiv 0$. No agent's income choice is disturbed by the joint reform.

Returning to the general case when $x'_{\text{het}}(z) \neq 0$, the changes to effective marginal income tax rates due to the reform do cause taxable incomes to change. Specifically:

$$\left.\frac{\partial z}{\partial \kappa}\right|_{\kappa=0} = \frac{z\varepsilon_z(z)}{1-\mathcal{T}'_z(z)} x'_{\text{het}}(z)\tau'_x(\hat{x}(z)), \tag{20}$$

where

$$\varepsilon_z(z) \equiv \frac{1-\mathcal{T}'_z(z)}{z} \left.\frac{\partial z}{\partial\left(1-\mathcal{T}'_z(z)\right)}\right|_z^c, \tag{21}$$

is the compensated elasticity of taxable income (ETI) with respect to the net-of-income-tax rate. Formal definitions of elasticities are included in the proof of Proposition 1.

Consumption of the Dirty Good (Indirect). The final behavioral response is the 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,

$$x'_{\text{inc}}(z)\frac{\partial z}{\partial \kappa}\bigg|_{\kappa=0} = x'_{\text{inc}}(z)\frac{z\varepsilon_z(z)}{1-\mathcal{T}'_z(z)}x'_{\text{het}}(z)\tau'_x(\hat{x}(z)), \tag{22}$$

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

$$\frac{\mathrm{d}x}{\mathrm{d}\kappa}\bigg|_{\kappa=0} = \frac{\partial \hat{x}(z)}{\partial \kappa}\bigg|_{z,\,\kappa=0} + x'_{\text{inc}}(z)\frac{\partial z}{\partial \kappa}\bigg|_{\kappa=0}. \tag{23}$$

## 2.7 Testing Local Pareto Efficiency: A Specific Reform

To characterize the set of Pareto-efficient tax systems, we proceed heuristically by considering a reform that raises the marginal dirty-good tax rate at a single income level. Our proofs in Appendix D provide a more formal treatment.[7] The reform raises the marginal dirty-good tax rate only for agents earning taxable income $z$:

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

The magnitude of the reform is carried by $\kappa$, as above. Distribution-neutrality then requires an offsetting reduction of those same agents' marginal income tax rates: by equation 12, $\tau'_z(z') = -\hat{x}'(z')$ if $z' = z$ and $\tau'_z(z') = 0$ otherwise.

## 2.8 Welfare Effects of the Reform

We obtain the welfare effect of this reform by multiplying each behavioral response (equations 15, 20, and 22) by its marginal impact on social welfare. For changes in income, this is the income tax rate $\mathcal{T}'_z(z)$. For changes in consumption of the dirty good (equations 15 and 22), this is the uninternalized portion of the marginal externality: the discrepancy between the marginal tax rate and the marginal social damage, $\mathcal{T}'_x(\hat{x}(z)) - \frac{\mathcal{D}'(X)}{\lambda}$. In total, the local impact on social welfare at income $z$ is the sum of three revenue effects:

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

where the components are defined as follows:

[7] In the appendix, we apply the fundamental lemma of the calculus of variations, similar to Jacquet and Lehmann (2021). Alternatively, our results can be formalized through a limiting argument—taking the point reform as the limit of reforms concentrated on a vanishing band around $z$—in the spirit of Gerritsen (2024).

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

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

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

The first of these effects, $\Delta\mathcal{R}^{z}$, captures the impact on government revenue from taxable income changes. The second term, $\Delta\mathcal{R}^{x|z}$, is the welfare effect of changes in consumption of the dirty good, holding income constant. Finally, $\Delta\mathcal{R}^{z\to x}$ is the welfare effect of further changes in consumption of $x$ that are induced indirectly by changes in taxable income.

Importantly, the effects that relate to taxable income changes ($\Delta\mathcal{R}^{z}$ and $\Delta\mathcal{R}^{z\to x}$) are only present in a model where preference heterogeneity implies that distribution-neutral reforms alter effective marginal income tax rates ($x_{\text{het}}'(z) \neq 0$). Absent this mechanism, we return to the separable benchmark of Atkinson and Stiglitz (1976), wherein there are no taxable income responses and only the own-price effect of the increased dirty-good tax rate, $\Delta\mathcal{R}^{x|z}$, remains. By itself, this term translates movements away from Pigouvian taxation into efficiency costs. The possibility that these costs might be offset by efficiency gains obtained via induced changes in taxable income is the reason that preference heterogeneity may lead the planner to deviate from a Pigouvian tax rate.

## 2.9 A Necessary Condition for Pareto Efficiency

Setting the local welfare effect $\Delta\mathcal{R}(z)$ to zero yields an optimal tax condition at each income level $z$. Because $z$ is arbitrary, Pareto efficiency requires that there be no net gain from a distribution-neutral reform at any income:

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

where $\eta_{x,z}^{\text{Taste}}\left(z\right)$ is the "taste elasticity": the portion of the elasticity of consumption of $x$ with respect to income that occurs through heterogeneity in agent tastes. In other words,

$$\eta_{x,z}^{\text{Taste}}\left(z\right) \equiv \frac{zx_{\text{het}}'(z)}{\hat{x}\left(z\right)}. \tag{30}$$

Equation 29 parallels the style of presentation of the analogous Pareto efficiency condition in Ferey et al. (2024). Rearranging yields Proposition 1.

**Proposition 1.** *If the income and dirty-good taxes are optimal and are allowed to be nonlinear, they must satisfy equation 31:*

$$\mathcal{T}_x'\left(\hat{x}\left(z\right)\right) = \frac{\mathcal{D}'(X)}{\lambda} + \frac{RR(z)\mathcal{T}_z'\left(z\right)}{1 - RR(z)x_{inc}'(z)} \tag{31}$$

*where:*

$$RR(z) \equiv \eta_{x,z}^{Taste}\left(z\right)\frac{\varepsilon_z\left(z\right)}{\varepsilon_{x|z}\left(z\right)}\frac{1+\mathcal{T}_x'\left(\hat{x}\left(z\right)\right)}{1-\mathcal{T}_z'\left(z\right)} \tag{32}$$

*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.*

Equation (31) reveals that the optimal tax structure retains the "additivity" principle of Sandmo (1975): The marginal social damage enters the optimal tax rate additively. However, the expression also contains a second term which adjusts the marginal tax rate to account for preference heterogeneity.

The taste elasticity, $\eta_{x,z}^{\text{Taste}}\left(z\right)$, is a sufficient statistic that determines the sign and scale of this new term. As noted above, in the absence of taste heterogeneity, i.e., $\eta_{x,z}^{\text{Taste}}\left(z\right) = 0$, the optimal corrective tax is the standard Pigouvian one, no matter how regressive the tax appears in the cross-section. But suppose that $\eta_{x,z}^{\text{Taste}}\left(z\right) < 0$, so that some of the regressivity of the tax is attributable to preference heterogeneity: Lower-income households have a stronger taste for the dirty good. In this scenario, a distribution-neutral reform that lowers the corrective tax rate while increasing the income tax rate will actually reduce the marginal effective income tax rate. It is this specific type of taste-based regressivity that creates a redistributive rationale for lowering corrective taxes.[8]

Conversely, if $\eta_{x,z}^{\text{Taste}}\left(z\right) > 0$, then preference heterogeneity is actually a force that pushes toward a higher optimal carbon tax. In such a case, increasing the tax on carbon while decreasing income taxes enhances income-earning incentives. This may paradoxically imply a redistributive motive for taxing carbon at a higher level than marginal damage even if the tax is regressive overall because income effects drive down the carbon intensity of households who have more income. Only the part of the overall regressivity/progressivity that relates to preference heterogeneity generates optimal tax implications.[9]

The magnitude of the optimal deviation from Pigouvian taxation depends on the relative size of key behavioral responses, as can be seen in the numerator of equation 31.

[8]Another way of seeing this is that, if $\eta_{x,z}^{\text{Taste}}\left(z\right) = 0$, there is no "tagging" role for carbon in the sense of Akerlof (1978): Because tastes for carbon-intensive goods do not vary along the income distribution, carbon intensity does not reveal anything about a household's type that is not already revealed by her income choice. Only if $\eta_{x,z}^{\text{Taste}}\left(z\right) \neq 0$ will taxing or subsidizing carbon be useful for redistribution.

[9]The discussion of the prior two paragraphs assumes that marginal income tax rates are positive. If they are negative over some income range, as in the phase-in region of the EITC, the implications of the sign of the taste elasticity for the direction of the deviation flip.

All else equal, a stronger demand response, $\varepsilon_{x|z}$, implies that the efficiency costs associated with deviations from Pigouvian taxation are greater. This pushes the optimal tax toward the Pigouvian level. By contrast, a higher value of the elasticity of income ($\varepsilon_z$) indicates that the efficiency costs of income taxation are greater and, all else equal, favors a larger deviation from Pigouvian taxation in a bid to offload some revenue raising responsibility onto the more (marginally) efficient tax instrument. The optimal deviation from Pigouvian taxation is also proportional to the marginal income tax rate, since the efficiency costs of income taxation are increasing in the marginal rate.

The denominator of equation 31 could reverse the sign of the preference heterogeneity term, but does not do so in empirically relevant cases for carbon taxation (see Section 4). Rather, it amplifies positive deviations from the Pigouvian level, and dampens negative deviations. To see why, suppose that the taste elasticity is positive, $\eta_{x,z}^{\text{Taste}}(z) > 0$, motivating a carbon tax greater than the Pigouvian level. Raising the carbon tax together with a compensating income tax reduction reduces the effective marginal income tax rate, increasing incomes. If the dirty good is a normal good, $x'_{\text{inc}}(z) > 0$, then this implies that consumption of the carbon-intensive good increases as well. Because we know the optimal tax is above the marginal externality in this case, $\mathcal{T}'_x(\hat{x}(z)) > \frac{\mathcal{D}'(X)}{\lambda}$, we know that this increased dirty-good consumption creates an additional net social benefit, since the marginal revenue gained exceeds the marginal damage.

Notably, Proposition 1 leaves open the possibility that a negative externality is optimally subsidized, or a positive externality is taxed. This occurs when the taste elasticity has the opposite sign of the marginal damage, and the implied preference heterogeneity adjustment exceeds the marginal damage in magnitude. This is more likely the smaller is the marginal externality and the greater is the magnitude of the taste elasticity, or the relative responsiveness of income $\varepsilon_z/\varepsilon_{x|z}$.[10]

A similar approach can be used to derive conditions characterizing the optimal levels of the externality and income taxes. Unlike the distribution-neutral reform we considered here, the optimal level involves trade-offs between individuals and requires assumptions about welfare weights. We discuss these conditions in Appendix B.

[10]Such reversals are always possible in the case of positive externalities, but not in the case of negative externalities. Provided that $\text{RR}(z)x'_{\text{inc}}(z) < 1$, as $\text{RR}(z) \to -\infty$ the deviation in equation (31) approaches $-\mathcal{T}'_z(z)/x'_{\text{inc}}(z)$. Subsidizing a negative externality therefore requires $\mathcal{D}'(X)/\lambda < \mathcal{T}'_z(z)/x'_{\text{inc}}(z)$, regardless of the strength of preference heterogeneity. The intuition behind this result is straightforward: a deviation of $-\mathcal{T}'_z(z)/x'_{\text{inc}}(z)$ is the level at which the effective income tax approaches zero. Further subsidization past this point can never achieve any further efficiency gains.

### 2.10 Restricting to Linear Taxation

Nonlinear pricing of carbon-intensive goods has real-world precedents for goods such as electricity. However, this may not be true in all settings or for all externality-generating goods. We therefore also consider the optimal linear tax and the extent to which this deviates from the level of marginal damage.

If the tax on the dirty good must be linear ($\mathcal{T}_x'(\hat{x}(z)) = t_x$), the Pareto efficiency condition for the linear tax rate is similar to the condition for the nonlinear corrective tax (equation 31). However, it averages over all levels of income $z$ and consumption of $x$. Specifically, the optimal linear tax is characterized by Proposition 2.

**Proposition 2.** *If the income and dirty-good taxes are optimal and the corrective tax is restricted to being linear, they must satisfy equation 33:*

$$t_x = \frac{\mathcal{D}'(X)}{\lambda} + \frac{\mathbb{E}_z\left[RR(z)\,\mathcal{T}_z'(z)\right]}{1-\mathbb{E}_z\left[RR(z)\,x_{inc}'(z)\right]} \tag{33}$$

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

$$RR(z) \equiv \eta_{x,z}^{Taste}(z)\,\frac{\varepsilon_z(z)\,\hat{x}(z)}{\mathbb{E}_z\left[\varepsilon_{x|z}(z)\,\hat{x}(z)\right]}\left(\frac{1+t_x}{1-\mathcal{T}_z'(z)}\right), \tag{34}$$

*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-good 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 33.

## 3 Multidimensional Heterogeneity

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.[11] We abstracted from this type of heterogeneity in our unidimensional model, but now extend the model to allow for it.

Within-income heterogeneity makes exact compensation through the income tax impossible: When households with the same income consume different amounts of the carbon-intensive good, any income-based adjustment will give some households too much compensation and others too little. Consequently, we consider the closest available analogue of the distribution-neutral reforms of Section 2: *vertically neutral* reforms, which

[11] 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 that case.

leave the average tax liability at each level of income unchanged. Vertically neutral reforms eliminate mechanical redistribution between income levels, but not within them.

Our results provide a characterization of the optimal tax system that is broadly similar to that of Section 2, but augmented in several ways to account for imperfect compensation. We show that, all else equal, within-income heterogeneity in $x'_{\text{inc}}$ results in an attenuation of the optimal deviation from the Pigouvian benchmark.

To simplify the discussion, we focus on the case of linear taxation of the dirty good. To adapt the model to incorporate multidimensional heterogeneity, let each agent have type $(w, \theta) \in \mathbb{R}_{++} \times \Theta$. Type $(w, \theta)$ agents make choices to maximize their utility $u(c, x, z; w, \theta)$. We largely maintain the assumptions of the unidimensional case in Section 2.3, which ensure that choices of $z(w, \theta)$ and $x(w, \theta)$ are unique and respond smoothly to reforms. In addition, consistent with Jacquet and Lehmann (2021), we assume that the mapping $w \mapsto z(w, \theta)$ is strictly monotone. This ensures that the conditional inverse, $\hat{w}(z; \theta)$, is well-defined. We further assume that, for all $z$, the conditional mean, $\bar{x}(z) \equiv \mathbb{E}[x(w, \theta)|z(w, \theta) = z]$, exists and is continuously differentiable.

## 3.1 Vertically Neutral Reforms

Formally, we define a vertically neutral tax reform as a pair of reform functions $\tau_x$ and $\tau_z$ such that, for each income level $z$:

$$\tau_z(z) = -\mathbb{E}[\tau_x(x(w, \theta))|z(w, \theta) = z]. \tag{35}$$

In the linear tax case, this is simply $\tau_z(z) = -\bar{x}(z)$, with an implied reform to marginal income tax rates of $\tau'_z(z) = -\bar{x}'(z)$.

The welfare effect of a vertically neutral reform is:

$$\begin{aligned} \left.\frac{\partial \mathcal{L}}{\partial \kappa}\right|_{\kappa=0} = & \underbrace{\int \mathbb{E}\left[(1 - g(z; \theta)) \cdot (\hat{x}(z, \theta) - \bar{x}(z)) \mid z\right] h_z(z)\, dz}_{\text{Welfare effect from mechanical redistribution}} \\ & + \underbrace{\int \left[\overline{\Delta\mathcal{R}}^{x|z}(z) + \overline{\Delta\mathcal{R}}^{z}(z) + \overline{\Delta\mathcal{R}}^{z \to x}(z)\right] h_z(z)\, dz}_{\text{Revenue effects from compensated behavioral responses}}, \end{aligned} \tag{36}$$

where $g(z; \theta)$ is the marginal social welfare weight of a type $\theta$ household with income $z$, and $\hat{x}(z, \theta) \equiv x(\hat{w}(z; \theta), \theta)$ is her consumption of the dirty good.

The first term is the direct consequence of imperfect compensation, and it has no counterpart in Section 2. Full compensation would require adjusting each household's tax liability in line with her own consumption of the dirty good. Instead, a vertically neutral reform adjusts it in line with average consumption at the household's income level. The discrepancy between the two is a change in tax liability of $(\hat{x}(z, \theta) - \bar{x}(z))\, \kappa$. Households that

consume more than average at their income level are undercompensated, and households that consume less are overcompensated. The first term values this within-income redistribution using the welfare weights. Equivalently, it equals $-\int \mathbb{C}\left[g(z;\theta), \hat{x}(z,\theta) \mid z\right] h_z(z)\, dz$: what matters is not within-income dispersion in consumption per se, but whether that dispersion covaries with the welfare weights.

The second term in equation 36 collects the same three compensated behavioral responses as in Section 2.6, averaged over the types present at each income level:

$$\overline{\Delta\mathcal{R}}^{x|z}(z) \equiv \mathbb{E}\left[-\left(t_x - \frac{\mathcal{D}'(X)}{\lambda}\right)\frac{\varepsilon_{x|z}(z;\theta)\hat{x}(z,\theta)}{1+t_x}\,\middle|\, z\right] \tag{37}$$

$$\overline{\Delta\mathcal{R}}^{z}(z) \equiv \mathbb{E}\left[\mathcal{T}_z'(z)\frac{\varepsilon_z(z;\theta)z}{1-\mathcal{T}_z'(z)}x'_{\text{het}}(z;\theta)\,\middle|\, z\right] \tag{38}$$

$$\overline{\Delta\mathcal{R}}^{z\to x}(z) \equiv \mathbb{E}\left[\left(t_x - \frac{\mathcal{D}'(X)}{\lambda}\right)x'_{\text{inc}}(z;\theta)\frac{\varepsilon_z(z;\theta)z}{1-\mathcal{T}_z'(z)}x'_{\text{het}}(z;\theta)\,\middle|\, z\right] \tag{39}$$

As before, $\overline{\Delta\mathcal{R}}^{x|z}$ captures changes in consumption of $x$ directly caused by the reform to $t_x$; $\overline{\Delta\mathcal{R}}^{z}$ captures the impact on government revenue from taxable income choices; and $\overline{\Delta\mathcal{R}}^{z\to x}$ is the effect of further changes in consumption of $x$ induced indirectly by changes in income. The objects inside each average are those of Section 2.6, indexed by $\theta$. In particular, the elasticities $\varepsilon_{x|z}(z;\theta)$ and $\varepsilon_z(z;\theta)$ and the causal income effect $x'_{\text{inc}}(z;\theta)$ may now vary across households with the same income.

As in Section 2, preference heterogeneity ensures that the tax reform we study will create changes in effective marginal income tax rates, due to a discrepancy between required and provided income tax compensation in response to a change in the dirty-good tax rate. In particular, the income tax offset reduces income tax liability at a marginally higher level of income by $\bar{x}'(z)\kappa$. But the reform increases consumption tax liability at a slightly higher income by $\kappa x'_{\text{inc}}(z;\theta)$ in the case of a type $\theta$ agent. The difference,

$$x'_{\text{het}}(z;\theta) \equiv \bar{x}'(z) - x'_{\text{inc}}(z;\theta), \tag{40}$$

is the type-specific analogue of the taste heterogeneity component of Section 2.6. A type $\theta$ agent with income $z$ sees her effective marginal income tax rate change by $-\kappa x'_{\text{het}}(z;\theta)$, and the taxable income responses this creates are what generate the welfare effects in equations (38) and (39).

## 3.2 Heterogeneity in Behavioral Responses

Because the statistics inside each conditional average vary jointly with $\theta$, the averaged behavioral responses depend not only on the average value of each statistic at a given income level, but also on the covariances between them.

Consider $\overline{\Delta\mathcal{R}}^{z}(z)$, the income tax revenue that results from the combined reform to $\mathrm{d}t_x$ and the income tax. The reform reduces a type $\theta$ agent's effective tax rate on additional income by $x'_{\text{het}}(z;\theta)\mathrm{d}t_x$. This varies across agents to the extent that there is heterogeneity in the causal income effect, $x'_{\text{inc}}(z;\theta)$. Then the response to that change varies with the ETI, $\varepsilon_z(z;\theta)$. We therefore cannot simply replace each behavioral parameter with its mean. Instead, we can write:

$$\overline{\Delta\mathcal{R}}^{z}(z) = \mathcal{T}'_z(z) \cdot \frac{\bar{\varepsilon}_z(z)z}{1-\mathcal{T}'_z(z)} \cdot \bar{x}'_{\text{het}}(z) + \frac{\mathcal{T}'_z(z)z}{1-\mathcal{T}'_z(z)} \mathbb{C}\left[\varepsilon_z(z;\theta), x'_{\text{het}}(z;\theta)|z\right] \tag{41}$$

where $\bar{x}'_{\text{het}}(z)$ and $\bar{\varepsilon}_z(z)$ are the average values of $x'_{\text{het}}(z;\theta)$ and $\varepsilon_z(z;\theta)$ for agents at income level $z$, and $\mathbb{C}$ indicates a covariance. Using only the first term would amount to applying the unidimensional formula to average statistics. The covariance is a correction that takes into account within-income heterogeneity.

Turning to the indirect consumption effect, $\overline{\Delta\mathcal{R}}^{z\to x}(z)$, we can again rewrite this as a function of the average values of each statistic and their covariances:

$$\begin{aligned}
\overline{\Delta\mathcal{R}}^{z\to x}(z) = &\left(\frac{t_x - \frac{\mathcal{D}'(X)}{\lambda}}{1-\mathcal{T}'_z(z)}\right) z\bar{\varepsilon}_z(z)\bar{x}'_{\text{inc}}(z)\bar{x}'_{\text{het}}(z) \\
&+ \left(\frac{t_x - \frac{\mathcal{D}'(X)}{\lambda}}{1-\mathcal{T}'_z(z)}\right) z\bar{\varepsilon}_z(z)\mathbb{C}\left[x'_{\text{inc}}(z;\theta), x'_{\text{het}}(z;\theta)|z\right] \\
&+ \left(\frac{t_x - \frac{\mathcal{D}'(X)}{\lambda}}{1-\mathcal{T}'_z(z)}\right) z\,\mathbb{C}\left[\varepsilon_z(z;\theta), x'_{\text{inc}}(z;\theta)x'_{\text{het}}(z;\theta)|z\right].
\end{aligned} \tag{42}$$

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 again involve covariances. These arise 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 $(w,\theta)$ 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 $x'_{\text{inc}}(z;\theta)$. The definition of $x'_{\text{het}}(z;\theta)$ in terms of $x'_{\text{inc}}(z;\theta)$ means that the magnitudes of these forces are *mechanically* related at the individual level. As a result, the first covariance collapses to the negative variance of causal income effects at $z$:

$$\mathbb{C}\left[x'_{\text{inc}}(z;\theta), x'_{\text{het}}(z;\theta) \mid z\right] = -\mathbb{V}\left[x'_{\text{inc}}(z;\theta) \mid z\right]. \tag{43}$$

Intuitively, agents with a larger $x'_{\text{het}}(z;\theta)$ experience a greater than average change in their effective marginal income tax rates, and those same agents have a smaller consumption response to any given increase in their incomes compared to the average.

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 ETI, $\varepsilon_z(z;\theta)$.

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 non-zero unless causal income effects are constant across individuals with the same income. This leads to attenuation of the indirect consumption effect.

### 3.3 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 36) 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 dirty good conditional on income, at least on average. That is, we assume that:

$$\mathbb{E}\left[\mathbb{C}(g(z;\theta), \hat{x}(z,\theta) \mid z)\right] = 0. \tag{44}$$

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. The assumption is exactly the condition under which the mechanical-redistribution term in equation 36 vanishes. It makes the multidimensional condition empirically implementable at the cost of abstracting from differences in levels of well-being between households with the same income, which are potentially relevant if correlated with carbon intensity.

Assuming that the effect on welfare of redistribution within income levels is negligible on average, rearranging equation 36 yields the following characterization of the optimal linear tax on the dirty good.

**Proposition 3.** *Assume that welfare weights are uncorrelated with consumption of the dirty good conditional on income:*

$$\mathbb{E}\left[\mathbb{C}(g(z;\theta), \hat{x}(z,\theta) \mid z)\right] = 0. \tag{45}$$

*Then the optimal linear tax on the dirty good satisfies:*

$$t_x - \frac{\mathcal{D}'(X)}{\lambda} = \frac{\mathbb{E}_z\left[\mathbb{E}_\theta\left[RR(z;\theta)\mathcal{T}'_z(z) \mid z\right]\right]}{1 - \mathbb{E}_z\left[\mathbb{E}_\theta\left[RR(z;\theta)x'_{inc}(z;\theta) \mid z\right]\right]}, \tag{46}$$

*where:*

$$RR(z;\theta) \equiv \eta^{Taste}_{x,z}(z;\theta)\left(\frac{\varepsilon_z(z;\theta)\bar{x}(z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)\hat{x}(z,\theta)]}\right)\left(\frac{1+t_x}{1-\mathcal{T}'_z(z)}\right), \tag{47}$$

*and the type-specific taste elasticity is:*

$$\eta^{\textit{Taste}}_{x,z}(z;\theta) \equiv \frac{z x'_{\textit{het}}(z;\theta)}{\bar{x}(z)} = \frac{z(\bar{x}'(z) - x'_{\textit{inc}}(z;\theta))}{\bar{x}(z)}. \tag{48}$$

Equation (46) closely parallels the optimal linear tax condition from the unidimensional case (Proposition 2), but with expectations taken over both income levels and $\theta$ types within each income level. The taste elasticity $\eta^{\text{Taste}}_{x,z}(z;\theta)$ now measures how a type $\theta$ individual's causal income effect diverges from the overall cross-sectional relationship between income and consumption.

## 3.4 Decomposing the Multidimensional Optimal Tax Condition

To guide our empirical implementation, we can decompose the expectations in equation 46 into average effects and within-income covariance terms. This follows the same logic as when we decomposed the averaged behavioral responses in Section 3.2.

First, the numerator of the right-hand side of equation 46 can be written as the expectation of $\frac{(1+t_x)\,\mathcal{T}'_z(z)}{1-\mathcal{T}'_z(z)}\,\mathbf{N}(z)$ where:

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

Here, $\bar{\eta}^{\text{Taste}}_{x,z}(z)$ and $\bar{\varepsilon}_z(z)$ are averages across $\theta$ 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 ETI and the causal income effect on $x$.

Consider a population where, conditional on $z$, agents with higher $\varepsilon_z$ also have higher $x'_{\text{inc}}$ (and thus lower $\eta^{\text{Taste}}_{x,z}$). This implies that the agents that respond most strongly to marginal income tax changes also experience the smallest changes in their effective marginal income tax rate from a vertically neutral reform. As a result, the aggregate impact of a vertically neutral reform on taxable income is lower than would be expected based on the average ETI and the average taste elasticity.

The denominator of the RHS of 46 can be written as $1 - \mathbb{E}_z\big[\frac{1+t_x}{1-\mathcal{T}'_z(z)}\,\mathbf{D}(z)\big]$ where:

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

This covariance reflects the correlation between the change in the effective marginal income tax rate induced by the reform and the effect of that change on dirty-good consumption through increased income. The first is governed by $x'_{\text{het}}(z;\theta) = \bar{x}'(z) - x'_{\text{inc}}(z;\theta)$, which at a given income level is a constant less the household's causal income effect; the

second is the product of the induced change in earnings, proportional to $\varepsilon_z(z;\theta)$, and the rate at which those earnings become dirty-good consumption, which is again $x'_{\text{inc}}(z;\theta)$.

Because $x'_{\text{inc}}$ enters both channels, the covariance contains a variance term that pushes it to be positive, which increases the denominator and dampens any optimal deviation from the Pigouvian level: the households whose earnings incentives are most affected by the reform are precisely those for whom the resulting change in earnings does least to raise dirty-good consumption. When the elasticity of taxable income does not vary with $\theta$, the covariance reduces exactly to this variance term, which is the $\text{Var}(x'_{\text{inc}}(z;\theta) \mid z)$ appearing in Corollary 1. Otherwise it is reinforced when households with a high $x'_{\text{inc}}(z;\theta)$ also have a high $\varepsilon_z(z;\theta)$, and offset or overturned when they have a low one.

## 3.5 A Tractable Case: Constant Elasticity of Taxable Income

In our empirical application, we assume that the ETI does not vary with $\theta$ at each income level. Corollary 1 presents the optimal tax condition for this empirically implementable case. As expected from Section 3.2, the condition features the variance of $x'_{\text{inc}}$. Holding all other empirical objects constant, this extra term attenuates any deviation of the optimal tax on $x$ from the Pigouvian level.[12] 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 (49) simplify to the unidimensional optimal linear tax condition in Proposition 2.

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

$$t_x - \frac{\mathcal{D}'(X)}{\lambda} = \frac{\mathbb{E}_z\left[\overline{\text{RR}}(z)\mathcal{T}'_z(z)\right]}{1 - \mathbb{E}_z\left[\overline{\text{RR}}(z)\bar{x}'_{\text{inc}}(z)\right] + \mathbb{E}_z\left[\widetilde{\text{RR}}(z)\mathbb{V}(x'_{\text{inc}}(z;\theta) \mid z)\right]}, \tag{49}$$

where:

$$\overline{\text{RR}}(z) \equiv \bar{\eta}^{\text{Taste}}_{x,z}(z)\left(\frac{\bar{\varepsilon}_z(z)\bar{x}(z)}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)\hat{x}(z,\theta)]}\right)\left(\frac{1+t_x}{1-\mathcal{T}'_z(z)}\right), \tag{50}$$

$$\widetilde{\text{RR}}(z) \equiv \frac{\bar{\varepsilon}_z(z)z}{\mathbb{E}[\varepsilon_{x|z}(z;\theta)\hat{x}(z,\theta)]}\frac{1+t_x}{1-\mathcal{T}'_z(z)}, \tag{51}$$

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 ETI is constant. First, we require the average value of several statistics at each income level: the

[12] A related result appears in Taubinsky and Rees-Jones (2018) and Farhi and Gabaix (2020). In these papers, inattention to a commodity tax creates a corrective motive for taxation because households consume more of the taxed good than would maximize their utility. The optimal commodity tax is therefore above the level under full attention, as part of the role of the tax is to reduce over-consumption. Heterogeneity in attention makes this correction imperfect, and attenuates the corrective component of the optimal tax.

cross-sectional slope, $\bar{x}'(z)$; the causal income effect, $\bar{x}'_{\text{inc}}(z)$; the ETI, $\bar{\varepsilon}_z(z)$; and the demand elasticities $\varepsilon_{x|z}(z;\theta)$, which enter through their consumption-weighted average. 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, 33, and 49 allow us to quantify the optimal carbon tax for the United States while taking into account preference heterogeneity. 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 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, $\eta^{\text{Taste}}_{x,z}(z)$. 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 19. This requires an estimate of $\hat{x}'(z)$, which we obtain from cross-sectional variation in the CEX; and an estimate of $x'_{\text{inc}}$ 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 measure the cross-sectional relationship between income and consumption of carbon-intensive goods ($\hat{x}'(z)$) using the Consumer Expenditure Surveys. Specifically, we use information from the public use microdata (PUMD) files of the 2018–2024 CEX surveys from the United States Bureau of Labor Statistics. We rely on the Interview survey, a rotating panel survey with approximately 6,000 observations per quarter, providing detailed data on expenditures, income, and characteristics of consumers. We restrict the CEX sample to respondents aged 25 to 54 to align it with our custom survey sample. This results in a sample composed of 59,238 quarterly observations across 25,888 households.[13] Because of the rotating nature of the panel, households may be sampled between one and four consecutive times (quarters) across the sample period (2018–2024).

We measure expenditure levels for two carbon-intensive expenditure categories: (i) an aggregate dirty good, and (ii) gasoline and other motor oils. Aggregate dirty-good

[13] See the detailed definition of a "consumer unit" at: https://www.bls.gov/cex/csxfaqs.htm#qc1.

consumption is defined as the sum of expenditure on electricity, gasoline and motor oil, heating fuels, and natural gas. These categories account for 93.7% of expenditures on goods with direct household carbon emissions. We convert quarterly expenditures to 2024 dollars and annualize them. Appendix A.1 shows the share of total consumption accounted for by these and other categories.

Figure 1 shows that consumption of carbon-intensive goods (in levels) increases with household taxable income but at a decreasing rate. The income-percentile bins show this cross-sectional relationship transparently.[14] We also fit the household-level microdata across years with a Poisson regression quadratic in the log of (centered) taxable income to recover a smoothed estimate of the *slope* of the cross-sectional relationship $\hat{x}'(z)$. See Appendix Table A.3 for details. Panel (a) of Figure 1 shows the average *dirty good* expenditure at each point in the income distribution. Panel (b) shows the same measure for *gasoline* only. Both the binning and fitting approaches abstract away from any within-income heterogeneity in dirty-good consumption. In Section 4.3.3, we relax this restriction.

Appendix Figure A.1 reports the corresponding expenditure shares (as a percent of total expenditures). 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 12.5% at the bottom of the income distribution to a low of approximately 3.9% at the very top. Focusing on gasoline only, displayed in panel (b), the expenditure share ranges from 5.7% at the bottom of the income distribution to 1.8% at the very top.

#### 4.1.2 Measuring Income Effects Using a Hypothetical Choice Survey

Next, we measure the causal response of additional income, $x'_{\text{inc}}$. 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*.[15] The sample comprises 1,551 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.8).[16]

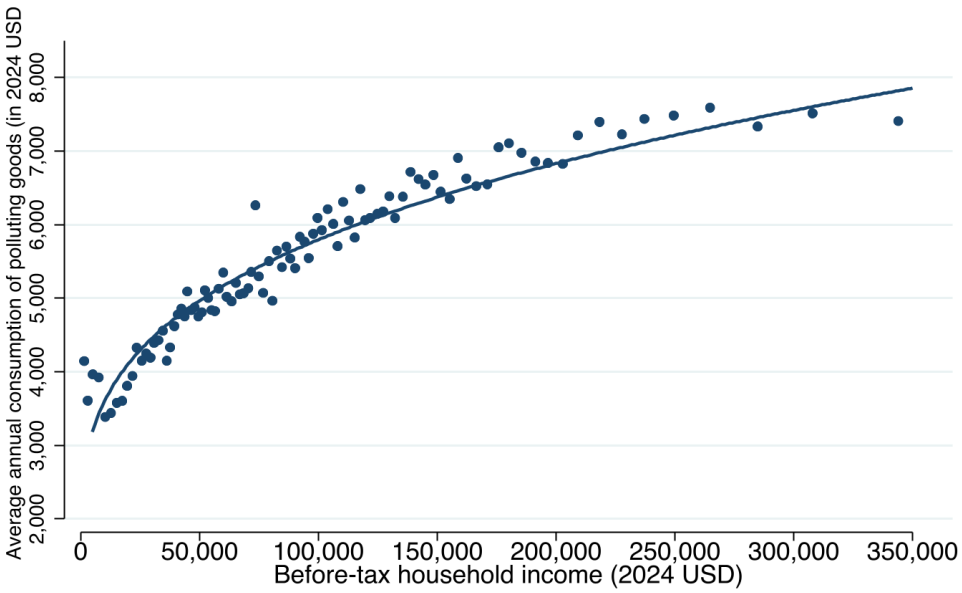
The primary purpose of the survey is to elicit how each household would respond to an unexpected increase in household (after-tax) income. Respondents were asked to allocate

[14] We assign each household to a year-specific income percentile (because taxable income is measured annually). We further decompose the top 1% of earners into a 99.9th percentile bin, corresponding to the top 0.1% of CEX respondents and leading to a total of 101 percentile bins each year. We then average consumption and income across year-bin cells first, using CEX-provided survey weights to ensure the sample is representative of the US population, and then across years.

[15] 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.

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

*(a)* Aggregate Dirty Good




*(b)* Gasoline

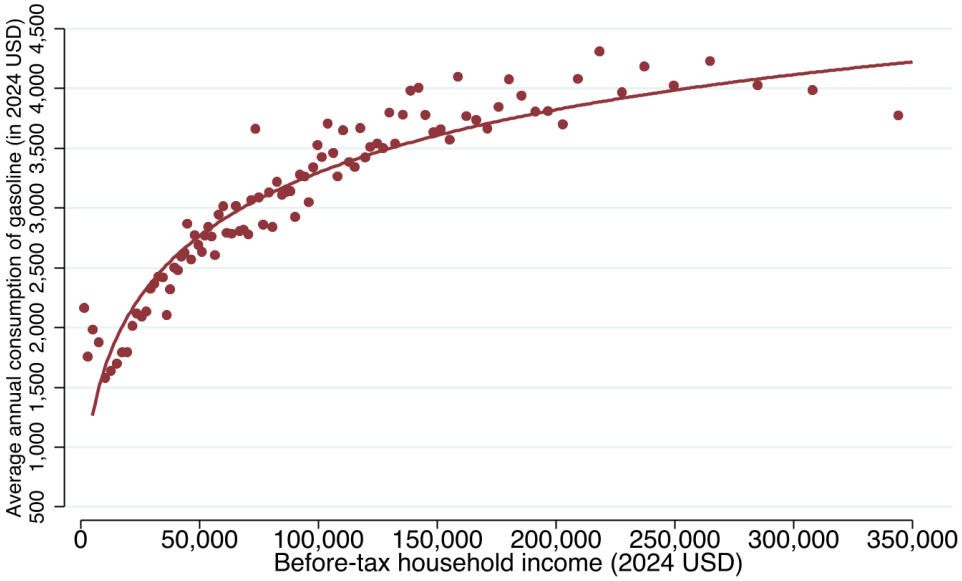



*Figure 1.* Consumption expenditure of carbon-intensive goods across the income distribution

**Note:** These binscatter plots show (i) survey-weighted average annualized expenditure, and (ii) corresponding average household taxable income within each income percentile bin (and a top 0.1% bin). The solid curve shows the centered-quadratic Poisson fit estimated from household-level data. 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 figure is truncated at $350k for readability.

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.[17] 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 their expenditures.

Response Quality. We took several steps to ensure the quality of responses. We screened out respondents based on (i) past studies' rejection rates, and (ii) the number of hours spent on the platform. In line with our pre-registration plan, we filter out participants who fail attention checks, complete the survey implausibly quickly, are extreme outliers, or provide inconsistent or nonsensical information regarding incomes or marginal propensities to consume (see Appendix Table A.9 for a breakdown of sample exclusions).

In addition, we have a powerful way to validate the representativeness of our survey sample. Figure A.2 plots the relationship between our respondents' reported status-quo consumption shares and their incomes. This relationship can be compared to that in the Consumer Expenditure Survey (Figure A.1). The figures align closely, suggesting that our responses are of high quality and our respondents quite representative.

Estimating Causal Income Effects. Our estimates for $x'_{\text{inc}}$ build on our survey respondents' marginal consumption choices in response to a hypothetical increase in their incomes. 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:

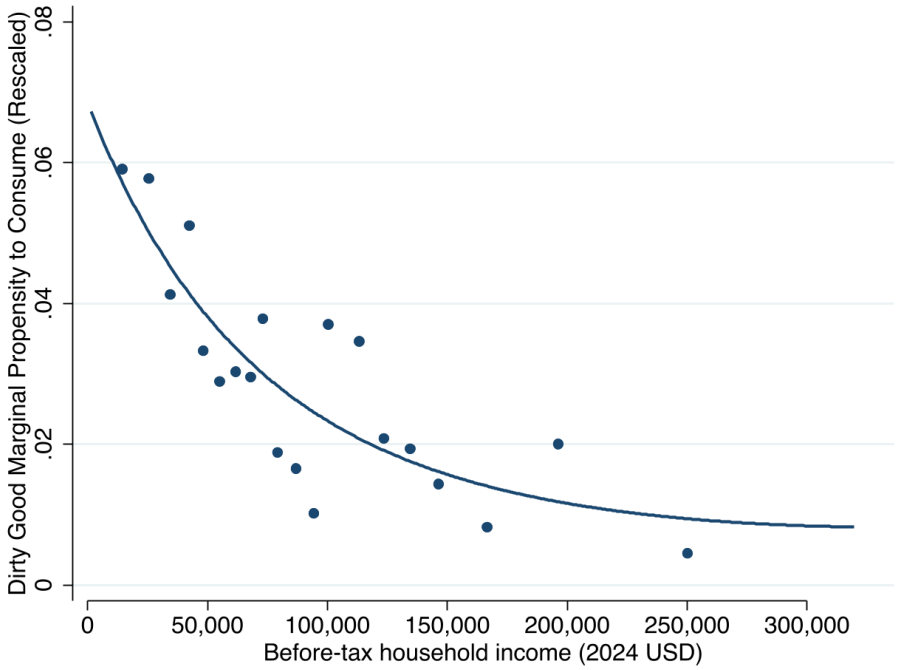
$$\frac{dx\left(\hat{w}\left(z\right);z\right)}{dI} = \eta_{x|z}(z) + \frac{\partial x\left(\hat{w}\left(z\right);z\right)}{\partial z}\,\eta_z(z), \tag{52}$$

where $\eta_z$ and $\eta_{x|z}$ are income effects due to a reform that lowers the tax burden slightly, while leaving both marginal tax rates unchanged (see Appendix D for formal definitions). Imposing weak separability between income and consumption choices lets us simplify this.[18] The causal effect of $z$ on $x$ then 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:

[17] See Figure E.1 for a screenshot of the online survey page.

[18] 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.

*(a)* Aggregate Dirty Good




*(b)* Gasoline

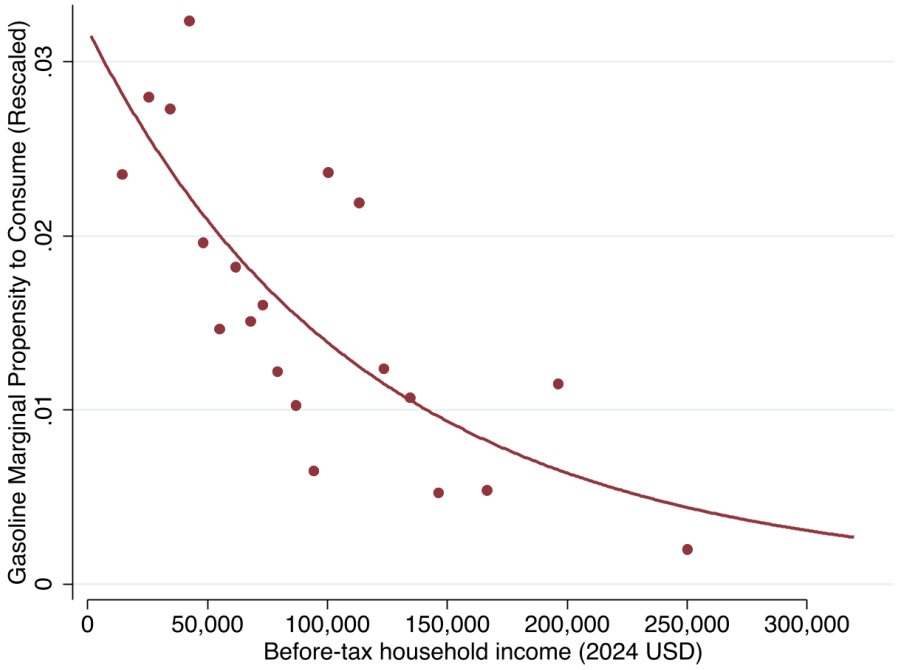



*Figure 2.* Causal income effect $x'_{\text{inc}}$ for carbon-intensive goods across the income distribution

**Note:** These binscatter plots show the causal income effect $x'_{\text{inc}}$ for carbon-intensive goods (corresponding to a rescaled marginal propensity to consume, see equation 55) across 20 equally-sized bins of household taxable income, together with our preferred Poisson-quadratic conditional-mean fit, using data from a custom representative survey of the US population ($N = 1,551$). Consumption responses are converted to 2024 dollars. The "dirty good" expenditure category includes gasoline, electricity, natural gas, and heating fuels.

$$\eta_{x|z}(z) = \frac{1}{1 - T_z'(z)} \frac{\partial x\left(\hat{w}\left(z\right); z\right)}{\partial z}. \tag{53}$$

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

$$\frac{dx\left(\hat{w}\left(z\right); z\right)}{dI} = \left(\frac{1}{1 - T_z'(z)} + \eta_z(z)\right) \frac{\partial x\left(\hat{w}\left(z\right); z\right)}{\partial z}. \tag{54}$$

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

$$x'_{\text{inc}}(z) \equiv \frac{\partial x\left(\hat{w}\left(z\right); z\right)}{\partial z} = \left(1 - T_z'(z)\right) \frac{dx\left(\hat{w}\left(z\right); z\right)}{dI}. \tag{55}$$

Figure 2 shows the causal income effect, $x'_{\text{inc}}$, across the income distribution obtained from our survey. It is calculated by equation 55, where $dx$ is the dollar amount of the permanent \$1,000 after-tax income increase allocated to the dirty good, and $dI$ is \$1,000. This MPC is rescaled by the net-of-tax rate obtained from the CEX, and matched to respondents' income (see Figure A.4). Figure 2 also plots the corresponding Poisson conditional-mean fit, quadratic in taxable income (see Appendix Table A.4). Panel (a) shows the aggregate results for all of the carbon-intensive goods, while panel (b) shows the results for gasoline. In both cases, the fitted response is positive and declines with income for both the aggregate dirty good and gasoline.

This approach for estimating $x'_{\text{inc}}$ relies on the unadjusted consumption response to the small income windfall, where respondents were allowed to save a portion of the small income increase. Since our model is static, it abstracts away from savings. As a result, the residual $x'_{\text{het}}$ may reflect in part heterogeneity in savings as well as heterogeneity in preferences for carbon-intensive goods. In Appendix section A.8, we show that our results are robust to an alternative "savings-adjusted" approach, in which we adjust both our measure for the cross-sectional slope $\hat{x}'(z)$ and the causal effect of income on dirty good consumption $x'_{\text{inc}}$ correspondingly, by rescaling the *share* of dirty good expenditures (as a percentage of total expenditures) with the level of after-tax income. Figures A.5 and A.6 show that these two adjustments result in almost identical results, leaving our main conclusions unchanged. If anything, the "savings-adjusted" optimal tax schedules are slightly closer to the Pigouvian benchmark (see Table A.7).

#### 4.1.3 Combining Cross-Sectional and Causal Variation to Estimate Taste Heterogeneity

Our next step is to evaluate the cross-sectional derivative and the causal income response on a common log-spaced income grid for numerical simulations, and calculate $x'_{\text{het}}(z)$ as the difference between $\hat{x}'(z)$ and $x'_{\text{inc}}(z)$. Starting from the binned CEX percentile distribution for income $z$ and dirty-good consumption $x$ (as plotted in Figure 1), we construct both a "continuous" distribution for taxable income $H(z)$ (smoothing the bin

weights at each averaged income level) and the corresponding cross-sectional consumption profile (in levels) by fitting a smoothing spline to $\log x$ as a function of $\log z$. We project the estimates from our regressions onto the same income grid to evaluate both (i) the CEX-based $\hat{x}'(z)$ from the fitted $x(z)$ schedule as plotted in Figure 1, with coefficient estimates shown in Table A.3, and (ii) the survey-based $x'_{\text{inc}}(z)$ from the fitted schedule plotted in Figure 2, with coefficient estimates (from the Poisson regression) shown in Table A.4.[19]

The results are depicted in Figure 3. For brevity, we restrict the remainder of our empirical analysis to the aggregate dirty good. However, we present results for each component in Appendix A, which are very similar in their implications.

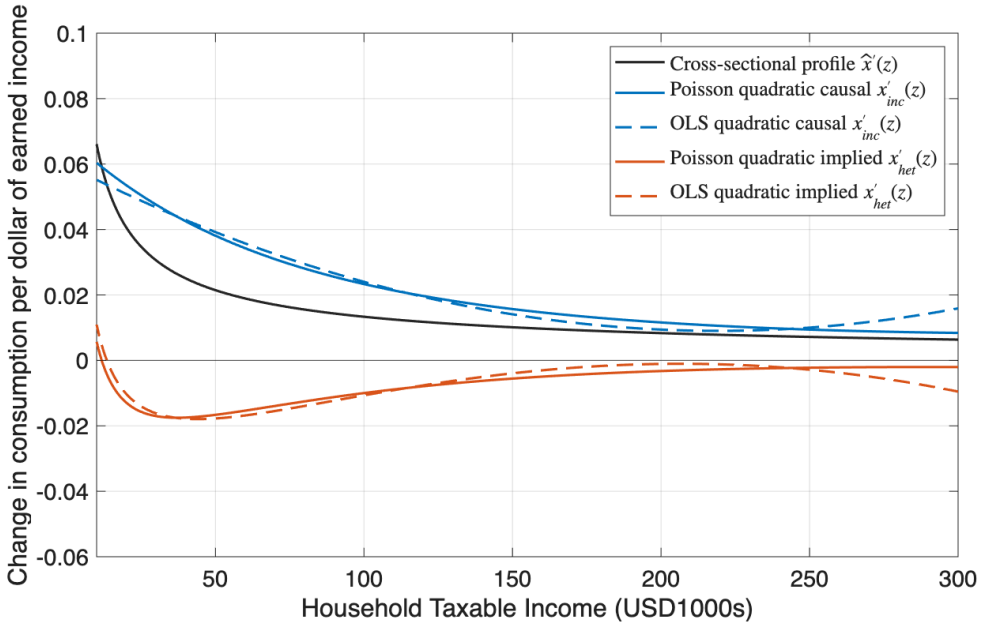



*Figure 3.* Decomposition of the cross-sectional profile $\hat{x}'(z)$ — Aggregate Dirty Good

**Note:** This figure shows how we decompose the cross-sectional relationship between income and carbon-intensive good consumption. The black line is a smoothed estimate of the cross-sectional slope, $\hat{x}'(z)$, at each level of income. The blue lines show estimates of the causal effect of income on expenditure on carbon-intensive goods. The orange lines subtract the income effect from the cross-section to obtain the part of the relationship between $x$ and $z$ that comes from preference heterogeneity. Solid blue and orange curves use the preferred Poisson specification, while the dashed curves show the OLS fit.

The black curve in Figure 3 depicts the smoothed cross-sectional profile $\hat{x}'(z)$. It is positive and 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.

[19]We also show coefficient estimates for the linear regression (OLS) with a quadratic in income. The Poisson regression is preferred because it is less sensitive to overfitting in the upper tail (see Figure 3) beyond the top income support of our custom survey. For the tax simulations, we further impose a smooth transition of the $x'_{\text{inc}}(z)$ schedule toward $\hat{x}'(z)$ between \$250,000 and \$450,000, setting the two equal thereafter. This is an upper-tail regularization assumption which sets $x'_{\text{het}}(z) = 0$ beyond \$450,000.

The blue curves show our fitted estimate of $x'_{\text{inc}}(z)$, which is decreasing throughout the income distribution. The orange curves show the implied (residual) $x'_{\text{het}}(z)$ which determines the sign of $\eta^{\text{Taste}}_{x,z}(z)$, a central input into the shape of our optimal and Pareto-efficient nonlinear tax schedules. Our estimates of $\hat{x}'(z)$ and $x'_{\text{inc}}(z)$ are close overall but the cross-sectional slope lies below the causal response across the income schedule, and more so at the lower end and middle of the income distribution i.e. for incomes up to \$150,000. This leads to estimates of $x'_{\text{het}}(z)$ that are negative, but converge towards zero as we move up the income distribution. This implies that differences in taste are mostly present at the bottom of the income distribution, and contribute relatively 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 an individual-level response (i.e., a change) from our custom hypothetical survey, while the second is computed as a slope using expenditures (i.e., levels) *across* households from the large CEX surveys.

For direct application in our quantitative analysis, Figure 4 translates this $x'_{\text{het}}(z)$ into the taste elasticity $\eta^{\text{Taste}}_{x,z}(z)$ that appears in equation 31, for our preferred Poisson specification. 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 1996–2021. We estimate the uncompensated elasticity to be $0.59$ (SE: $0.30$) in our preferred specification with month $\times$ year fixed effects and household-level controls (see Appendix A.6 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 ETI $\varepsilon_z(z)$ 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.

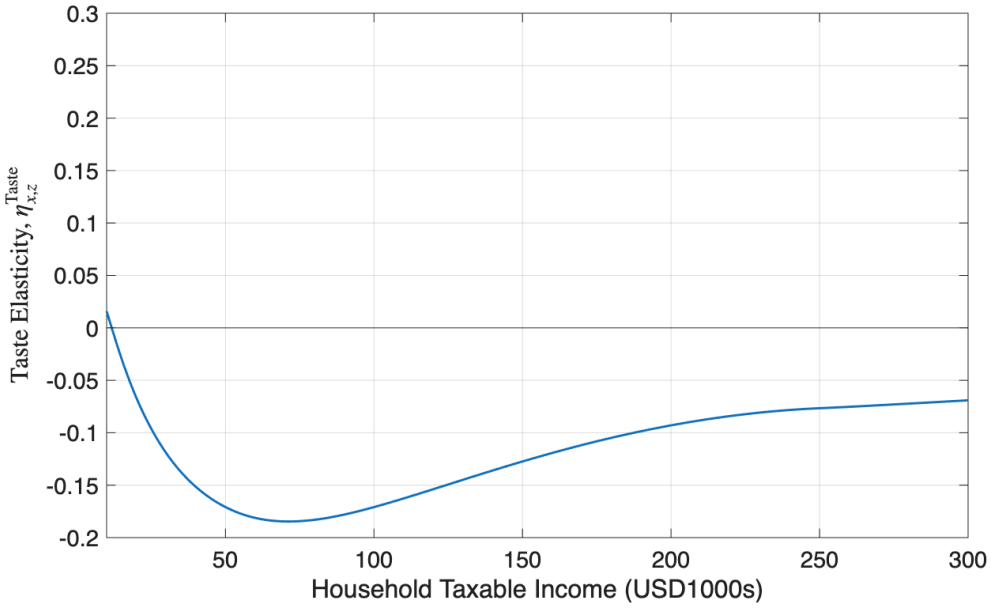



*Figure 4.* Elasticity of consumption of $x$ with respect to income via taste heterogeneity only

**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 orange line in Figure 3 (the Poisson specification), 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). The aggregate dirty good generates an estimated average of 2.462 of $CO_2$ per dollar of final expenditure.[20] This results in a marginal damage of 49 cents per dollar of final expenditure, which implies a baseline Pigouvian tax of 49% in our model.[21] We note that our Pareto-efficient tax schedules are calibrated relative to this marginal damage estimate. A reader who prefers an alternative estimate of the SCC can approximately shift the levels of all of our optimal tax schedules (e.g., Figure 6) to match their higher or lower estimate of the marginal damage from carbon-intensive consumption.

Marginal income tax rate schedule. We estimate the relationship between after-tax and before-tax household income, in a similar way to the cross-sectional profile of carbon-intensive goods, by fitting the 2018 to 2024 CEX data[22] (converted to 2024 USD using the average annual Consumer Price Index (CPI)) with a Poisson regression quadratic in the log of taxable income, with year fixed effects and CEX household weights.[23] We compute the

[20]This is 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).

[21]This magnitude is in line with the gasoline tax in France (TICPE) in 2024 (DILA, 2024), while in our data (see Appendix Section A.6 for data sources) the average gasoline tax across US states was 18.5% in 2021.

[22]We exclude observations with missing after-tax income data, which includes the 2024 Q2 to Q4 waves.

[23]The sample is restricted to households whose reference person is aged 25–54, with before-tax income of

marginal tax schedule as one minus the analytical derivative of the fitted relationship. The blue curve in Figure A.4 shows the resulting marginal income tax rates across the income distribution used in our simulations.

## 4.3 Empirically Optimal Tax Rates

Figure 5 plots the optimal marginal tax rates on the carbon-intensive good implied by our Pareto efficiency conditions for the unidimensional case. Panels (a) and (c) present the nonlinear case (equation 29). Panels (b) and (d) show the case where the tax is restricted to be linear (equation 33). Panels (a) and (b) assume an ETI $\varepsilon_z = 0.33$, which is most consistent with Chetty's (2012) meta-analysis. Panels (c) and (d) assume $\varepsilon_z = 0.7$, a high ETI scenario. 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 Nonlinear Taxation of Carbon — Unidimensional Case

We focus first on the nonlinear case. Under the benchmark calibration with $\varepsilon_z = 0.33$ and $\varepsilon_{x|z} = 0.5$, the Pareto-efficient marginal tax rate lies below marginal damage (the Pigouvian level) across the income distribution. It is 46.6 cents around \$40,000, falls to 44.7 cents around \$100,000, and then rises to 46.7 cents around \$250,000, compared with a marginal damage level of 49 cents per dollar of expenditure.

The downward adjustment follows from the negative preference-heterogeneity component in Figure 3. Households at lower incomes have relatively higher taste for carbon-intensive goods. A carbon tax slightly below the Pigouvian level can thus achieve redistribution at a lower distortionary cost than through the income tax. The departure from the Pigouvian level becomes larger when dirty-good consumption is less responsive (e.g. with $\varepsilon_{x|z} = 0.5$) or when taxable income is more elastic (e.g. with $\varepsilon_z = 0.7$) because using the carbon tax for redistribution is then less costly relative to using the income tax.

### 4.3.2 Optimal Linear Taxation of Carbon — Unidimensional Case

Panels (b) and (d) show the results when we restrict to linear taxation. In our benchmark scenario with $\varepsilon_z = 0.33$ and $\varepsilon_{x|z} = 0.5$, the Pareto-efficient rate is 46 cents, equal to 93.4% of marginal damage. The distributional adjustment lowers the optimal linear rate by only 3.3 cents per dollar of expenditure. In this sense, distributional concerns have a modest impact on the optimal carbon tax if it is required to be linear. As before, the adjustment is larger if taxable income is more elastic or dirty-good demand is less elastic.

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 32). Departures

at least \$5,000, positive total expenditures and after-tax income.

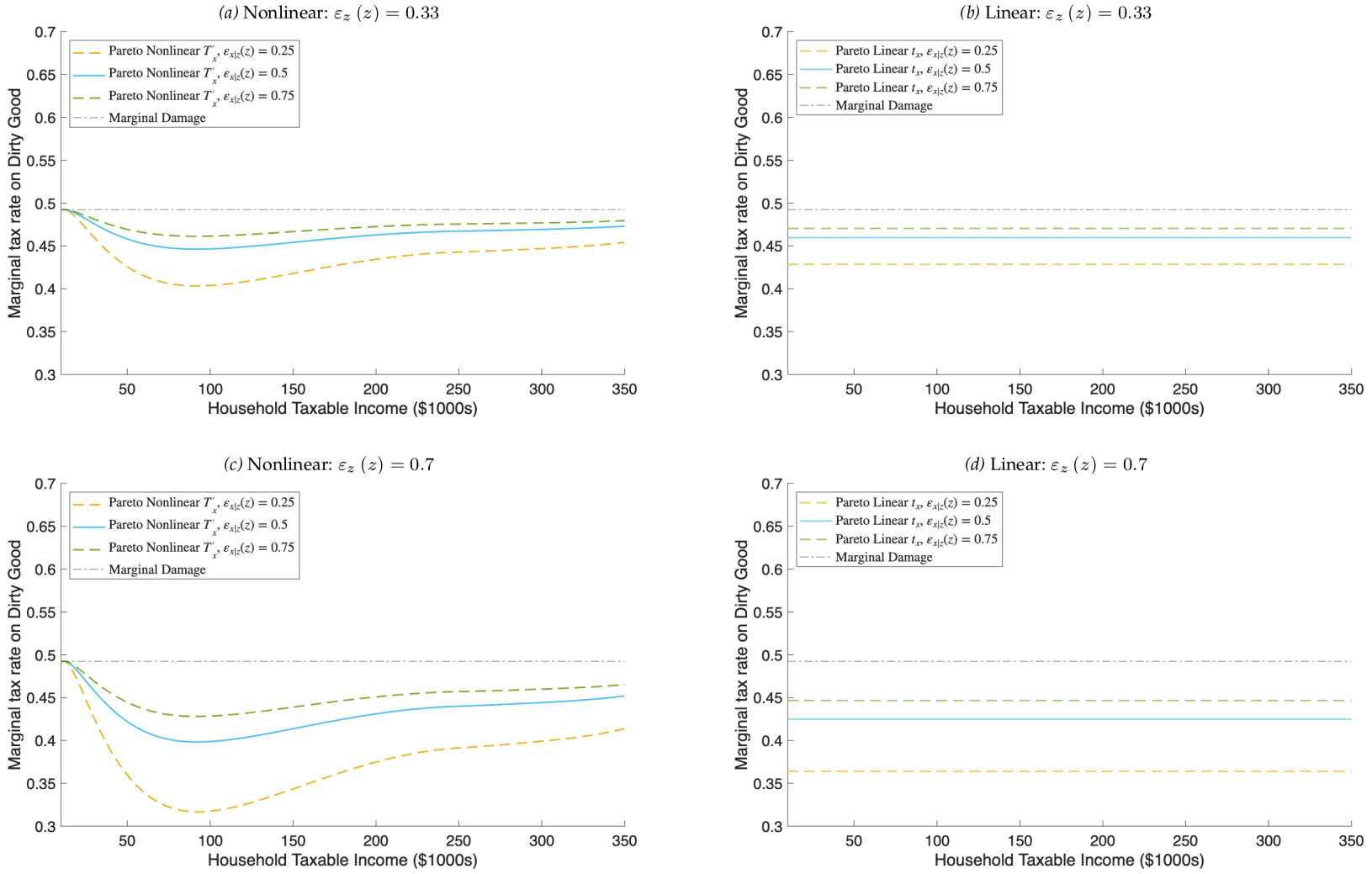



*Figure 5.* Dirty-Good Tax Schedule: Marginal Rates Implied by Pareto-Efficient Formulas

**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.

from the Pigouvian level thus depend on a weighted average of the preference heterogeneity term from equation 31. As Figure 4 shows, $\eta_{x,z}^{\text{Taste}}(z)$ is negative across the income distribution and rises toward zero at higher incomes. Integrating over the distribution, the Pareto-efficient linear tax is therefore below but near the Pigouvian level.

#### 4.3.3 Optimal Linear Taxation of Carbon — Multidimensional Case

To quantify the optimal carbon tax with multidimensional heterogeneity, we need the conditional variance of the income effect. To estimate this, we exploit our individual-level survey. Specifically, we use the squared residuals from the estimation of the conditional mean $x'_{\text{inc}}(z)$ with our Poisson-quadratic specification.[24] This lets us estimate $\text{Var}(x'_{\text{inc}} \mid z)$ by fitting the squared residuals with a quadratic generalized linear regression in taxable income (see Appendix A.5). We project this fitted conditional variance schedule onto the log-spaced income grid we used for the unidimensional numerical simulations.

Table A.5 shows the regression results, while Figure A.3 plots the fitted schedule. The conditional variance of the causal income effect is largest for households at the bottom of the income distribution, and declines moderately over the income range shown, implying that multidimensional heterogeneity in the causal income response is most significant at lower incomes and smaller at higher income levels.[25]

Aside from $\text{Var}(x'_{\text{inc}} \mid z)$, we re-use most of the empirical quantities from the unidimensional case, with the difference that point estimates used for calibration of the unidimensional case are interpreted as conditional averages. For example, $\eta_{x,z}^{\text{Taste}}(z) = \mathbb{E}[\eta_{x,z}^{\text{Taste}}(z,\theta) \mid z]$ 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.

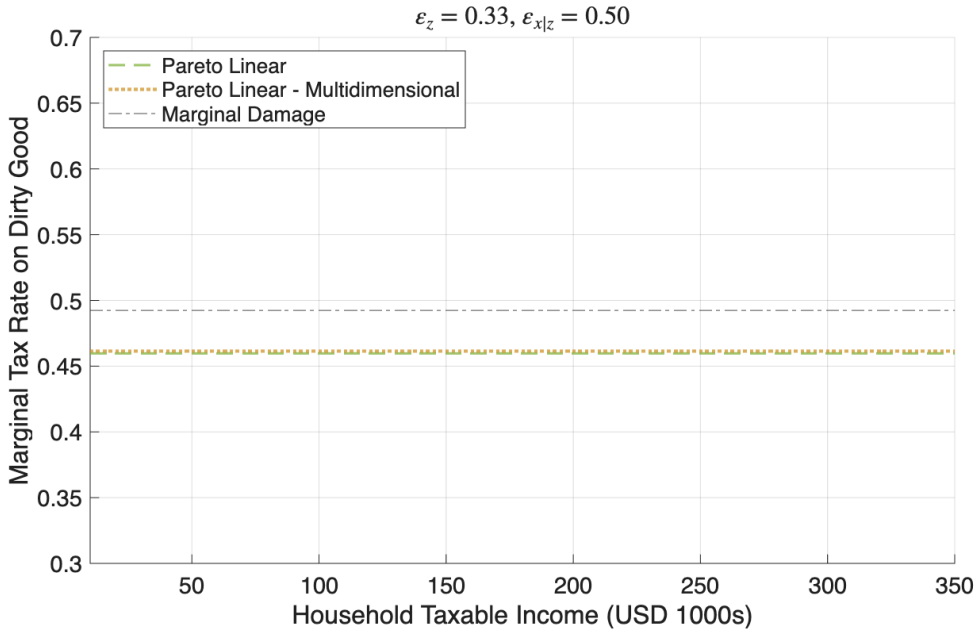
Figure 6 plots the optimal tax rate for the multidimensional case (orange dotted line). For comparison, we also include the linear tax rate implied by the Pareto-efficient linear tax formula in the unidimensional case (green dashed line). Panel (a) assumes an ETI $\varepsilon_z = 0.33$, and an elasticity 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 on our theoretical results, within-income variation of the causal effect of income leads to attenuation: The optimal tax is pushed closer to the Pigouvian benchmark.

The degree of attenuation is nonetheless very small in our baseline calibration. It is slightly larger in the lower panel, where the deviation from Pigouvian taxation was larger without multidimensional heterogeneity. This is because the amount of attenuation due to

[24] Said differently, we exploit the identity $e_i^2 = \left(x'_{\text{inc},i} - \widehat{\mathbb{E}}[x'_{\text{inc}} \mid z_i]\right)^2$.

[25] Computing the raw variance of the causal effect at deciles of the income distribution yields similar values to the fitted curve. We note that both approaches may yield upwardly biased estimates of $\text{Var}(x'_{\text{inc}} \mid z)$ if there is significant measurement error in $x'_{\text{inc}}$, which is plausible.

*(a)* Baseline calibration




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

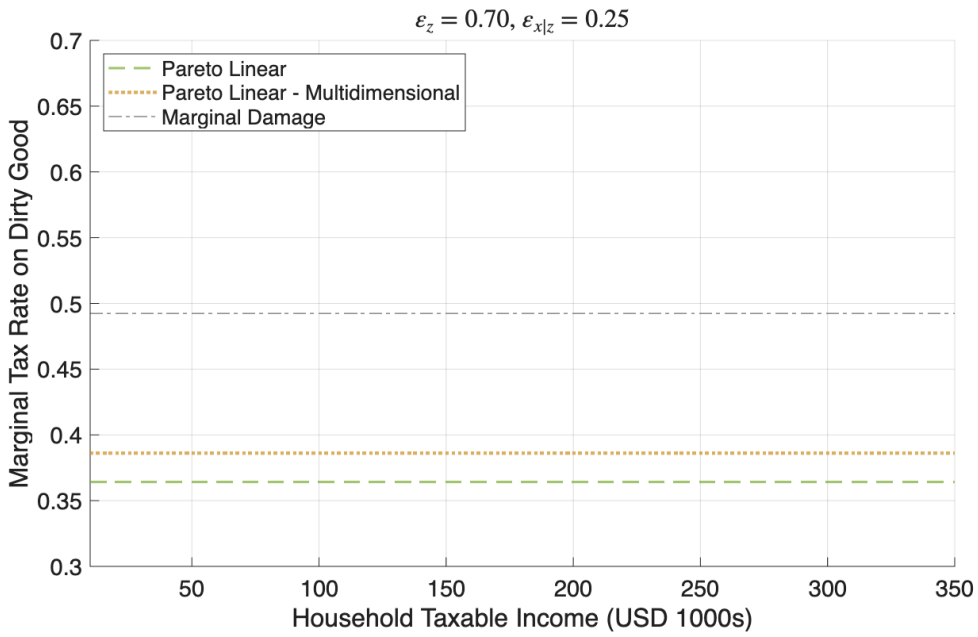



*Figure 6.* Dirty-Good Marginal Linear Tax Rate: Unidimensional vs. Multidimensional Heterogeneity

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

$\mathrm{Var}(x'_{\mathrm{inc}} \mid z)$ is proportional to the ratio of elasticities $\varepsilon_z/\varepsilon_{x|z}$ (see equation 49). Recall that this same ratio scales the intensity of the preference heterogeneity adjustment in the first place (see equation 32), which implies that the degree of attenuation from within-income heterogeneity in $\partial x/\partial z$ is greatest precisely when the numerator of equation 49 is largest. As a result, there is only minimal attenuation in Figure 6(a) where the optimal linear rate in the unidimensional case was already close to the Pigouvian level. There is more attenuation in panel (b), where the higher ratio of elasticities led to a larger deviation.

#### 4.3.4 Back-of-the-Envelope Welfare Calculation

A back-of-the-envelope calculation suggests that the additional surplus from adjusting the tax on the dirty good from the Pigouvian level to the Pareto-efficient level is small. In our benchmark calibration, moving to the Pareto-efficient rate adds \$1.05 per household per year for the linear tax, \$1.22 for the nonlinear schedule, and \$0.99 for the linear tax when within-income heterogeneity is accounted for. The linear-tax gain is 0.0013% of mean consumption expenditure (see Appendix C and Table C.1 for details).

For comparison, moving from a zero tax rate to the Pigouvian level yields a surplus of \$267.36 per household per year. Taxing at the Pigouvian rate therefore captures 99.61% of the calculated surplus from moving from a zero tax to the Pareto-efficient tax rate in the linear case (for our baseline elasticity scenario), and at least 89.2% across the elasticity combinations considered (across linear and nonlinear tax schedules). While we focus on different mechanisms, this is similar to Douenne et al.'s (2025) result that there is only a 0.008% consumption-equivalent welfare gain from moving from Pigouvian taxation to their optimum in their dynamic model with incomplete markets.

## 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.

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. What matters is whether this regressivity reflects income effects or preference heterogeneity. Differences in taste for carbon-intensive goods at different income levels can make the carbon tax a lower-cost redistributive instrument relative to the income tax. 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 tax on carbon-intensive goods depend on whether nonlinear taxation is feasible. If only linear carbon taxation is possible, then the optimal tax is slightly below but remains close to the classic Pigouvian prescription of setting the tax equal to the marginal damage from emitted carbon. The Pareto-efficient rate is 46 cents in our baseline scenario, compared with marginal damage of 49 cents per dollar of expenditure. Allowing for heterogeneity within income levels leads to attenuation and moves this rate closer toward the Pigouvian benchmark.

The conclusion is more nuanced if nonlinear carbon taxation is possible. Our benchmark scenario implies that optimal marginal tax rates are below the Pigouvian level and more so for lower- and middle-income households, but rise back toward 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.

Our conceptual and empirical frameworks could be applied to many settings that involve both distributional concerns and externalities. This could include other environmental externalities, public health problems, or even criminal behavior. Our results 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. However, this result may differ for other externalities.

# 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. (2026). Public goods and optimal taxation under multidimensional heterogeneity. Working paper.

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. (2010). Environmental tax design with endogenous earning abilities (with applications to france). *Journal of Environmental Economics and Management*, 59(1):82–93.

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.

Djalalova, F. and Rothschild, C. (2025). Putting the "re" in redistributive taxation. Working paper.

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., Dyrda, S., Hummel, A. J., and Pedroni, M. (2025). Optimal climate policy with incomplete markets. CEPR Discussion Paper 20820, CEPR.

Douenne, T., Hummel, A. J., and Pedroni, M. (2026). Optimal fiscal policy in a climate-economy model with heterogeneous households. *The Economic Journal*, 136(677):1761–1801.

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

Farhi, E. and Gabaix, X. (2020). Optimal taxation with behavioral agents. *American Economic Review*, 110(1):298–336.

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. *Review of Economic Dynamics*, 30:30–46.

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.

IGM Forum (2011). Carbon tax. Economic Experts Panel (US), Initiative on Global Markets, University of Chicago Booth School of Business, December 20.

IGM Forum (2012). Carbon taxes ii. Economic Experts Panel (US), Initiative on Global Markets, University of Chicago Booth School of Business, December 4.

IGM Forum (2018). Climate change policies. Economic Experts Panel (US), Initiative on Global Markets, University of Chicago Booth School of Business, November 13.

IGM Forum (2025). Low-carbon energy R&D. Economic Experts Panel (US), Kent A. Clark Center for Global Markets, University of Chicago Booth School of Business, October 1.

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. (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):1635–1682.

Lyubich, E. (2025). The role of people versus places in individual carbon emissions. *American Economic Review*, 115(5):1439–1484.

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. (1999). 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.

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*.

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. *The Review of Economic Studies*, 83(3):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.

Taubinsky, D. and Rees-Jones, A. (2018). Attention variation and welfare: Theory and evidence from a tax salience experiment. *The Review of Economic Studies*, 85(4):2462–2496.

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.

# Appendix - Table of Contents


**A Empirical Appendix** **47**

A.1 Average Expenditure Shares and $CO_2$ Emissions across Categories . . . . . 47

A.2 Estimating a Smoothed Cross-Sectional Profile $\hat{x}'(z)$ . . . . . 48

A.3 Cross-Sectional Profile: Carbon-Intensive Consumption in *Shares* . . . . . 49

A.4 Share of Carbon-Intensive Goods in Status Quo Expenditures . . . . . 50

A.5 Estimating Income-Conditional Moments of the Causal Consumption Response $\mathbb{E}[x'_{\text{inc}}|z]$ and $\text{Var}[x'_{\text{inc}}|z]$ . . . . . 51

A.6 Estimating the Uncompensated Elasticity of Gasoline Consumption . . . . . 52

A.7 Status Quo Income Marginal Tax Rates . . . . . 54

A.8 Adjusting for Savings . . . . . 55

A.9 Pareto-Efficient Marginal Tax Schedules — Gasoline Only . . . . . 58

A.10 Pareto-Efficient Marginal Tax Schedules — Electricity, Natural Gas, and Heating Fuels Only . . . . . 59

A.11 Online Survey Demographics Breakdown & Sample Definition . . . . . 60

**B Optimal Levels of the Income and Externality Taxes** **61**

**C Surplus Forgone at the Pigouvian Tax** **63**

**D Proofs of Theoretical Results** **66**

**E Survey Questionnaire Appendix** **79**

E.1 Survey Link and Metadata . . . . . 79

E.2 Prolific-Provided Respondent Demographics . . . . . 79

E.3 Survey Questions . . . . . 79

# 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 | 0.175 |
| Housing (excluding utilities) | 0.279 |
| Electricity, natural gas, and home heating fuels | 0.039 |
| Water, sewage, and other public services | 0.012 |
| Healthcare | 0.061 |
| Gasoline and other motor fuels | 0.043 |
| Flights / Airfare expenditures | 0.005 |
| Other Expenditures | 0.386 |
| Observations | 59,238 |

**Note:** This table shows shares allocated to each consumption category among consumers in the Consumer Expenditure Survey before-tax/expenditure working sample: 2018–2024, including all four quarters of 2024, reference-person age 25–54, and annual before-tax income of at least $600 in 2024 dollars.

*Table A.2.* Average $CO_2$ Emissions per Dollar of Final Energy Expenditure in the United States

| | **Unit** | **Price per Unit (2023 $)** | **$CO_2$ Emissions per Unit (kg)** | **$CO_2$ Emissions (kg per $)** |
|---|---|---|---|---|
| Gasoline | Gallon (3.785 litres) | 3.635 | 8.778 | 2.415 |
| Electricity | kWh delivered | 0.1600 | 0.363 | 2.270 |
| Natural gas | 1,000 cubic feet (Mcf) | 15.40 | 54.440 | 3.535 |
| Air travel | Passenger-mile | 0.1566 | 0.182 | 1.161 |

**Note:** Prices are nominal 2023 U.S. average final consumer prices from the U.S. Energy Information Administration (EIA). Gasoline uses the EIA U.S. annual retail price for all grades. Electricity and natural gas use EIA residential prices, consistent with household final expenditure categories. Emissions are direct $CO_2$ emissions only and exclude upstream, refining, transmission, and other life-cycle emissions. Electricity emissions use EPA eGRID2023's U.S. average $CO_2$ output rate, adjusted for grid gross losses. Gasoline and natural gas emissions use default combustion factors. Air travel emissions are calculated using jet fuel emission factors (EIA/EPA) and passenger-mile fuel economy from the U.S. Transportation Energy Data Book. The price per passenger-mile uses FAA passenger yield data.

## A.2 Estimating a Smoothed Cross-Sectional Profile $\hat{x}'(z)$

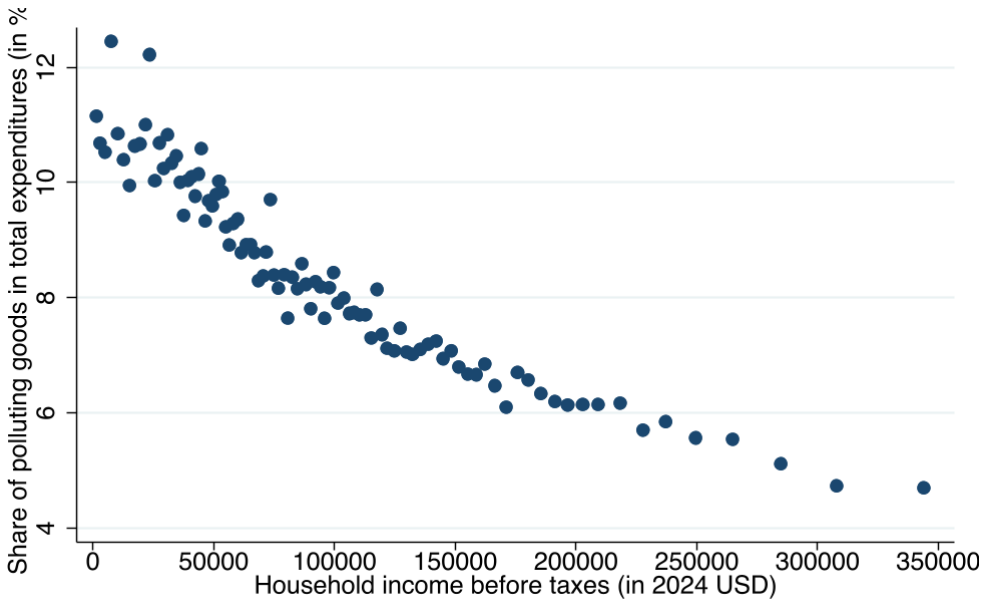
*Table A.3.* Coefficient Estimates – Conditional Mean Carbon-Intensive Consumption

| | Dirty goods | Elec. and heating fuels | Gasoline |
|---|---|---|---|
| Centered log income, $\widetilde{z}_i$ | 0.228*** | 0.217*** | 0.240*** |
| | (0.005) | (0.006) | (0.007) |
| Centered log income squared, $\widetilde{z}_i^2$ | 0.010** | 0.052*** | -0.029*** |
| | (0.004) | (0.005) | (0.006) |
| Constant | 8.653*** | 7.799*** | 8.101*** |
| | (0.012) | (0.013) | (0.016) |
| Year fixed effects | Yes | Yes | Yes |
| Households | 25,199 | 25,199 | 25,199 |
| Observations | 57,345 | 57,345 | 57,345 |

**Note:** This table reports coefficient estimates from separate Poisson pseudo-maximum-likelihood regressions of annualized consumption (in levels) of each good on centered log before-tax household income. Let $x_i$ denote household $i$'s annual consumption expenditure, $z_i$ its annual before-tax income, and $t$ its survey year. The conditional mean is $\mathrm{E}[x_i \mid z_i, t] = \exp\big(\alpha + \beta_1 \widetilde{z}_i + \beta_2 \widetilde{z}_i^2 + \delta_t\big)$, where $\widetilde{z}_i = \ln\Big(\frac{z_i}{90{,}000}\Big)$ and $\delta_t$ denotes year fixed effects, with 2024 as the omitted year. Consumption is measured in 2024 dollars. The sample consists of Consumer Expenditure Survey observations from 2018–2024 for households whose reference person is aged 25–54 and whose before-tax income is at least $5,000 in 2024 dollars (to avoid overfitting observations with both very low income, likely transitory, and relatively large consumption levels). The regressions use CEX household weights. Estimates are then used to construct the consumption-level fits as plotted in Figure 1 and analytical derivatives $\hat{x}'(z)$ passed to numerical simulations for the optimal tax schedules. Standard errors are clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

## A.3 Cross-Sectional Profile: Carbon-Intensive Consumption in *Shares*

*(a)* Aggregate Dirty Good




*(b)* Gasoline

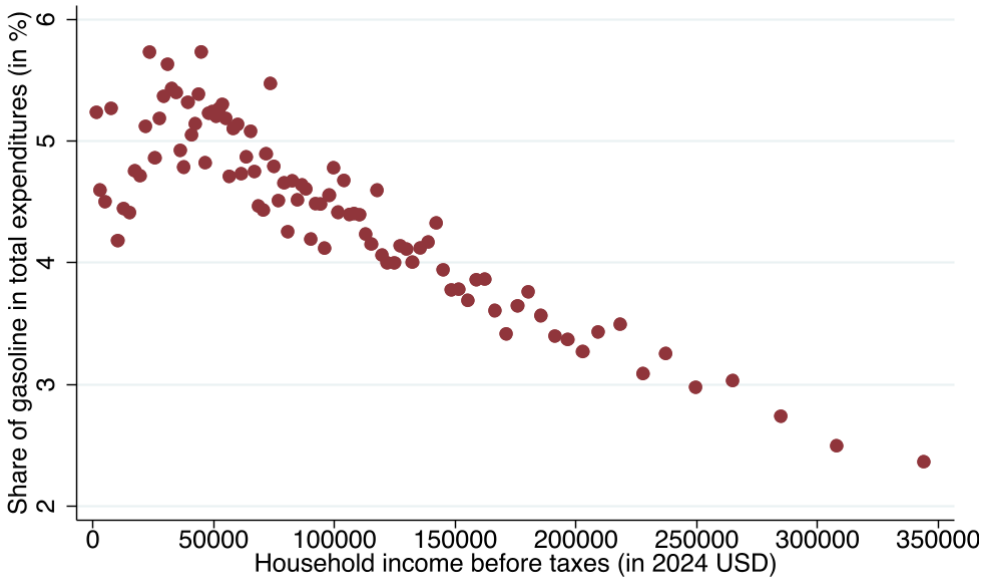
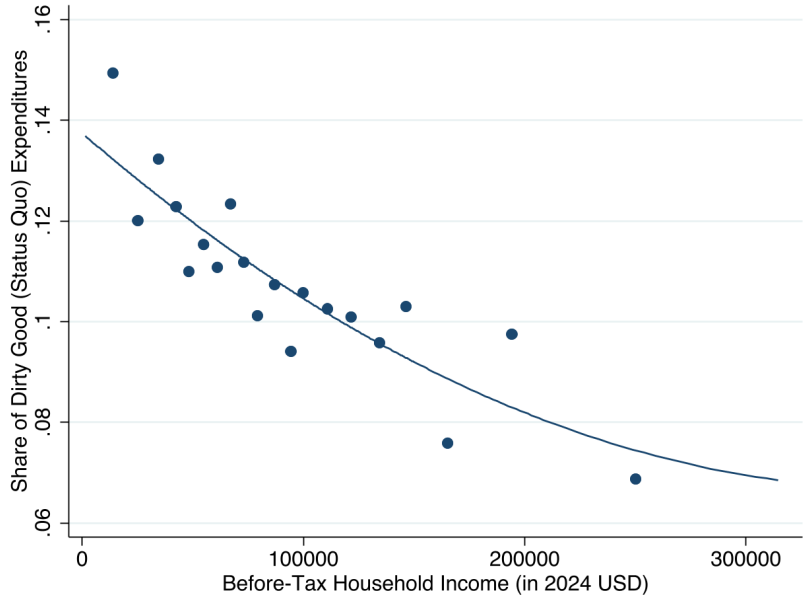



*Figure A.1.* Expenditure shares of carbon-intensive goods across the income distribution

**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 2018–2024 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. The figure is truncated at $350k for readability. This excludes the top 3% of earners.

## A.4 Share of Carbon-Intensive Goods in Status Quo Expenditures

*(a)* Aggregate Dirty Good




*(b)* Gasoline

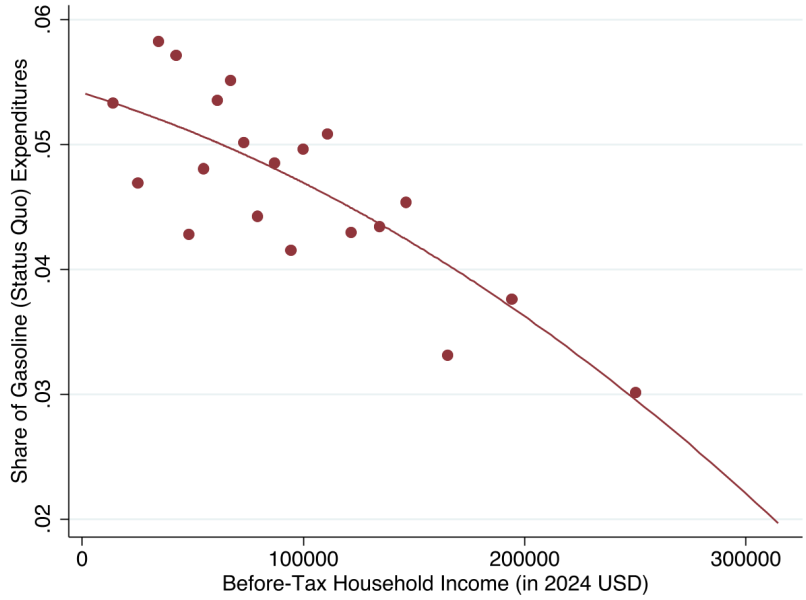



*Figure A.2.* Share of carbon-intensive goods in total (monthly) expenditures in custom survey

**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 ($N = 1,442$). The "dirty good" expenditure category includes gasoline, electricity, natural gas, and heating fuels.

## A.5 Estimating Income-Conditional Moments of the Causal Consumption Response $\mathbb{E}[x'_{\text{inc}}|z]$ and $\text{Var}[x'_{\text{inc}}|z]$

*Table A.4.* Coefficient Estimates – Conditional Mean of the Causal Income Effect $x'_{\text{inc}}(z)$

| | Dirty goods | | Elec. and heating fuels | | Gasoline | |
|---|---|---|---|---|---|---|
| | OLS | Poisson | OLS | Poisson | OLS | Poisson |
| Income (10,000 USD) | -0.00462*** | -0.126*** | -0.00264*** | -0.173*** | -0.00198*** | -0.0855** |
| | (0.000930) | (0.0319) | (0.000572) | (0.0388) | (0.000590) | (0.0386) |
| Income squared | 0.000105*** | 0.00188 | 0.0000672*** | 0.00369** | 0.0000382** | 0.000251 |
| | (0.0000303) | (0.00132) | (0.0000185) | (0.00147) | (0.0000189) | (0.00163) |
| Constant | 0.0597*** | -2.682*** | 0.0296*** | -3.293*** | 0.0301*** | -3.447*** |
| | (0.00589) | (0.148) | (0.00366) | (0.187) | (0.00377) | (0.185) |
| Mean squared error | 0.00408 | 0.00407 | 0.00128 | 0.00127 | 0.00162 | 0.00162 |
| Observations | 1,551 | 1,551 | 1,551 | 1,551 | 1,551 | 1,551 |

**Note:** This table reports coefficient estimates from the regression of $x'_{\text{inc},i}$ on before-tax income $z$ and before-tax income squared $z^2$. Columns are estimated separately for each good using our custom survey sample. Income is measured in tens of thousands of 2024 dollars. *OLS* columns correspond to a linear regression with a quadratic in income; *Poisson* columns correspond to a log-link quadratic specification. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

*Table A.5.* Coefficient Estimates – Conditional Variance of the Causal Income Effect $\text{Var}(x'_{\text{inc}} \mid z)$

| | Dirty goods | Gasoline | Elec. and heating fuels |
|---|---|---|---|
| Income (million USD) | -7.933 | -7.193 | -12.71* |
| | (5.631) | (6.985) | (6.738) |
| Income squared | -4.073 | -9.150 | 7.603 |
| | (21.46) | (27.25) | (26.26) |
| Constant | -4.815*** | -5.751*** | -5.731*** |
| | (0.313) | (0.382) | (0.370) |
| Mean squared error | 0.00026 | 0.00007 | 0.00004 |
| Observations | 1,551 | 1,551 | 1,551 |

**Note:** This table reports Gamma GLM coefficient estimates with a log link for the squared residuals from the baseline Poisson-quadratic conditional-mean model of $x'_{\text{inc}}$, shown in Table A.4. The fitted values provide the paper's estimate of $\text{Var}(x'_{\text{inc}} \mid z)$ as a function of before-tax household income $z$. Income is measured in millions of 2024 dollars (for ease of exposition of the coefficient estimates' units in the table). Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

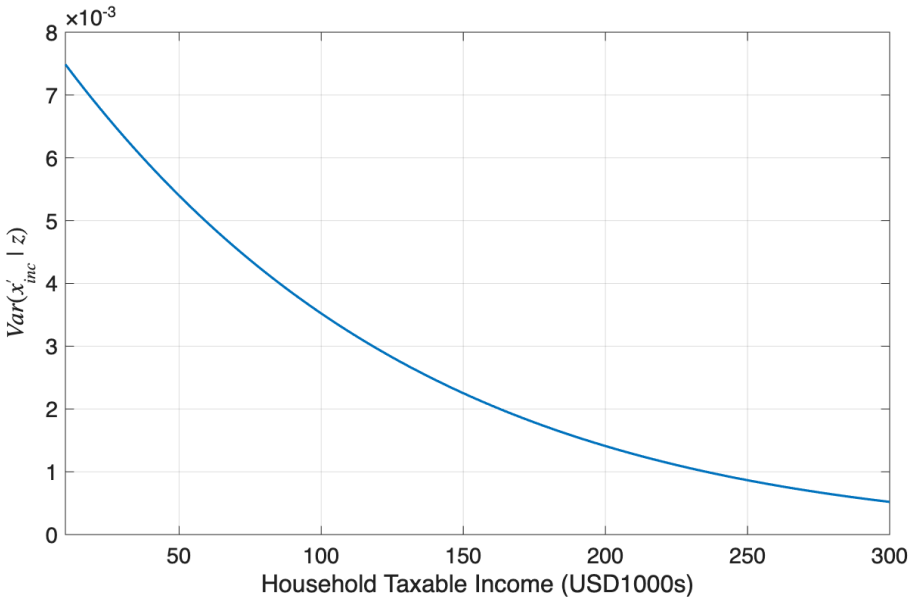



*Figure A.3.* Conditional variance of $x'_{\text{inc}}(z)$ across the income distribution

**Note:** This figure shows the fitted conditional variance of the causal effect of income on dirty-good consumption, $\text{Var}(x'_{\text{inc}} \mid z)$, estimated from our custom survey ($N = 1,551$). The fitted curve comes from a log-link Gamma GLM for squared residuals from the preferred Poisson quadratic mean specification and is evaluated on the same income grid used in the tax simulations.

## A.6 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:

$$\log(q_{ist}) = \alpha \log(p_{st}) + \beta \log(1 + \frac{\tau_{st}}{p_{st}}) + \Theta X_{ist} + \delta_s + \phi_t + e_{ist} \tag{56}$$

where $q_{ist}$ is the household-level consumption of gasoline (in quantities), $p_{st}$ is the tax-exclusive price of gasoline, $\tau_{st}$ is the gasoline excise tax (in dollars); $X_{ist}$ is a vector of household-level controls; and $\delta_s$ and $\phi_t$ 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.6.* Elasticity of gasoline consumption with respect to tax rate: 1996–2021

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $\log(1 + \text{tax ratio})$ | -0.379 | -0.382 | -0.585** | -0.589** |
| | (0.322) | (0.322) | (0.297) | (0.297) |
| $\log(\text{gas price})$ | -0.453*** | -0.454*** | -0.414*** | -0.415*** |
| | (0.145) | (0.145) | (0.136) | (0.136) |
| 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 | 0.090 | 0.201 | 0.202 |
| 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 (56). 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 1996–2021 on household gasoline consumption are from the interview module of the Consumer Expenditure Surveys 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.7 Status Quo Income Marginal Tax Rates

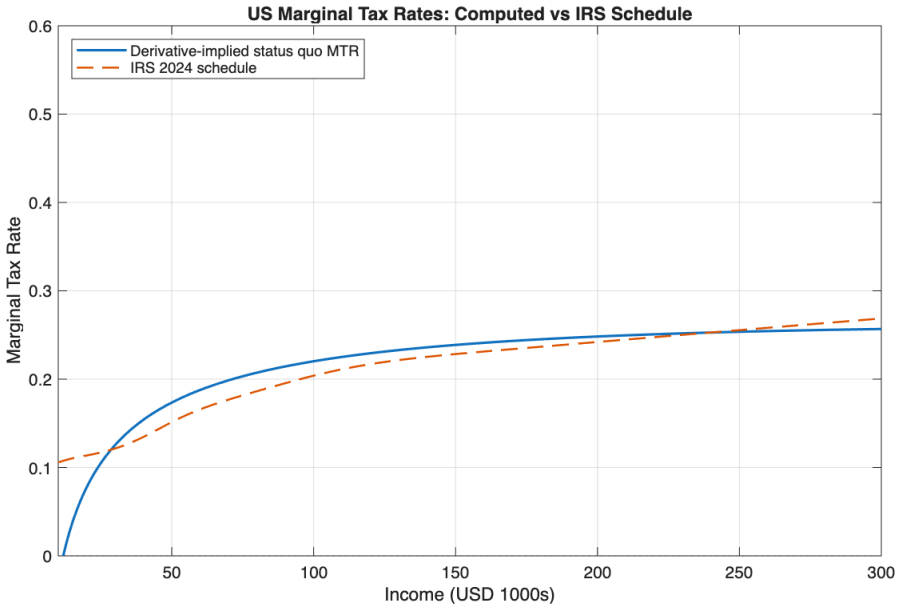



*Figure A.4.* Status quo income marginal tax rates across the income distribution

**Note:** This figure shows the status quo income marginal tax rates across the household income distribution. The blue curve is obtained from a Poisson regression of after-tax income on a quadratic in log (centered) before-tax income and year fixed effects, using CEX data for 2018 to 2024, converted to 2024 USD, and CEX household weights. The sample is restricted to households whose reference person is aged 25-54, with positive total expenditures and after-tax income, and non-missing tax information. The marginal tax rate is computed as 1 minus the derivative of the fitted relationship. 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.8 Adjusting for Savings

Our model is static, and thus abstracts away from savings. However, households do save a fraction of their disposable incomes, and may have positive (and potentially heterogeneous) marginal propensities to save. We therefore use a *savings-adjusted* approach. First, when constructing the cross-sectional profile $\hat{x}'(z)$ with the CEX data, we can use *shares* (as in Figure A.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. Formally, we can write $\tilde{x} = (x/E) \times I = s_x \times I$, where $x$ is observed consumption on the dirty good, $E$ total consumption, $s_x$ the share spent on dirty good consumption, and $I = z - T(z)$ disposable income. This adjustment allocates disposable income across goods in proportion to current expenditure shares. Using the CEX household data, we construct $\tilde{x}$ for each respondent and estimate its cross-sectional slope in the same way as for the unadjusted consumption schedule.

Second, we calculate the causal effect of income on this savings-adjusted level of consumption $\tilde{x}'_{\text{inc}}(z)$, which can be obtained using equation 55 applied to $\tilde{x}$ together with the product rule:

$$\begin{aligned} \tilde{x}'_{\text{inc}}(z) &= (1 - T'_z(z)) \frac{d\tilde{x}}{dI} \\ &= (1 - T'_z(z)) \big[ s_x + \frac{I}{E} (\frac{dx}{dI} - s_x \frac{dE}{dI}) \big], \end{aligned} \tag{57}$$

where the marginal propensities to consume for the dirty good composite $dx/dI$ and total expenditures $dE/dI$ are elicited directly in our custom survey. We match each respondent's taxable income to smoothed CEX expenditure shares and to the ratio of CEX disposable income and total expenditure. This uses the status quo variables at each income level while retaining variation in respondents' marginal consumption choices.[26]

[26]The savings adjustment can produce very slightly negative $\tilde{x}'_{\text{inc}}$ even when the reported marginal propensities to consume are nonnegative. As a result, for the preferred Poisson specification, we set negative adjusted responses to zero, while the OLS specification retains them (see Figure A.5 which shows how the level is virtually unaffected by this adjustment). Matching CEX quantities by income also abstracts away from within-income covariance between status quo expenditures and marginal consumption responses.

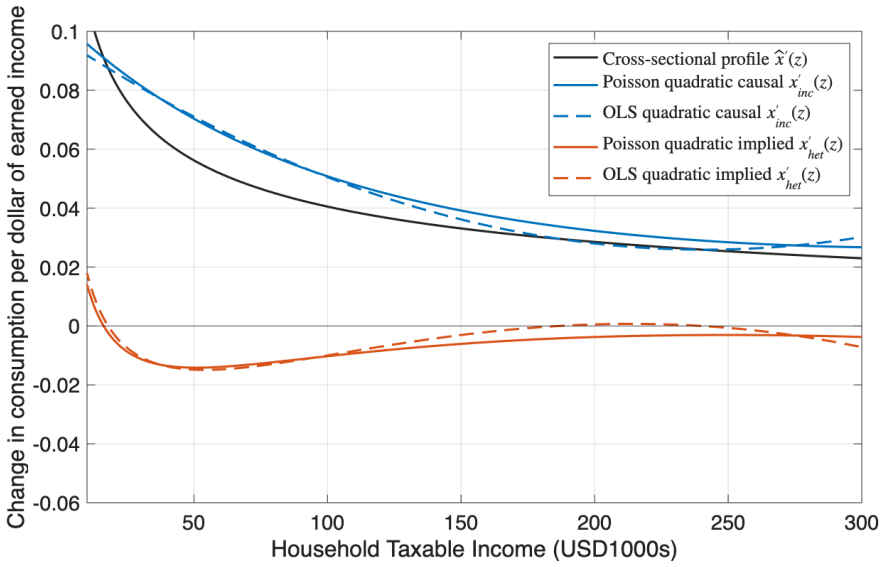



*Figure A.5.* Savings-Adjusted Decomposition of the Cross-Sectional Profile — Aggregate Dirty Good

**Note:** This figure shows how we decompose the cross-sectional relationship between income and *savings-adjusted* consumption, $\tilde{x} = (x/E) \times I$, for the aggregate dirty good (gasoline, electricity, natural gas, and heating fuels). The black line shows the cross-sectional slope estimated from the CEX data. The blue lines show the causal income response in equation 57, using survey marginal propensities to consume and CEX status quo expenditures and shares, matched on taxable income. The orange lines show the residual heterogeneity component.

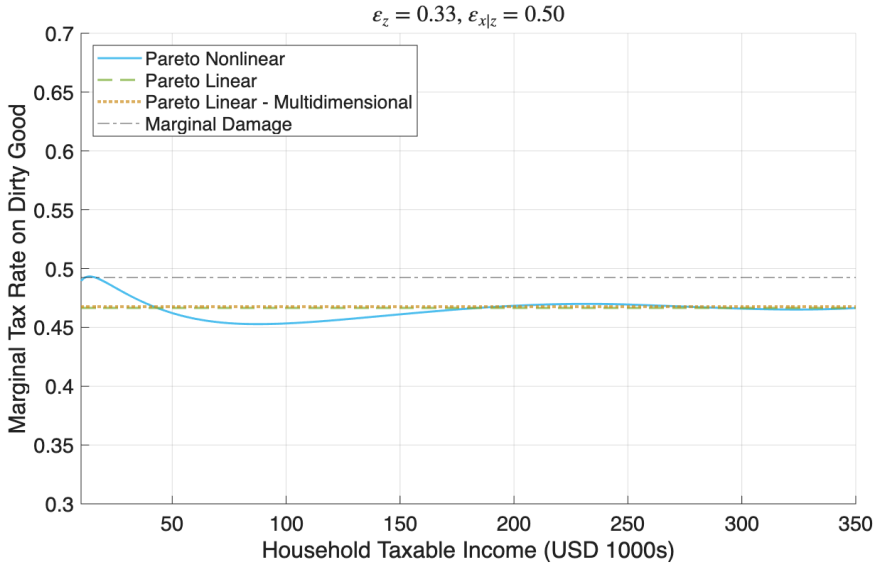



*Figure A.6.* Savings-Adjusted Dirty-Good Marginal Tax Rates: Unidimensional vs. Multidimensional Heterogeneity

**Note:** This figure plots the Pareto-efficient nonlinear and linear carbon taxes in the unidimensional case, and the optimal linear tax allowing for multidimensional heterogeneity, using the savings adjustments described in Appendix A.8 when the elasticity of taxable income is $\varepsilon_z = 0.33$, and the elasticity of consumption of the carbon-intensive good $\varepsilon_{x|z} = 0.5$.

*Table A.7.* Baseline and Savings-Adjusted Dirty-Good Tax Rates

| | | Nonlinear, unidimensional | | | Linear | |
|---|---|---|---|---|---|---|
| $(\varepsilon_z, \varepsilon_{x\|z})$ | Construction | \$50k | \$150k | \$300k | Unidim. | Multidim. |
| $(0.33, 0.25)$ | Baseline | 42.59 | 41.80 | 44.69 | 42.86 | 43.47 |
| | Savings-adjusted | 43.38 | 43.14 | 44.06 | 44.19 | 44.56 |
| | Difference (pp) | +0.80 | +1.34 | -0.63 | +1.32 | +1.09 |
| **(0.33, 0.50)** | **Baseline** | **45.81** | **45.41** | **46.93** | **45.97** | **46.14** |
| | **Savings-adjusted** | **46.22** | **46.11** | **46.60** | **46.66** | **46.76** |
| | **Difference (pp)** | **+0.40** | **+0.69** | **-0.33** | **+0.69** | **+0.62** |
| $(0.33, 0.75)$ | Baseline | 46.93 | 46.67 | 47.69 | 47.04 | 47.12 |
| | Savings-adjusted | 47.20 | 47.13 | 47.46 | 47.50 | 47.55 |
| | Difference (pp) | +0.27 | +0.47 | -0.23 | +0.46 | +0.43 |
| $(0.70, 0.25)$ | Baseline | 35.98 | 34.36 | 39.92 | 36.42 | 38.61 |
| | Savings-adjusted | 37.60 | 36.97 | 38.71 | 39.02 | 40.42 |
| | Difference (pp) | +1.62 | +2.62 | -1.21 | +2.60 | +1.80 |
| $(0.70, 0.50)$ | Baseline | 42.21 | 41.37 | 44.42 | 42.49 | 43.17 |
| | Savings-adjusted | 43.05 | 42.78 | 43.76 | 43.89 | 44.31 |
| | Difference (pp) | +0.84 | +1.41 | -0.67 | +1.40 | +1.14 |
| $(0.70, 0.75)$ | Baseline | 44.45 | 43.89 | 45.99 | 44.66 | 44.99 |
| | Savings-adjusted | 45.02 | 44.86 | 45.53 | 45.62 | 45.82 |
| | Difference (pp) | +0.57 | +0.97 | -0.46 | +0.96 | +0.83 |

**Note:** This table shows a comparison of optimal tax rates for the aggregate dirty good under the baseline and savings-adjusted approaches to constructing $x'_{\text{het}}$. Tax rates are in percent, while differences are computed as the savings-adjusted tax rates minus baseline tax rates, in percentage points. Nonlinear rates are evaluated at the three levels of household taxable incomes (2024 USD). The bold rows highlight the values for our preferred elasticity scenario, namely $\varepsilon_z = 0.33$ and $\varepsilon_{x|z} = 0.5$. Marginal damage, corresponding to the Pigouvian level, is assumed to be equal to 49%.

## A.9 Pareto-Efficient Marginal Tax Schedules — Gasoline Only

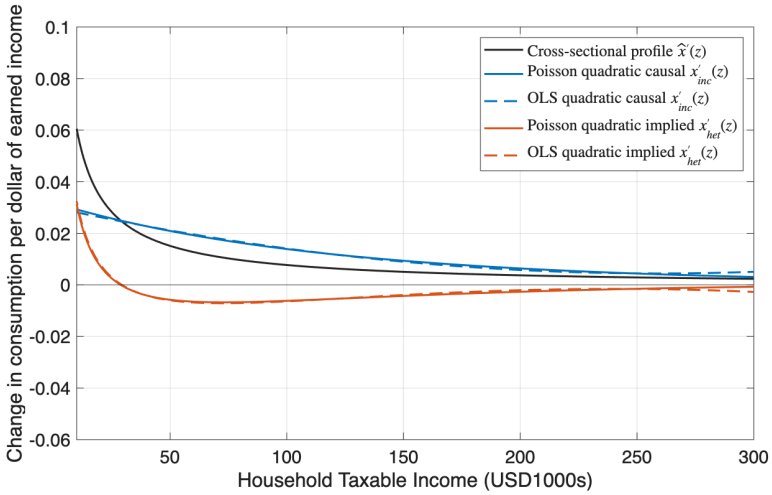



*Figure A.7.* Gasoline: Decomposition of the cross-sectional profile $\hat{x}'(z)$

**Note:** This figure shows how we decompose the cross-sectional relationship between income and **gasoline consumption**. The solid black line is a smoothed estimate of the cross-sectional slope, $\hat{x}'(z)$, at each level of income (from the 2018–2024 CEX surveys). The blue lines show estimates of the causal effect of income on expenditure on gasoline (from our custom hypothetical survey). The orange lines subtract the income effect from the cross-section to obtain the part of the relationship between $x$ and $z$ that comes from preference heterogeneity.

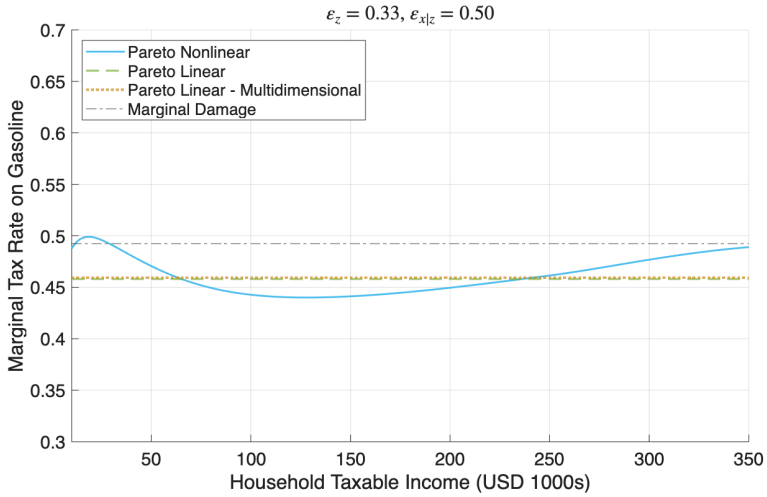



*Figure A.8.* Gasoline Marginal Tax Rates: Unidimensional vs. Multidimensional Heterogeneity

**Note:** This figure plots the Pareto-efficient carbon tax for **gasoline only** in the unidimensional case (Linear and Nonlinear) vs. the optimal 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.10 Pareto-Efficient Marginal Tax Schedules — Electricity, Natural Gas, and Heating Fuels Only

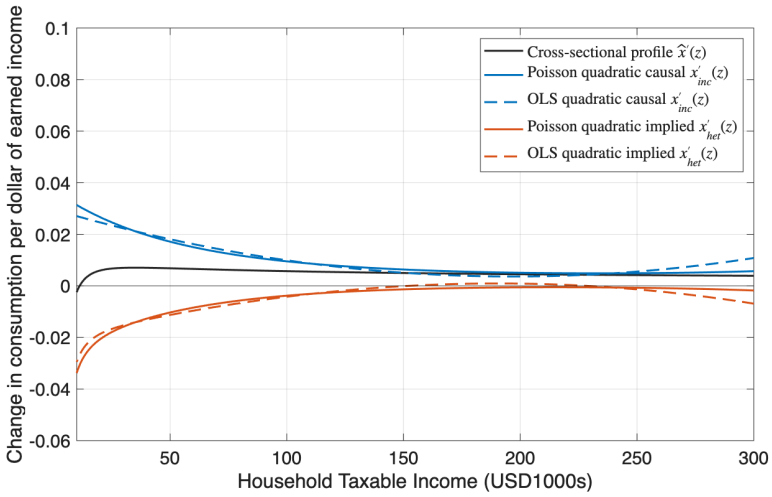



*Figure A.9.* Electricity, Natural Gas, and Heating Fuels: Decomposition of the cross-sectional profile $\hat{x}'(z)$

**Note:** This figure shows how we decompose the cross-sectional relationship between income and consumption of **electricity, natural gas, and heating fuels**. The solid black line is a smoothed estimate of the cross-sectional slope, $\hat{x}'(z)$, at each level of income (from the 2018–2024 CEX surveys). The blue lines show estimates of the causal effect of income on expenditure on the same goods (from our custom hypothetical survey). The orange lines subtract the income effect from the cross-section to obtain the part of the relationship between $x$ and $z$ that comes from preference heterogeneity.

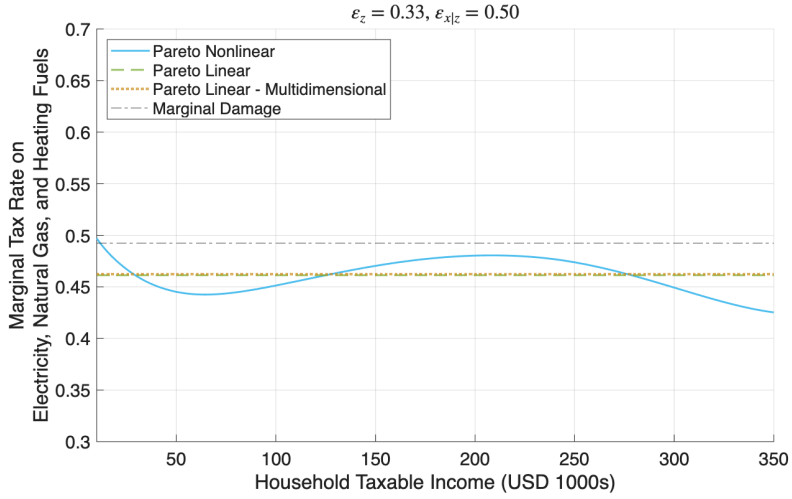



*Figure A.10.* Electricity, Natural Gas, and Heating Fuels Marginal Tax Rates: Unidimensional vs. Multidimensional Heterogeneity

**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 optimal 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.11 Online Survey Demographics Breakdown & Sample Definition

| Ethnicity | Freq. | % | (C) |
|---|---|---|---|
| White | 1178 | 75.9% | 73.9% |
| Black | 209 | 13.5% | 14.3% |
| Asian | 94 | 6.1% | 7.5% |
| Mixed | 47 | 3.0% | 2.5% |
| Other | 24 | 1.5% | 1.7% |

| Age | Freq. | % | (C) |
|---|---|---|---|
| 25-29 | 276 | 17.8% | 16.9% |
| 30-34 | 296 | 19.1% | 18.0% |
| 35-39 | 316 | 20.4% | 17.3% |
| 40-44 | 215 | 13.9% | 16.8% |
| 45-49 | 243 | 15.7% | 15.2% |
| 50-54 | 206 | 13.3% | 15.9% |

| Sex | Freq. | % | (C) |
|---|---|---|---|
| Female | 798 | 51.4% | 49.7% |
| Male | 754 | 48.6% | 50.3% |

*Table A.8.* 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.9.* Online Survey Response Quality: Sample Exclusions

| | N | % |
|---|---|---|
| **Total # of Respondents** | 2,214 | 100.00 |
| Suspected Bots | 18 | 0.81 |
| Speedsters | 421 | 19.02 |
| Failed Attention Check | 44 | 1.99 |
| Respondents Outside the Age Range | 9 | 0.41 |
| Inconsistent or Outlier Income Response | 49 | 2.21 |
| Zero Expenditure or Inconsistent MPC | 121 | 5.47 |
| **Final Sample** | 1,552 | 70.10 |

**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 Optimal Levels of the Income and Externality Taxes

An approach very similar to the one used to derive the Pareto-efficiency conditions in the main text can be used to characterize the optimal levels of the externality and income taxes. The fact that a reform to either tax schedule is not by itself distribution-neutral means that such reforms 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:

$$g(w) = \gamma(w)\frac{u_c(c(w), x(w), z(w) \mid w)}{\lambda} + \left[\mathcal{T}_x'\left(x\left(w\right)\right) - \frac{\mathcal{D}'\left(X\right)}{\lambda}\right]\left.\frac{\partial \hat{x}(w)}{\partial I}\right|_{z=z(w)} + \left\{\mathcal{T}_z'(z(w)) + \left[\mathcal{T}_x'(x\left(w\right)) - \frac{\mathcal{D}'\left(X\right)}{\lambda}\right] x_{\text{inc}}'(z(w))\right\}\frac{\partial z(w)}{\partial I}, \tag{58}$$

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 $\lambda$ 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 4. We defer the derivation of these conditions to Appendix D.

**Proposition 4.** *The optimal tax on the externality-generating good must satisfy equation 59:*

$$\frac{\mathcal{T}_x'\left(\hat{x}\left(z\right)\right) - \frac{\mathcal{D}'(X)}{\lambda}}{1 + \mathcal{T}_x'\left(\hat{x}\left(z\right)\right)} = \eta_{x,z}^{\textit{Taste}}\left(z\right)\frac{1}{\varepsilon_{x|z}\left(z\right)}\frac{1 - H_z\left(z\right)}{z h_z\left(z\right)}\left(1 - \bar{g}_+\left(z\right)\right), \tag{59}$$

*and the optimal tax on income must satisfy equation 60:*

$$\frac{\mathcal{T}_z'\left(z\right) + \left[\mathcal{T}_x'\left(x\left(\hat{w}\left(z\right)\right)\right) - \frac{\mathcal{D}'(X)}{\lambda}\right] x_{inc}'(z)}{1 - \mathcal{T}_z'\left(z\right)} = \frac{1}{\varepsilon_z\left(z\right)}\frac{1 - H_z\left(z\right)}{z h_z\left(z\right)}\left(1 - \bar{g}_+\left(z\right)\right), \tag{60}$$

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

Equation 59 closely resembles the standard "ABC" expression for the optimal nonlinear income tax (Mirrlees, 1971), with dirty-good consumption as the taxed base. Because $x$ is strictly increasing in $z$, agents with incomes above $z$ are exactly those consuming more than $\hat{x}(z)$: $1 - H_z(z) = 1 - H_x(\hat{x}(z))$, $h_z(z) = \hat{x}'(z)h_x(\hat{x}(z))$, and $\bar{g}_+(z)$ is the average welfare weight of agents consuming more than $\hat{x}(z)$. Substituting into equation 59 gives

$$\frac{\mathcal{T}_x'(\hat{x}(z)) - \frac{\mathcal{D}'(X)}{\lambda}}{1 + \mathcal{T}_x'(\hat{x}(z))} = \frac{x_{\text{het}}'(z)}{\hat{x}'(z)} \cdot \frac{1}{\varepsilon_{x|z}(z)} \cdot \frac{1 - H_x(\hat{x}(z))}{\hat{x}(z)h_x(\hat{x}(z))}\left(1 - \bar{g}_+(z)\right).$$

When the taste elasticity is positive, a higher marginal rate at $\hat{x}(z)$ is optimal if redistribution away from those consuming more than $\hat{x}(z)$ is valued ($\bar{g}_+(z)$ small), demand is inelastic ($\varepsilon_{x|z}(z)$ small), and/or much consumption lies above $\hat{x}(z)$ relative to the base distorted locally. The adjustment is scaled by $x'_{\text{het}}(z)/\hat{x}'(z)$—the share of the cross-sectional consumption gradient attributable to preferences; a negative taste elasticity reverses each comparative static, and a zero one restores the Pigouvian rate.

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 $x$, $\mathcal{D}'(X)$. The second is the presence of the taste elasticity, $\eta_{x,z}^{\text{Taste}}(z)$, which captures the same intuition as above.

Equation 60 is almost identical to the usual "ABC" formula. A higher marginal tax rate on income is optimal if redistribution is valued ($\bar{g}_+(z)$ is small), 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 $\left[\mathcal{T}'_x\left(x\left(\hat{w}\left(z\right)\right)\right) - \mathcal{D}'\left(X\right)/\lambda\right] x'_{\text{inc}}(z)$ 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.

# C Surplus Forgone at the Pigouvian Tax

The Pareto-efficient carbon tax in our calibration is slightly below marginal damage. We now ask how much is gained by making this adjustment rather than simply setting the tax at the Pigouvian level. This appendix applies a back-of-the-envelope calculation to approximate this difference in surplus, measured in dollars per household per year, defined as government revenue net of external damage. In the benchmark linear case, the additional gain is approximately $1.05 per household per year.

Write $d \equiv \mathcal{D}'(X)/\lambda$ for marginal damage per dollar of dirty expenditure (held fixed). Consider a small change in the linear carbon tax $t_x$, together with an income-tax adjustment that offsets the change in each household's tax liability at its original choices. This is the distribution-neutral reform of Section 2.5.

We can use equation 25 to identify the marginal effect of such a reform, evaluated at income level $z$ and initial tax rate $t_x$, which we denote by $\Delta\mathcal{R}(z; t_x)$. Taking its expectation across $z$ yields the change in surplus per household per unit increase in the carbon tax.

At $t_x = d$, the two channels with the tax-minus-damage wedge (i.e. the compensated dirty consumption effect $\Delta\mathcal{R}^{x|z}$, and the consumption effect induced by changes in taxable income $\Delta\mathcal{R}^{z\to x}$) are equal to zero, and thus we get

$$\mathbb{E}_z[\Delta\mathcal{R}(z;d)] = \mathbb{E}_z[\Delta\mathcal{R}^z(z;d)] = \mathbb{E}_z\left[\frac{\mathcal{T}_z'(z)}{1-\mathcal{T}_z'(z)}\,\varepsilon_z(z)\,z\,x'_{\text{het}}(z)\right]. \tag{61}$$

This expectation is negative in our calibration. Starting at the Pigouvian tax, a marginal reduction together with the compensating income-tax adjustment raises surplus.

Next, let $G(t_x)$ denote the additional annual surplus per household from moving from the Pigouvian level $d$ to a candidate tax $t_x$. By definition, $G(d) = 0$. Summing up the marginal surplus along the tax change gives

$$G(t_x) \equiv \int_d^{t_x} \mathbb{E}_z[\Delta\mathcal{R}(z;s)]\,ds. \tag{62}$$

By construction, we can write $G(t_x^*)$ as the additional surplus at the Pareto-efficient tax level, denoted by $t_x^*$. It measures what is forgone by choosing the Pigouvian level rather than making the optimal adjustment accounting for taste heterogeneity.

Holding the estimated sufficient statistics and marginal income tax fixed, we can evaluate this integral analytically using equation 25:

$$G(t_x) = (t_x - d)\mathbb{E}_z[\Delta\mathcal{R}^z(z;d)] \tag{63}$$

$$
\begin{aligned}
&- \mathbb{E}_z[\varepsilon_{x|z}\hat{x}(z)] \left[ t_x - d - (1+d)\log\left(\frac{1+t_x}{1+d}\right)\right] \\
&+ \frac{(t_x-d)^2}{2}\mathbb{E}_z\left[\frac{\varepsilon_z(z)z}{1-\mathcal{T}_z'(z)}x'_{\text{inc}}(z)x'_{\text{het}}(z)\right].
\end{aligned}
$$

In our preferred baseline elasticity scenario ($\varepsilon_z = 0.33$ and $\varepsilon_{x|z} = 0.5$), marginal damage is 49.24 cents per dollar of dirty expenditure. The Pareto-efficient linear tax is 3.27 cents below the Pigouvian benchmark. Evaluating the integral at this tax rate gives an additional annual surplus of \$1.05 per household, or 0.0013% of mean consumption expenditure.

The multidimensional calculation for the linear tax accounts for variation in the causal income response among households with the same income. In equation 63, we replace the product of the mean slopes by

$$
\mathbb{E}[x'_{\text{inc}}x'_{\text{het}} \mid z] = \bar{x}'_{\text{inc}}(z)\bar{x}'_{\text{het}}(z) - \mathbb{V}(x'_{\text{inc}} \mid z). \tag{64}
$$

The variance term reduces the benchmark additional gain to \$0.99 per household per year. This calculation compensates households on average at each income level, rather than compensating every household. It does not take into account redistribution within income groups and the income effects caused by unequal compensation (see Appendix B).

For the nonlinear tax of Proposition 1, we repeat the calculation separately at each income level, treating $\mathcal{T}_x'(\hat{x}(z))$ as the marginal rate to adjust at that income. We choose each adjustment separately and then average the resulting gains across the income distribution. This gives a gain of \$1.22 per household per year in our preferred baseline elasticity scenario.

The last two columns of Table C.1 apply the same calculation over a wider tax change. We integrate the marginal surplus formula from a zero tax rate to the Pigouvian level, and from a zero tax rate to the Pareto-efficient tax rate. For the benchmark linear tax, we obtain surplus values of \$267.36 and \$268.41 per household per year. These values illustrate that the benefit of introducing a tax at the Pigouvian level is much larger than the additional benefit of adjusting it to account for distributional concerns. These comparisons hold the estimated sufficient statistics fixed over a much larger tax change than for the calculation comparing the Pigouvian level to the Pareto-efficient level. They should therefore be read as illustrative.

*Table C.1.* Approximate Annual Surplus Forgone at the Pigouvian Tax

| | | | Surplus forgone at Pigou (per household) | | Illustrative surplus ($ per household) | |
|---|---|---|---|---|---|---|
| $\varepsilon_z$ | $\varepsilon_{x\|z}$ | Instrument | Dollars | % of consumption | Zero to Pigou | Zero to optimum |
| 0.33 | 0.25 | Linear | 2.06 | 0.0025 | 118.45 | 120.51 |
| | | Multidimensional | 1.86 | 0.0023 | 131.82 | 133.68 |
| | | Nonlinear | 2.38 | 0.0029 | 118.45 | 120.83 |
| **0.33** | **0.50** | **Linear** | **1.05** | **0.0013** | **267.36** | **268.41** |
| | | **Multidimensional** | **0.99** | **0.0012** | **280.73** | **281.73** |
| | | **Nonlinear** | **1.22** | **0.0015** | **267.36** | **268.58** |
| 0.33 | 0.75 | Linear | 0.70 | 0.0009 | 416.27 | 416.97 |
| | | Multidimensional | 0.68 | 0.0008 | 429.64 | 430.32 |
| | | Nonlinear | 0.82 | 0.0010 | 416.27 | 417.09 |
| 0.70 | 0.25 | Linear | 8.92 | 0.0109 | 84.29 | 93.22 |
| | | Multidimensional | 7.33 | 0.0089 | 112.66 | 119.99 |
| | | Nonlinear | 10.22 | 0.0124 | 84.29 | 94.52 |
| 0.70 | 0.50 | Linear | 4.63 | 0.0056 | 233.21 | 237.83 |
| | | Multidimensional | 4.15 | 0.0051 | 261.57 | 265.72 |
| | | Nonlinear | 5.34 | 0.0065 | 233.21 | 238.55 |
| 0.70 | 0.75 | Linear | 3.13 | 0.0038 | 382.12 | 385.24 |
| | | Multidimensional | 2.90 | 0.0035 | 410.48 | 413.38 |
| | | Nonlinear | 3.62 | 0.0044 | 382.12 | 385.73 |

**Note:** All dollar amounts are per household per year and measure revenue net of external damage. We highlight in bold our preferred baseline elasticity scenario.

# D 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:

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

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:

$$\varepsilon_{x|z}(w) \equiv -\frac{1+\mathcal{T}_x'(x(w))}{\tau_x'(x(w))\,x(w)}\left.\frac{\partial x(w;z(w))}{\partial \kappa}\right|_{z(w),\kappa=\tau_z(z(w))+\tau_x(x(w))=\tau_z'(z(w))=0}. \tag{66}$$

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.

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

This implies marginal income tax rate changes of

$$\tau_z'(z) = -\hat{x}'(z)\,\tau_x'(\hat{x}(z)). \tag{68}$$

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 2.8 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, namely the following three components multiplied by $\tau_x'(\hat{x}(z))$:

$$\Delta\mathcal{R}^z(z) = \mathcal{T}_z'(z)\frac{z\varepsilon_z(z)}{1-\mathcal{T}_z'(z)}x_{\text{het}}'(z)$$

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

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

Integrating over all types in the economy, the social welfare impact is:

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

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

A Necessary Condition for Optimality

If the tax system is 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,[27] we obtain equation 29, which we restate here for convenience

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

It is then straightforward to rearrange this to obtain equation 31, also restated:

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

where:

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

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. □

[27] 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 $z$. Letting $f(z) \equiv (\Delta\mathcal{R}^z(z) + \Delta\mathcal{R}^{x|z}(z) + \Delta\mathcal{R}^{z\to x}(z))h_z(z)$ and $\rho(z) \equiv \tau'_x(\hat{x}(z))$, equation 29 follows immediately.

*Proof of Proposition 4.* 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 by 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:

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

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

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

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

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

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

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:

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

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

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

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

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

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

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

Welfare Effects of a Tax Reform

For any given pair of continuously differentiable reform functions $\tau_z$ and $\tau_x$, an infinitesimal reform away from $\mathcal{T}_z$ and $\mathcal{T}_x$ has the following effect on social welfare:

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

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

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

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

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

where $\lambda$ is the multiplier on the planner's budget constraint, and $\mathcal{D}'\left(X\right)/\lambda$ 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 is 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 $\left(\tau_z\left(\cdot\right),\tau_x\left(\cdot\right)\right)$ 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 $\tau_x\left(x\right)=0$ 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 $\tau_z(\cdot)$:

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

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:[28]

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

[28]Here we use a slightly different variant of the 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).

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

$$1-\bar{g}_{+}(z)=\frac{\int_{z}^{\infty}\left(1-g\left(\hat{w}\left(s\right)\right)\right)h_{z}\left(s\right)\mathrm{d}s}{1-H_{z}(z)}, \tag{77}$$

we obtain the optimality condition stated in Proposition 4.

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 $\tau_z\left(z\right)=0$ 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 $\tau_x(\cdot)$:

$$\int\left(1-g(\tilde{w}(x))\right)\tau_{x}\left(x\right)h_{x}\left(x\right)\mathrm{d}x-\int\left[\frac{\mathcal{T}_{x}'\left(x\right)-\frac{\mathcal{D}'(X)}{\lambda}}{1+\mathcal{T}_{x}'\left(x\right)}\varepsilon_{x|z}\left(\tilde{w}\left(x\right)\right)x\ldots\right.$$
$$\left.+\left(\frac{\mathcal{T}_{z}'(\tilde{z}(x))+\left[\mathcal{T}_{x}'(x)-\frac{\mathcal{D}'(X)}{\lambda}\right]x_{\text{inc}}'(\tilde{z}(x))}{1-\mathcal{T}_{z}'\left(\tilde{z}(x)\right)}\right)x_{\text{inc}}'(\tilde{z}(x))\varepsilon_{z}(\tilde{w}(x))\tilde{z}(x)\right]\tau_{x}'(x)h_{x}(x)\mathrm{d}x=0,$$

where $\tilde{w}\left(x\right)$ and $\tilde{z}\left(x\right)$ 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:

$$\begin{aligned}&\frac{\mathcal{T}_{x}'\left(x\right)-\frac{\mathcal{D}'(X)}{\lambda}}{1+\mathcal{T}_{x}'\left(x\right)}\varepsilon_{x|z}\left(\tilde{w}\left(x\right)\right)xh_{x}\left(x\right)\\&\qquad+x_{\text{inc}}'(\tilde{z}(x))\left(\frac{\mathcal{T}_{z}'(\tilde{z}(x))+\left[\mathcal{T}_{x}'(x)-\frac{\mathcal{D}'(X)}{\lambda}\right]x_{\text{inc}}'(\tilde{z}(x))}{1-\mathcal{T}_{z}'\left(\tilde{z}(x)\right)}\right)\varepsilon_{z}(\tilde{w}(x))\tilde{z}(x)h_{x}(x)\\&\qquad=\int_{x}^{\infty}\left(1-g\left(\tilde{w}\left(y\right)\right)\right)h_{x}\left(y\right)\mathrm{d}y.\end{aligned} \tag{78}$$

We can then write equation 78 in terms of the distribution of taxable income. Because both $x(w)$ and $z(w)$ are strictly increasing in $w$ (consistent with our empirical estimates), the mapping from incomes to consumption preserves orientation, and therefore we can write:

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

Using this result, multiplying both sides of equation 78 by $x_{\text{inc}}'(z)+x_{\text{het}}'(z)$, and inserting

the optimality condition for the income tax (equation 76) we can rewrite equation 78 as:

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

which further simplifies to

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

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 proposition. □

*Proof of Proposition 2.* The proof here proceeds in a very similar manner to the fully nonlinear case. Each agent now maximizes her utility subject to her slightly modified budget constraint, producing indirect utility $v(w)$.

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

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, $\bar{\mathcal{R}}$.

$$\max\boldsymbol{\mathcal{W}}=\int\gamma(w)v(w)dF-\mathcal{D}(X) \tag{81}$$
$$\text{s.t.}\quad\mathcal{R}\equiv\int\left(T_z(z(w))+t_xx\right)dF(w)=\bar{\mathcal{R}} \tag{82}$$
$$\text{and}\quad X=\int x(w)dF$$

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.

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

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

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

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

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

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: $\tau_x\left(x\left(\hat{w}\left(z\right);z\right)\right)=\rho\times\hat{x}(z)$ for some constant $\rho$; and $\tau_z\left(z\right)=-\rho\times\hat{x}(z)$. For such a reform, we have $\tau_x'\left(x\left(\hat{w}\left(z\right);z\right)\right)=\rho$ and:

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

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

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

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

$$\left.\cdots+\frac{t_x-\frac{\mathcal{D}'(X)}{\lambda}}{1+t_x}\varepsilon_{x|z}\left(w\right)x\left(w\right)\right\}f\left(w\right)\mathrm{d}w. \tag{85}$$

In this case, this simplifies to:

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

$$\cdots - \left\{ \frac{t_x - \frac{\mathcal{D}'(X)}{\lambda}}{1 + t_x} \varepsilon_{x|z}(w)\, x(w) \right\} f(w)\, \mathrm{d}w. \tag{86}$$

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 3.* Setup and Behavioral Responses

Consider a population of agents indexed by type vectors $(w, \theta)$ 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 $\kappa$ 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:

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

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 $dx/d\kappa$:

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

It is convenient to separate each response into a compensated component and an income-effect component. Let

$$\begin{aligned}\left.\frac{dz^c(w,\theta)}{d\kappa}\right|_{\kappa=0} &\equiv -\frac{\varepsilon_z(w,\theta)z(w,\theta)\tau_z'(z(w,\theta))}{1-\mathcal{T}_z'(z(w,\theta))} + \chi(w,\theta)\tau_x'(x(w,\theta)), \\ \left.\frac{dx^c(w,\theta)}{d\kappa}\right|_{\kappa=0} &\equiv -\frac{\varepsilon_{x|z}(w,\theta)x(w,\theta)}{1+\mathcal{T}_x'(x(w,\theta))}\tau_x'(x(w,\theta)) + \frac{\partial x}{\partial z}\left.\frac{dz^c}{d\kappa}\right|_{\kappa=0},\end{aligned}$$

so that

$$\begin{aligned}\left.\frac{dz}{d\kappa}\right|_{\kappa=0} &= \left.\frac{dz^c}{d\kappa}\right|_{\kappa=0} - \eta_z(w,\theta)[\tau_z(z)+\tau_x(x)], \\ \left.\frac{dx}{d\kappa}\right|_{\kappa=0} &= \left.\frac{dx^c}{d\kappa}\right|_{\kappa=0} - \left[\eta_{x|z}(w,\theta) + \frac{\partial x}{\partial z}\eta_z(w,\theta)\right][\tau_z(z)+\tau_x(x)].\end{aligned}$$

Welfare Effects of a Reform

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

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

where $\frac{\partial v}{\partial c}$ denotes the household's marginal utility of disposable income, evaluated at her optimal bundle. 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 follows. Collecting the income-effect components of these responses with the mechanical term converts the welfare weight $\frac{\gamma(w,\theta)}{\lambda}\frac{\partial v}{\partial c}$ into the augmented social marginal welfare weight $g(w,\theta)$ defined below, so that

the remaining behavioral terms involve only the compensated responses:

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

where the social marginal utility of income is:

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

Change of Variables to $(z,x)$

Assume that, conditional on $\theta$, $z(w,\theta)$ 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 $H_x(x|z) = P(x(w,\theta) \le x \mid z(w,\theta) = z)$ 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:

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

Vertically Neutral Reform

Consider a linear tax reform $\tau_x(x) = x$ with compensating income tax changes $\tau_z(z) = -\mathbb{E}[\tau_x(x(w,\theta)) \mid z(w,\theta) = z] = -\bar{x}(z)$, 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:

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

This is the residual of the cross-sectional relationship between $x$ and $z$ after removing the causal income effect for type $\theta$ 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:

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

Under the assumption that $\mathbb{E}[\mathbb{C}(g(w,\theta), x(w,\theta)|z)] = 0$, the optimal tax satisfies equation (46):

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

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

□

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

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

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

Substituting into the general formula and collecting terms yields equation (49):

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

where:

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

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

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. □

# E Survey Questionnaire Appendix

## E.1 Survey Link and Metadata

**Survey link:** https://umich.qualtrics.com/jfe/form/SV_egEUDcSVdWIjkTY

**Survey flow:** Landing Screen; Actual Expenditure Block; Attention Check; Causal Response Elicitation; Demographics; Commute Block.

The questionnaire below reproduces the respondent-facing text, answer options, and relevant display logic. Administrative Qualtrics timing fields and embedded Prolific/session fields are omitted; graphic-only display elements are noted where they appear.

## E.2 Prolific-Provided Respondent Demographics

Respondents were recruited through Prolific, and the survey records were linked to Prolific profile data using the Prolific participant IDs.

- **Age:** integer age in years.
- **Sex:** Female; Male.
- **Ethnicity:** Asian; Black; Mixed; Other; White.
- **Country of birth, country of residence, and nationality:** country fields from the respondent's Prolific profile.
- **Language:** respondent language recorded by Prolific.
- **Student status:** Yes; No.
- **Employment status:** Full-Time; Part-Time; Unemployed (and job seeking); Not in paid work (e.g. homemaker, retired or disabled); Due to start a new job within the next month.

## E.3 Survey Questions

### Landing Screen

1. **Welcome**

   Thank you for your interest in taking this survey. Your participation is valuable for our research, which will help understand how people make spending decisions.

   Before proceeding, please review our Participant Information Sheet.

   By moving forward, you confirm that:

   1. You have read and understood the Participant Information Sheet
   2. You voluntarily agree to participate in this research
   3. You understand you can withdraw at any time

   Click → to begin the survey.

### Actual Expenditure Block

1. Please indicate your total **annual** household income **before** taxes paid to the government, transfers received from the government, and deductions in the current fiscal year (2024). Please include income from wages, salaries, self-employment, investments, social security, retirement accounts, and any other sources.

   *If you are not sure, it is okay to provide your best estimate. However, an accurate estimate of your **annual** before-tax income would be very valuable for our research.*

   *Response format:* Open numeric entry, whole dollars, minimum 0. The live page displays implied household monthly before-tax income from the annual entry.

2. Please indicate your total **annual** household income **after** taxes paid to the government, transfers received from the government, and deductions in the current fiscal year (2024). You can think of this as your "take-home" income.

   *If you are not sure, it is okay to provide your best estimate. To reiterate, please enter your **annual** after-tax income, but note that your corresponding **monthly** household "take-home" income will automatically be displayed below.*

   *Response format:* Open numeric entry, whole dollars, minimum 0. The live page displays implied household monthly after-tax income from the annual entry.

   *Display element:* The survey displays a graphic-only expenses icon before the next item.

3. In the rest of the survey, we will ask you about your consumption across different expenditure categories such as food, housing expenditures, or utilities. But for now, we want to know whether you are paying attention, so please select "Healthcare (health insurance, medical services, drugs, and medical supplies)."

   *Response options:*

   - Food (both at and away from home) and beverages (including alcoholic beverages)
   - Water, sewage, and other public services (e.g. trash and garbage collection)
   - Gasoline, diesel and/or other motor fuels
   - Housing expenditure excluding utilities (includes rent, owned dwelling expenditures, mortgage payments, home insurance, furnishing and appliances, and domestic services)
   - Healthcare (health insurance, medical services, drugs, and medical supplies)
   - Electricity, natural gas (for heating and cooking), fuel oil and other home heating fuels

4. How much of your household's [monthly after-tax income amount] **monthly after-tax income** did you spend (as opposed to save) on each of the following expenditure categories, on average over the last 3 months?

   *These categories, including savings, are exhaustive, and the total amount (at the bottom) will automatically sum up to your household's total monthly after-tax income. Your household's average monthly savings will automatically be calculated and displayed below.*

*Response format:* Constant-sum dollar entries by category. Savings and total are displayed and calculated on screen. The live page displays household average monthly savings and uses a "Validate and Continue" button.

*Note:* The QSF/PDF placeholder is "MonthExp"; the live survey replaces it with the monthly dollar amount implied by the respondent's annual after-tax income.

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

### Attention Check

*Display element:* The survey displays the permanent-income-increase graphic before the questions in this block. The same graphic appears again before the causal response elicitation question; Figure E.1 shows that screen.

1. Earlier you told us your after-tax household income last year was [annual after-tax income].

   Now imagine that you or someone in your household receives an **unexpected, permanent pay raise** (or a similar windfall such as a new government benefit, or a tax credit) that **boosts your after-tax income by exactly $1,000** forever.

   - Starting now your after-tax income becomes [annual after-tax income] **+ $1,000**
   - Every future year it will be **$1,000 higher than whatever amount you would otherwise have expected**.

   In this scenario, will your household income be higher **three years from now** than what you originally expected?

   *Note:* The QSF/PDF placeholder is "AnnAfterTaxInc"; the live survey replaces it with the respondent's annual after-tax income.

   *Response options:*

   - Yes
   - No

2. By how much has your household income increased in this scenario, compared to what you originally expected?

   *Response options:*

   - $500 of total lifetime income
   - $1,000 of total lifetime income
   - $500 each year
   - $1,000 each year

### Causal Response Elicitation

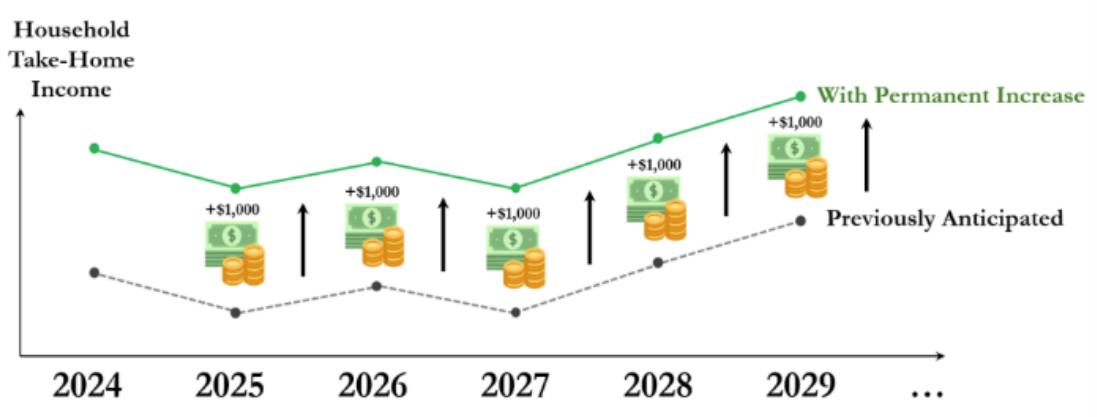
*Display element:* The survey displays the permanent-income-increase graphic immediately above the following question.

1. 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.*

   *Response format:* Constant-sum dollar entries by category. Savings and total are displayed and calculated on screen, with the total summing to $1,000. The live page displays the additional amount saved each year and uses a "Validate and Continue" button.

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




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 E.1.* Permanent-income-increase graphic and survey screen for the causal consumption response elicitation block

## Demographics

1. How would you describe the area where you live?
   *Response options:*
   - Urban (e.g., city or a large town)
   - Suburban (e.g., a residential area near a city)
   - Rural (e.g., a countryside, a small town, or a village)

2. In which state do you currently reside?
   *Response format:* Dropdown menu: Alabama; Alaska; Arizona; Arkansas; California; Colorado; Connecticut; Delaware; District of Columbia; Florida; Georgia; Hawaii; Idaho; Illinois; Indiana; Iowa; Kansas; Kentucky; Louisiana; Maine; Maryland; Massachusetts; Michigan; Minnesota; Mississippi; Missouri; Montana; Nebraska; Nevada; New Hampshire; New Jersey; New Mexico; New York; North Carolina; North Dakota; Ohio; Oklahoma; Oregon; Pennsylvania; Rhode Island; South Carolina; South Dakota; Tennessee; Texas; Utah; Vermont; Virginia; Washington; West Virginia; Wisconsin; Wyoming; Outside the US

3. What is the highest level of education you have completed or the highest degree you have received?
   *Response options:*
   - High school, high school equivalent, or lower
   - Some college, no degree
   - Associate's degree in college (e.g., AA, AS)
   - Bachelor's degree (e.g., BA, BS)
   - Masters (MA, MS, MEd), professional (JD, MD, DDS) or doctorate degree (PhD, EdD)

4. Please indicate your agreement with the following statement: I spend my entire income on paper clips.
   *Response options:*
   - Strongly Disagree
   - Disagree
   - Agree
   - Strongly Agree

5. Which category best represents your household?
   *Response options:*
   - Married Couple only
   - Married Couple, with children
   - One parent, with children
   - Single, without children
   - Other (please specify) _____

6. How many children are part of your household and younger than 18 years of age?
   *Response format:* Open numeric entry, integer from 0 to 50.

   *Note:* Displayed if the respondent selected Married Couple, with children; One parent, with children; or Other (please specify) in the household-type question.

7. How many of these children part of your household are 5 years or younger?
   *Response format:* Open numeric entry, integer from 0 to 50. The live page displays a reminder that the value should be no more than the reported number of children under 18 and enforces that cap.

   *Note:* Displayed if the respondent selected Married Couple, with children; One parent, with children; or Other (please specify) in the household-type question.

8. What is your spouse/partner's age? Enter a number only, in years.
   *Response format:* Open numeric entry, integer from 1 to 130.

   *Note:* Displayed if the respondent selected Married Couple only; Married Couple, with children; or Other (please specify) in the household-type question.

### Commute Block

1. How far is your commute (one-way) from your home to your work location? Enter your best estimate of the **distance in miles**. If you do not work, or work from home, simply enter 0.

   *Response format:* Open numeric entry, minimum 0.

2. What is your primary mode of transportation/commuting?
   *Response options:*
   - Car
   - Public transportation (bus/train)
   - Ride-sharing app
   - Bicycle
   - Walking
   - Other (please specify) _____

3. How easy or difficult would it be for you to switch to a more environmentally friendly mode of transportation (one that produces less carbon emissions)?
   *Response options:*
   - Very Difficult
   - Difficult
   - Nor Easy, nor Difficult
   - Easy
   - Very Easy