CESifo WORKING PAPERS

11174
2024

June 2024

Pigouvian Income Taxation

Lassi Ahlvik, Matti Liski, Mikael Mäkimattila

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11174*

Pigouvian Income Taxation

Abstract

This paper develops a mechanism design approach to study externalities and re-distribution. The mechanism screens individuals' social weights to strike a balance among broad distributional objectives, incentives to work, and incentives to reduce externalities. The welfare-optimal allocation can be decentralized through income taxation, defining income-dependent externality payments. Two applications use individual-level administrative data on incomes, pollution measures, and financial burdens to demonstrate how population characteristics shape the optimal policy on carbon emissions.

JEL-Codes: D820, H210, H230, Q540, Q580.

Keywords: Pigouvian taxation, optimal income taxation, inequality, climate change.

Lassi Ahlvik
Department of Economics and Management
University of Helsinki / Finland
lassi.ahlvik@helsinki.fi

Matti Liski
Department of Economics
Aalto University and Helsinki GSE /
Helsinki / Finland
matti.liski@aalto.fi

Mikael Mäkimattila
Department of Economics
Aalto University and Helsinki GSE / Finland
mikael.makimattila@aalto.fi

Pigouvian Income Taxation

Lassi Ahlvik
Department of Economics and Management, University of Helsinki and Helsinki GSE

Matti Liski
Department of Economics, Aalto University and Helsinki GSE

Mikael Mäkimattila
Department of Economics, Aalto University and Helsinki GSE

1. Introduction

Grand externality problems call for grand policy solutions. However, these solutions may encounter difficulties in gaining policy traction if the distribution of policy costs and benefits is perceived as unfair. Economists generally accept that equity considerations can distort policies when individuals' economic outcomes are influenced by exogenous factors such as background, ability, and health. The same principle can apply when society shifts course to fight grand externality problems like climate change, resulting in unanticipated and unequally distributed costs.

We thank Felix Bierbrauer, Jukka Pirttilä, Rick van der Ploeg, and participants at EEA 2023, NAERE 2022 and 2023, Toulouse energy workshop 2023, and at seminars of Bergen NHH, Cambridge MIT, Cologne EWI, Helsinki GSE, Paris PSE, and Stockholm SSE for valuable comments and discussions. We thank Elisabeth Isaksen for sharing elasticity estimates in the vehicle application, and Arttu Ahonen for assistance with the register data build-up. This project has been funded by Research Council Finland, the Finnish Cultural Foundation, and the European Union (ERC GRETA, 101116659). Views and opinions expressed are how-ever those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

To reduce inequality, efficiency may have to be compromised; for instance, setting a corrective tax below the level considered efficient would reduce its burden. One could also compensate individuals through direct transfers but, without complete information, scarce funds might not reach those with the greatest needs. From a broader perspective, society already has tools to address inequality, notably progressive income taxation and the associated transfers for redistribution. Could these also serve to tackle the inequality from externalities and policies designed for them? The answer to this question, as we find in this paper, has far-reaching consequences: efficiency-equity tradeoffs for incomes and externalities are interconnected and must be addressed jointly. To do so, we develop a mechanism design approach to jointly determine income and externality taxation.

In our model, the policy maker lacks knowledge of (i) how the externality affects each individual, (ii) what financial burden the externality policy places on individuals' private economy, and (iii) what the income-earning potential of each individual is. The policy maker has preferences for the distribution of outcomes, which we capture by welfare weights. The weights encompass broad objectives, including efficiency-equity tradeoffs for both incomes and externalities.

Our main result is that the optimal mechanism ties together individuals' incomes and corrective taxes, resulting in income-dependent pricing of externalities. Conditioning the externality tax on income leads to gains in both efficiency and equity.[1] For an inequality averse policy maker, the optimal income-dependent externality tax schedule can be either progressive or regressive,[2] and it can lie below or above the efficient, Pigouvian level. These results are new and they can explain when there is a case for stricter climate policies for the wealthy, as proposed in the World Inequality Report (Ch 6, 2022): "*To accelerate carbon emissions reductions among the wealthiest, progressive carbon taxes can become a useful instrument.*"

The policy maker can infer the redistributive value of deviating from the efficient corrective tax on externalities such as pollution by noting that either wealthier or poorer people may find it easier to reduce tax burdens by polluting less. When there is a classical preference for aversion to income inequality, we show that if the poor find it easier to pollute less, the externality tax schedule generally deviates upwards for a transfer from the rich to poor. Conversely, if the rich pollute less, the tax schedule deviates downwards to limit an opposite transfer.

Another, distinct reason to deviate from the efficient corrective tax is aversion to inequality of cost burdens. For instance, the policy maker can be concerned about cost burdens due to externality taxes, regardless of incomes of individuals who bear them. This preference alone leads to a reduced corrective tax. The concern can depend on income, as phenomena such as energy or transport poverty are amplified at lower incomes.[3] In such situations, the policy maker could have a classical aversion to income

[1] In practice, one can make externality-related expenses deductible in income taxation, cover externality-related expenses through social programs, or make the tax income-dependent directly. We discuss examples of these below.

[2] We call an externality tax *progressive* if the tax rate increases in income and *regressive* in the opposite case.

[3] Energy poverty refers to a situation where a household cannot afford to pay their energy bills. In the UK, a household is classified as "fuel poor" if the required fuel costs are at least 10% of the household's

equality and a parallel aversion to inequality of cost burdens, with the latter diminishing in individuals' incomes. Formally, we show that welfare weights that satisfy a certain intuitive decreasing differences condition can capture such preferences.

The level of the corrective tax is intimately linked to the optimal shape of the tax schedule, that is, to its progressivity or regressivity. Starting with a simple tax reform, we develop three separate determinants of externality tax progressivity (regressivity), all functions of income: (i) aversion to inequality of cost burdens, (ii) behavioral response of emissions, and (iii) of earnings. For example, when the aversion to inequality of cost burdens declines in income, the planner cares more about low-income polluters' cost burdens than those of the high-income polluters. This alone would lead the planner to implement a progressive reform, that is, to choose a higher emissions tax for high income brackets. Similarly for the other two determinants, when the behavioral responses of emissions and earnings vary across income groups, the relevant efficiency-equity tradeoffs vary with income, which leads to a progressive reform under intuitive conditions. Building on this preliminary analysis, we then develop determinants of progressivity (regressivity) for the general optimal tax system.

For the general optimal tax system, we find a strong result on how the income tax progression changes due to the externality. The income tax schedules are adjusted to maintain, on average, the progressivity of the standalone income tax schedule that was optimal prior to the externality problem. Thus, whether the efficiency-equity tradeoffs call for progressive or regressive externality taxation, the mean progression of the tax system is preserved. Intuitively, the externality problem does not change society's taste for redistribution between income groups, reflected by the original tax progression. To preserve the overall progressivity, a reform that introduces progressive externality taxes should be accompanied by a decrease in progressivity of income taxation, leading to an income tax formulas that differ from those in the theory of optimal income taxation (e.g., Diamond, 1998, Saez, 2001). Notably, our formulas can lead to income tax progression, even without a classical aversion to income inequality.

Real-life corrective policies, broadly interpreted, are increasingly conditioned on income. California plans to introduce a system of income-based electricity charges to consumers.[4] In the U.S., household with gross income exceeding a certain limit are not qualified to receive clean vehicle tax credits.[5] In France, there has been an additional electric vehicle subsidy to buy or lease a vehicle for low-income households.[6] Carbon pricing increases costs of electricity and natural gas to incentivize emission reductions, but many countries grant low-income households subsidized rates even if they distort incentives.[7]

income before housing costs (see UK Office for National Statistics, 2023) Similarly, transport poverty refers to situations where people do not have access to affordable transport options.

[4]The proposal is prepared by California Public Utilities Commission (2024a).

[5]The inflation reduction act grants vehicle tax credit, but households with gross income above a certain threshold (e.g. \$300,000 for married couples) are not eligible (Internal Revenue Service, 2024).

[6]The "ecological bonus" supports financial assistance to households who buy or lease a low-emissions vehicle. Drivers with annual income less than a threshold (14,100€) receive an elevated subsidy (European Commission, 2024).

[7]For instance, Low Income Home Energy Assistance Program (LIHEAP, Department of Health and Human Services, 2024) California Alternate Rates for Energy (CARE, California Public Utilities Commission,

Our results contribute to a broader search for principles to assist policy makers in tailoring policies to different income levels.

To help policy makers to see how the theory results can be applied, we consider two empirical cases, both building on individual-level administrative data on incomes, pollution measures, and financial burdens from the policy. In both applications, we use behavioral elasticities by income groups and a distribution of welfare weights in sufficient statistics test for evaluating a tax reform.

The first application uses data for new vehicles, their owners, kilometers driven, and, the incidence of taxes from Finland. People with lower incomes are more likely to buy a polluting vehicle and thus polluters have relatively high welfare weights in the population. In this application, the optimal corrective tax is set below the level considered efficient. Our sufficient statistics test shows that reforming the tax to become progressive is welfare-enhancing. The second application uses data for electricity consumption metered at a household level, and the type of contract. Households whose contracts expired during the crisis were exposed to a price shock, allowing us to estimate the semi-elasticity of consumption responses to a consumption tax among different income groups. Wealthier households consume relatively more and, because they receive lower weights than low-income households, the optimal tax on consumption is set above the level considered efficient. The sufficient statistics test shows that in this application reforming the tax to become regressive increases welfare.

Literature. Our theory and empirical results amplify the relevance of earlier empirical literature showing that carbon pricing policies can have potentially large distributional ramifications (e.g., Känzig, 2023). The environmental economics literature has mostly focused on inequality between income groups (vertical inequality) created by uncompensated carbon pricing (West, 2004, Hassett *et al.*, 2009, Grainger and Kolstad, 2010, Williams III *et al.*, 2015), but, more recently, large variation of policy cost incidence has been documented also within income groups (Fischer and Pizer, 2019, Davis and Knittel, 2019, Cronin *et al.*, 2019, Pizer and Sexton, 2020, Douenne, 2020). This horizontal inequality is often argued to be quantitatively important to the design of climate and energy policies. Our results show *why* horizontal equity matters and *how* it interacts with vertical equity in the optimal policy design. Both dimensions are critical when questions such as "Should the policy treat low and high income polluters differently?" are considered.

The current paper belongs to an emerging strand of the mechanism design literature emphasizing redistribution (e.g., Dworczak *et al.*, 2021, Akbarpour *et al.*, 2024b,a). In this literature, Pai and Strack (2022) develop an optimal externality tax which is nonlinear in the agent's consumption, depending on a relationship between the willingness to pay and the social value of allocating consumption to the agent. The main difference between our setting and the literature is that in our model incomes are observable and endogenous depending on privately known abilities. The behavioral response related to incomes is elementary for the mechanisms driving our results, differentiating us in the

2024b) provide a price reduction for low-income consumers. These subsidies increase production and emissions, see Hahn and Metcalfe (2021).

literature on redistribution mechanisms. In addition, our welfare objective allows representing quite broad vertical and horizontal equity concerns. In developing this broad objective, we build on the literature in assuming that agents' behavior screens their welfare weights.

The designer in our setting faces a screening problem in multiple dimensions, but the core screening is about agents' abilities and costs to participate in emissions reductions. Technically, this random participation model comes close to those in Rochet and Stole (2003), Kleven *et al.* (2009), Lehmann *et al.* (2014), Ahlvik and Liski (2022). In this main model, choices of earnings are continuous and participations to emissions reduction are discrete, but, as an extension, we develop a sufficient statistics result for continuous choices in both dimensions.[8]

Our finding that the welfare-optimal externality payment is income-dependent is new in the public finance literature. The main difference lies in our more general mechanism design approach without exogenous restrictions on the instruments available. Pigouvian tax can be interpreted as a form of commodity taxation which should not be distorted for equity reasons when consumption preferences are homogeneous according to the famous Atkinson-Stiglitz theorem (1976). The result is known to break down if marginal welfare weights correlate with consumption preferences (Saez, 2002).[9] Similar correlations arise in our model but in a different setting because taxes are directly conditioned on incomes, resulting in nonlinear externality payments in income; in a very different setting without externality policies, Ferey *et al.* (2023) derive jointly optimal tax formulas for income and savings allowing earnings-dependent savings taxation.[10]

Organization of the paper. In Section 2, we introduce the direct mechanism and develop the primitives of the allocation problem. In Section 3, we present our first theorem for decentralization in which an income tax schedule and a linear tax on the externality are optimally designed under the restriction that the two taxes are independent. This benchmark result serves to introduce rudimentary efficiency-equity tradeoffs, some of which extend to the general setting. In Section 4, we start by developing a sufficient statistics for reforming the constant externality tax to exhibit dependence on income. The formula developed serves to introduce three measures linked to welfare weights and

[8] For recent advances in multidimensional mechanism design and their application to income taxation, see Spiritus *et al.* (2022) and Golosov and Krasikov (2023).

[9] See Allcott *et al.* (2019) for a corrective motive for taxation (internalities) in this setting.

[10] Related to the literature on the Atkinson-Stiglitz theorem, Kaplow (2012) and Jacobs and De Mooij (2015) assume homogeneity of preferences with respect to the externality, which leads to a separation of Mirrleesian and Pigouvian taxation. Cremer *et al.* (1998) and Cremer *et al.* (2003) characterize the optimal tax also without homogeneity within income groups and show that the optimal externality taxation can become nonlinear in quantities, but independent of income. Feger and Radulescu (2020) study the socially optimal electricity pricing in an Atkinson-Stiglitz framework, and derive a model where (constant) externality prices deviate from the Pigouvian level. Hänsel *et al.* (2022) also note the importance of horizontal inequality and the tradeoff between efficiency and equity that may arise from distorting the carbon tax from the standard Pigouvian level. Douenne *et al.* (2023) study climate policies and inequality in a dynamic Ramsey framework, in contrast to the mechanism design model considered in this paper. Bierbrauer (2023) develops a sufficient statistics test for deviations from the uniform pricing of externalities between sectors of the economy in an equilibrium setting.

behavioral responses. The measures feature strongly in the main theorem and they will be quantified in the applications. The main result of the section is the characterization of the system of income taxes, defining also the externality payments. The decentralization leads to a set of relatively general results on externality tax levels and tax progression both for direct income and externality taxes. In Section 5, we present extensions to continuous choices, heterogeneity in emissions, tagging, and consider robustness with respect to functional forms and welfare specifications. In Section 6, the results from the two applications are summarized. The main appendix contains the proofs of the theorems and details of the empirical applications. Other proofs are in the online appendix.

2. THE SET-UP

Assumptions. We consider a unit mass of individuals heterogeneous in their ability $n \in [\underline{n}, \overline{n}] \subset \mathbb{R}_+$, cost $q \in \mathbb{R}$ of switching from dirty to clean consumption, and benefit $b \in \mathbb{R}_+$ from a marginal reduction in aggregate pollution. An individual's type $\theta = (n, q, b)$ is privately known and distributed (independently of other individuals' types) according to a cumulative distribution function F and differentiable density f with full support on $[\underline{n}, \overline{n}] \times \mathbb{R} \times \mathbb{R}_+$. Marginal distributions are denoted by F_i, conditional distributions by $F_{i|j}$, and hazard rates of marginal and conditional distributions by $h_i = f_i/(1 - F_i)$ and $h_{i|j} = f_{i|j}/(1 - F_{i|j})$ for $i, j \in \{n, q, b\}$.

Each individual makes observable (i.e., taxable) choices on how much to earn, $y \in \mathbb{R}_+$, and whether to reduce pollution, $x \in \{0, 1\}$. Utility from choice (y, x) for type $\theta = (n, q, b)$ is quasi-linear,

$$u(\theta; y, x, t, \overline{x}) = y - k(y, n) - xq + b\overline{x} - t, \tag{1}$$

in which $t \in \mathbb{R}$ is a transfer to the agent, and $\overline{x} = \mathbb{E}[x] \in [0, 1]$ is the mean pollution reduction in the population. Ability n links to income y through cost of effort function $k(y, n)$. For exposition, we assume an iso-elastic parametric class, $k(y, n) = \frac{\epsilon}{1+\epsilon}(\frac{y}{n})^{\frac{1+\epsilon}{\epsilon}}$, with elasticity $\epsilon > 0$, and show in the extensions that the main results hold under general assumptions on earnings.[11] Here y/n can be interpreted as the time an individual needs to work to earn income y.

The revelation principle allows restricting attention to incentive-compatible direct mechanisms. The mechanism $(y, x, t) : \Theta \to \mathbb{R}_+ \times \{0, 1\} \times \mathbb{R}$ assigns earnings, actions and transfers to each private type so that individuals will want to self-select the treatment designed for them:

$$\theta \in \arg\max_{\theta'} u(\theta; y(\theta'), x(\theta'), t(\theta'), \mathbb{E}[x]) \tag{2}$$

for all $\theta \in \Theta$. Here, $\mathbb{E}[x] = \mathbb{E}_{\theta \sim F}[x(\theta)]$ so the individual's payoff is affected by both the action assigned to them and the actions assigned to others, but the economy is "large" in that no single individual has a noticeable effect on the public-good provision (pollution reductions), in contrast to Clarke (1971) and Groves (1973).

[11] The extensions also cover the case in which pollution emissions depend on earnings as, for example, commuting can be an input to earnings.

Each individual has a social welfare weight, $\omega \in \mathbb{R}_{++}$, with the mean over all individuals normalized to one, $\mathbb{E}[\omega] = 1$. An individual's welfare weight is not observable but the government knows the joint distribution of ω and θ when designing (y, x, t) to maximize social welfare

$$\mathbb{E}[\omega u(\theta)] \tag{3}$$

subject to incentive compatibility in (2) and budget constraint:[12]

$$\mathbb{E}[t(\theta)] = 0. \tag{4}$$

That the government cannot observe the individual's welfare weight is important: an equity-efficiency tradeoff arises since the government has to rely on the joint distribution for (ω, θ) when attempting to identify those in need.[13] We assume that the joint density distribution for (ω, θ) is continuously differentiable and that objective (3) remains bounded for all admissible policies.

The distribution of weights allows representing society's broad preferences for allocations.

ASSUMPTION 1. *(Vertical equity)* $\mathbb{E}[\omega|n]$ *is strictly decreasing in* n.

Throughout the analysis, unless otherwise explicitly stated, we hold this assumption which captures aversion to inequality created by differences in ability. In any incentive-compatible mechanism higher ability types will select into earning higher incomes than lower types, and therefore the assumption captures aversion to income inequality in a traditional sense. But even broader preferences for redistribution can be represented by the distribution of weights:

ASSUMPTION 2. *(Horizontal equity) For given* n*, the conditional weight* $\mathbb{E}[\omega|n, q]$

- **(a)** *is strictly increasing in* q
- **(b)** *is strictly decreasing in* q

on its domain $[\underline{n}, \overline{n}] \times \mathbb{R}$.

Horizontal preferences are thus related to differences in individuals' cost of reducing emissions. Assumption 2a places a high weight on individuals who have high costs in a given ability (income) class. This could hold, for example, for emissions of commuters who cannot shift to public transportation but the same may not hold for emissions from air miles for those who have a preference to fly for leisure, and then Assumption 2b may be more appropriate. The dependence on q captures broad preferences for horizontal

[12] This can be alternatively seen as a resource constraint, because the aggregate consumption should not exceed aggregate output: $\mathbb{E}[c(n,q)] = \mathbb{E}[y(n,q)]$ where we define consumption $c(n,q) = y(n,q) - t(n,q)$.

[13] That individuals cannot observe their welfare weight is not important. It can be shown that even if individuals could observe their own welfare weights, there is an optimal mechanism which does not ask individuals to report them. This is intuitive; individual's own welfare weight does not affect private preferences. A similar result appears in, e.g., Dworczak *et al.* (2021).

equity, including the case that q correlates with some other underlying characteristics that the policy maker cares about.[14]

We focus on mechanisms in which agents only report two dimensional type (n, q). This is without loss of generality in the sense that welfare cannot be improved by reports on b because private b does not affect preferences over allocations (y, x, t) assigned to the individual, as shown by Lemma C.1 in Online Appendix C.9.

We assume that the mechanism cannot randomize the treatment.[15] Then, from the form of utility u, the Spence-Mirrlees condition holds in the y-dimension conditional on x, implying that higher types n choose higher earnings than lower types n in both clean $(x = 1)$ and dirty classes $(x = 0)$. The condition is also satisfied in the emissions choice dimension, $\partial^2 u/\partial x \partial q < 0$, which together with separated preferences with respect to earnings and emissions implies a cut-off cost $\overline{q}(n)$ for emissions reductions in each ability class n. Those with $q \leq \overline{q}(n)$ choose $x = 1$ and others choose $x = 0$.

Allocation problem. The allocation is characterized by $(y^0(n), y^1(n), \overline{q}(n))_{n\in[\underline{n},\overline{n}]}$, with earnings depending on emissions choices $x \in \{0, 1\}$ and the associated transfers $(t^0(n), t^1(n))$ satisfying the incentive-compatibility and budget constraints. Let $v(n, q) = u - b\overline{x}$ be the private utility, net of the externality, which can be expressed as a sum of "ground-level utility" and rents due to private information:

$$v(n, q) = \underline{v} + \mathcal{R}(n, q) \tag{5}$$

where $\underline{v}$ is the utility for the lowest ability type $\underline{n}$ having high cost $q > \overline{q}(\underline{n})$, who thereby receives no information rents.[16] Term $\mathcal{R}(n, q)$ encapsulates the information rents held by all other types:

$$\mathcal{R}(n, q) = \Big[-\int_{\underline{n}}^{n} k_n(y^0(s), s)ds \Big] \mathbb{1}_{q>\overline{q}(n)} + \Big[\overline{q}(\underline{n}) - q - \int_{\underline{n}}^{n} k_n(y^1(s), s)ds \Big] \mathbb{1}_{q\leq\overline{q}(n)}. \tag{6}$$

If $q > \overline{q}(n)$, the agent does not reduce emissions and earns only labor-related rents, the first integral term.[17] If $q \leq \overline{q}(n)$, a second dimension of rents emerges if the agent can reduce emissions with a cost lower than the marginal type. Intuitively, an individual having a preference for the clean alternative, such as driving an electric vehicle, does not incur a cost due to the policy. Rents earned in both dimensions n and q shape the welfare objective of the designer.

[14] One interpretation is that the weights arise from a local approximation of a welfare function (see Section 5), although an independent content for the horizontal equity concern in the welfare function representation is subject to well-known challenges (Kaplow, 1989, Musgrave, 1990, Auerbach and Hassett, 2002).

[15] The mechanism treats individuals symmetrically: the allocation is function of type (n, q), not on individual's identity, and the allocation is deterministic. The assumption is common in random-participation models; see Rochet and Stole (2003).

[16] With some abuse of the terminology, we call any such type the "lowest type", though there is a continuum.

[17] This is a well-known expression for information rents obtained from the incentive-compatibility condition for earnings. In general, the payoff from integrating the IC condition takes the form $u(\theta) = \underline{u} + \int_{\underline{\theta}}^{\theta} u_\theta(s)ds$ for some utility u, type θ, and $\underline{u} = u(\underline{\theta})$. Here, in our specific expression, the type varies only in the ability dimension n having a direct impact on utility only through the cost function $k(y, n)$.

For the welfare objective, the ground-level utility is endogenous: it is set at a level that exhausts the resources of the economy (budget) while respecting agents' incentives. We can solve the ground-level utility by taking the average over all individuals in equation (5), using the budget constraint (4) as well as the definition of $v(n,q)$:

$$\underline{v} = \mathbb{E}[y(n,q) - k(y(n,q),n) - x(n,q)q - \mathcal{R}(n,q)] \tag{7}$$

in which $y(n,q) = y^0(n)\mathbb{1}_{q>\overline{q}(n)} + y^1(n)\mathbb{1}_{q\leq\overline{q}(n)}$. Intuitively, $\underline{v}$ is the average earned income, net of labor supply costs and abatement costs, minus the average rents. Using equations (5) and (7), we can rewrite the welfare objective in (3) as:

$$\mathbb{E}[\omega u(\theta)] = \mathbb{E}\big[\underbrace{y(n,q) - k(y(n,q),n)}_{\text{output}} + \underbrace{(\mathbb{E}[\omega b] - q)\mathbb{1}_{q\leq\overline{q}(n)}}_{\text{externality}} + \underbrace{(\omega-1)\mathcal{R}(n,q)}_{\text{redistribution}}\big]. \tag{3'}$$

The optimal earnings $y^1(n)$ and $y^0(n)$, and the margin $\overline{q}(n)$ maximize welfare with three terms. The net output equals total earnings minus labor-supply costs. The externality term is the weighted benefit, net of the cost of producing the externality. The redistribution term is the weighted information rent, with mean-deviations in weights driving departures from the efficient allocation of resources. That is, if all individuals had the same weight, $\omega = 1$, the choices would not be distorted from the efficient allocation.

Decentralization. With an appeal to the taxation principle we decentralize the allocation through transfer function $t(y,x)$ decomposed into income tax $T(y)$ and income-dependent externality tax $\tau(y)$,[18]

$$t(y,x) = T(y) + (1-x)\tau(y). \tag{8}$$

To decentralize earnings $\big(y^0(n), y^1(n)\big)$, the marginal income tax should be equal to the the marginal income net of marginal costs from earning the income[19]

$$T'(y^1(n)) = 1 - k_y(y^1(n),n) \tag{9}$$

for those who reduce emissions and avoid the externality payment, and for polluters the same condition is

$$T'(y^0(n)) + \tau'(y^0(n)) = 1 - k_y(y^0(n),n). \tag{10}$$

Additionally, the tax system incentivizes emission reductions $x(\theta) = 1$ if

$$y^1(n) - k(y^1(n),n) - q - T(y^1(n)) \geq y^0(n) - k(y^0(n),n) - T(y^0(n)) - \tau(y^0(n)) \tag{11}$$

and otherwise $x(\theta) = 0$, with type $q = \overline{q}(n)$ being indifferent.

[18] The form of tax function in (8) is general enough to cover all possible allocations to be decentralized, given our assumption that allocations cannot be randomized.

[19] The expressions here assume interior solutions and differentiable of tax functions. These assumptions will hold by the properties of the allocations to be decentralized, given the parametric class for $k(y,n)$.

3. The first theorem: separated taxation

As a starter, the first theorem collapses earning schedules to one $y(n) = y^0(n) = y^1(n)$ by introducing a stand-alone income tax $T(y)$ and a constant externality tax τ. While this scheme is specific, the theorem serves to introduce important benchmarks for the optimal mechanism.[20]

THEOREM 1. *Restrict attention to a constant τ. For all n such that $y(n)$ is strictly increasing, the optimal allocation can be decentralized by tax schedule $T(y)$ satisfying*

$$\frac{T'(y)}{1-T'(y)} = \left(1+\frac{1}{\epsilon}\right)\left(1-\mathbb{E}[\omega|n' \geq n]\right)\frac{1-F_n(n)}{f_n(n)n}, \tag{12}$$

and it holds for any finite optimal $\tau > 0$ that

$$\tau = B^\omega + \frac{1-\mathbb{E}[\omega|q > \tau]}{h_q(\tau)} \tag{13}$$

in which $B^\omega = \mathbb{E}[\omega b]$ is the welfare-weighted Pigouvian price.

The income tax schedule follows Diamond's famous ABC formula (Diamond, 1998), modified here for our definition of welfare weights. Thus, $A = 1 + \frac{1}{\epsilon}$ captures elasticity-related efficiency distortion, $B = 1 - \mathbb{E}[\omega|n' \geq n]$ represents the preference for redistribution between ability (income) groups, and $C = \frac{1-F_n(n)}{f_n(n)n}$ characterizes the relative number of individuals responding to a marginal tax change. Importantly, while the constant externality tax contributes to public funds and influences the level of income taxation, it leaves income tax progression unaffected. Thus, because $T(y)$ optimally trades off the redistribution of rents and efficiency in earnings, the budgetary impacts of environmental taxation do not lead to "double dividend" by reducing income tax distortions in addition to environmental gains.[21] All impacts of externality taxation on income taxation are "budgetary", a result that will change in our main theorem in Section 4.

The formula for the externality tax τ encapsulates three main motives for redistribution in this setting.

First, $B^\omega = \mathbb{E}[\omega b]$ is the social benefit from potentially dispersed externality benefits which leads to a redistributive deviation from the Pigouvian principle that "the tax should reflect the harm". Without any preference for redistribution, the Pigouvian "price" on emissions would be $\mathbb{E}[b]$.

REMARK 1. *The welfare-weighted Pigouvian price exceeds the unweighted price, $B^\omega > \mathbb{E}[b]$, if and only if ω covaries positively with b, $Cov(\omega, b) > 0$.*

[20] In the corresponding direct mechanism the restriction means that action x only depends on q (then $\overline{q}(n) = \tau$ for all $n \in [\underline{n}, \overline{n}]$), and earnings y only depend on ability n (so $y^0(n) = y^1(n) =: y(n)$).

[21] This point is consistent with the literature; see, e.g., Jacobs and De Mooij (2015). A large literature on the double dividend hypothesis considers other margins for distortions; see e.g. Goulder (1995), Bovenberg and van der Ploeg (1994), Bovenberg (1999), Barrage (2020), Douenne *et al.* (2023).

This follows directly from definitions, $B^{\omega} = \mathbb{E}[\omega b] = \mathbb{E}[\omega]\mathbb{E}[b] + Cov(\omega, b)$. If those with high welfare weight are hurt more by the externality ($Cov(\omega, b) > 0$), then the externality tax exceeds the expected benefit, all else equal. As b is not "contractable", the policy maker cannot reach welfare-improvements by targeted transfers.[22]

Second, term $1 - \mathbb{E}[\omega|q > \tau]$ captures *differentials in cost burdens* due to τ itself. It represents the welfare effect when policies collect funds from polluters and distribute them evenly throughout the entire population. Weight $\mathbb{E}[\omega|q > \tau]$ is the mean welfare-weight of those who continue polluting. The semi-elasticity of emissions with respect to the tax, $h_q(\tau) = \frac{f_q(\tau)}{1-F_q(\tau)}$, captures the behavioral response: if emissions reductions are inelastic, h_q is low and the tax has a small impact on the externality relative to the impact on cost inequality.

PROPOSITION 1. *(Horizontal equity) Suppose that either (i) $n \perp\!\!\!\perp q$ or (ii) $\mathbb{E}[\omega|n]$ is invariant in n. Then, the externality tax deviates from welfare-weighted Pigouvian price B^{ω}:*

- **(a)** *If horizontal equity Assumption 2a holds, $\tau < B^{\omega}$.*
- **(b)** *If horizontal equity Assumption 2b holds, $\tau > B^{\omega}$.*

A policy maker who has a horizontal equity concern as captured by Assumption 2a gives a higher weight to polluters than to clean individuals, thus preferring to suppress polluters' costs by $\tau < B^{\omega}$. However, polluters may receive a lower welfare weight, and then Assumption 2b holds and the implications for the tax level are reversed.

The two premises in Proposition 1 are distinct. Independence $n \perp\!\!\!\perp q$ implies that the share of clean individuals is the same across all income levels. The externality tax cannot then screen ability and is therefore dependent only on how the policy maker weighs cost burdens. The second premise is a failure of Assumption 1, introduced here to isolate the pure horizontal equity concern. This assumption is consistent with a possibly different distribution of q across incomes thus, for instance, allowing for more extreme costs at higher incomes. Hence, the externality tax could screen ability, but if the policy maker does not care about income inequality, it is again the horizontal concern that dictates how the externality tax should be set.

Third, although the income tax schedule is the primary tool for dealing with the inequality due to ability differences, the externality tax can help to achieve the same goal when abilities correlate with costs.

PROPOSITION 2. *(Vertical equity) Suppose that $\mathbb{E}[\omega|n, q]$ is invariant in q. It follows from Assumption 1 that*

- **(a)** *If $n' > n''$ implies $F_{q|n}(q|n') \geq F_{q|n}(q|n'')$ for all $n', n'' \in [\underline{n}, \overline{n}]$, then $\tau \leq B^{\omega}$.*
- **(b)** *If $n' > n''$ implies $F_{q|n}(q|n') \leq F_{q|n}(q|n'')$ for all $n', n'' \in [\underline{n}, \overline{n}]$, then $\tau \geq B^{\omega}$.*

[22] This informational assumption is one key difference to the broader environmental economics literature which has noted the importance of welfare-weighting for the social cost of carbon (Anthoff *et al.*, 2009, Dennig *et al.*, 2015, Errickson *et al.*, 2021) and for environmental valuation (Nurmi and Ahtiainen, 2018).

The assumptions state that the policy maker has aversion to income inequality but the welfare weights are invariant in the horizontal dimension. The stochastic dominance condition indicates if the number of individuals who choose to become clean is larger at low or high incomes when facing tax τ. In the first situation, lower incomes are associated with higher emissions reduction costs which calls for lowering the tax below B^{ω} as this mitigates a transfer from the poor to rich. In the second situation, relatively more high-income earners self-select to pay the externality tax, and the transfer from the rich to poor has a positive impact on redistribution. This result echoes the finding that Atkinson-Stiglitz theorem breaks down when productivity and preferences are correlated (see e.g. Saez 2002, Jacobs and van der Ploeg 2019, Feger and Radulescu 2020, and Ferey *et al.* 2023).

4. The second theorem: optimal tax schedule

4.1 *A simple reform: a sufficient statistics approach*

We next move on to present our main results: how externality taxes and income taxes should be linked in an unrestricted mechanism. We start by developing a sufficient statistics for a simple reform that ties together income and externality taxation. To this end, consider a simple reform increasing the marginal externality tax τ of Theorem 1 for some small interval $[n, n+dn]$.[23] This makes the externality tax progressive: all incomes $y \geq y(n+dn)$ face a higher tax rate, and thus higher costs of polluting by small amount $d\tau$. The simple reform is small and, by the envelope theorem, all direct utility impacts from changes in behavior vanish, thereby leaving the potential gains to depend only on the impacts of the reform on tax revenues (that are used for redistribution) and environmental benefits. The tax revenues change through two channels. First, there is a behavioral response of earnings to a higher marginal income tax. Second, more will be collected from those who do not change behavior, that is, high-income individuals who continue polluting. The joint welfare impact of these two effects together with the third effect, the environmental benefit, can be expressed with the help of the following definitions:

$$\xi_1 = \frac{1 - F_{q|n}(\tau|n)}{1 - \mathbb{E}[F_{q|n}(\tau|n')|n' \geq n]}, \xi_2 = \frac{h_{q|n' \geq n}(\tau, n)}{h_q(\tau)}, h_{q|n' \geq n}(\tau, n) = \frac{\mathbb{E}[f_{q|n}(\tau|n')|n' \geq n]}{1 - \mathbb{E}[F_{q|n}(\tau|n')|n' \geq n]}.$$

Term ξ_1 is the relative share of polluters at n compared to the average share for $n' \geq n$. We have $\xi_1 > 1$ if high-income earners above the cut-off, $n' \geq n$, pollute less on average than incomes at the cut-off. Term ξ_2 measures the behavioral response of emissions as the mean semi-elasticity among high incomes in comparison to that in the full population, with $\xi_2 > 1$ for more response among the high-income earners. Also, to save on notation, $\overline{\omega} = 1$ is the mean weight in the full population, $\overline{\omega}^0 = \mathbb{E}[\omega | q > \tau]$ is the mean weight of all polluters, $\overline{\omega}_{n' \geq n} = \mathbb{E}[\omega | n' \geq n]$ is the mean weight of all high-income earners, and $\overline{\omega}^0_{n' \geq n} = \mathbb{E}[\omega | n' \geq n, q > \tau]$ is the mean weight of high-income polluters.

[23] Our simple reform agrees with the common definition given, e.g., in Bierbrauer *et al.* (2021), although in our case the benefits of the reform include the impact on the externality.

PROPOSITION 3. *(Sufficient statistics) For the separated-tax policy* $(T(y), \tau)$ *in Theorem 1, a simple reform marginally increasing externality tax for* $n' \geq n$ *(progressive externality tax reform) is welfare-improving if:*

$$\underbrace{(\overline{\omega}_{n'\geq n} - \overline{\omega}^0_{n'\geq n}) - (\overline{\omega} - \overline{\omega}^0)}_{(i)} + \underbrace{(\xi_2 - 1)(\overline{\omega}^0 - 1)}_{(ii)} + \underbrace{(\xi_1 - 1)(\overline{\omega}_{n'\geq n} - 1)}_{(iii)} > 0.$$

A marginal decrease in the externality tax for $n' \geq n$ *(regressive externality tax reform) is welfare-improving if the condition holds with less-than inequality (<). For a constant carbon tax to be welfare optimal, the condition must hold with equality for all* n.

The three terms comprise three distinct reasons for a progressive (regressive) reform. Term (i) captures an intuitive social preference for caring more about low-income polluters' cost burdens than those of the high-income polluters. Formally, the term is positive if weights satisfy a decreasing differences condition: the gap in weights between the polluters and the rest declines in income. This condition is satisfied, for example, if polluters are assigned higher weights than non-polluters at low income levels but lower weights than non-polluters at high income levels. This preference alone leads to optimality of a progressive reform if n and q are not correlated (i.e., if $\xi_1 = \xi_2 = 1$).

Term (ii) isolates the reason linked to the environmental gain. When $\overline{\omega}^0 > 1$, the optimal tax τ from Theorem 1 is below the welfare-weighted Pigouvian level B^{ω}, and the size of this deviation depends on the behavioral response of emissions to the tax in the full population. The reform allows the designer to look at this behavioral response separately for the sub-population of high-income earners. If this response is bigger than in the full population, $\xi_2 > 1$, a progressive reform would correct the tax upwards, towards B^w, to reduce the deviation. Analogously, if the optimal tax τ from Theorem 1 exceeds B^w, the same behavioral response, $\xi_2 > 1$, would justify a regressive reform.

Finally, term (iii) links to the earnings distortion. The optimal income tax $T(y)$ schedule in (12) trades off tax distortions at n against the redistribution benefits for the entire mass $n' \geq n$. The simple tax reform creates a similar tradeoff, but only for those who pollute. When $\xi_1 < 1$ there is a small mass of polluters at n compared to the mass for which the distributional benefits occur. As $\overline{\omega}_{n'\geq n} < 1$ by Assumption 1, there is a preference for vertical redistribution and this would favor a progressive externality tax reform. Likewise, $\xi_1 > 1$ would support a regressive reform.

4.2 *Pigouvian income tax schedule*

The main theorem characterizes the welfare-optimal tax schedule in "ABC+D" format in which D stands for the modification of ABC due to the externality. The tax schedule decentralizes the optimal allocation $\left(y^0(n), y^1(n), \overline{q}(n)\right)_{n\in[\underline{n},\overline{n}]}$ for the welfare objective given in equation (3′). The Pigouvian income tax schedule is $T(y) + \tau(y)$ in which only the polluters pay the income-dependent externality tax $\tau(y)$, defining the effective externality payment for type n as

$$\tau^n(n) = T(y^0(n)) + \tau(y^0(n)) - T(y^1(n)).$$

To simplify notation, let $\overline{\omega}_n^0 = \mathbb{E}[\omega|n, q > \overline{q}(n)]$ be the welfare weight of polluters at n, and define $\overline{\omega}_n^1$ similarly for non-polluters.

THEOREM 2. *(Pigouvian income tax schedule) A welfare-optimal allocation maximizing objective (3) exists. For all $n \in [\underline{n}, \overline{n}]$ such that y^0 and y^1 are strictly increasing, the optimal allocation can be decentralized by tax schedules $T(y)$ and $\tau(y)$ satisfying*

$$\frac{T'(y^0) + \tau'(y^0)}{1 - T'(y^0) - \tau'(y^0)} = \left(1 + \frac{1}{\epsilon}\right) \frac{\mathbb{E}[(1 - \overline{\omega}_{n'}^0)(1 - F_{q|n}(\overline{q}|n')) + (B^\omega - \tau^n) f_{q|n}(\overline{q}|n')|n' \geq n]}{n(1 - F_{q|n}(\overline{q}|n))h_n(n)} \tag{14}$$

$$\frac{T'(y^1)}{1 - T'(y^1)} = \left(1 + \frac{1}{\epsilon}\right) \frac{\mathbb{E}[(1 - \overline{\omega}_{n'}^1)F_{q|n}(\overline{q}|n') - \left(B^\omega - \tau^n\right) f_{q|n}(\overline{q}|n')|n' \geq n]}{nF_{q|n}(\overline{q}|n)h_n(n)} \tag{15}$$

in which $\tau^n = \tau^n(n')$.

The optimal tax schedules can be derived by showing that no small tax reform at any n can improve welfare, as detailed in the Appendix (the full proof using optimal-control methods is in the online appendix). Again, as for Proposition 3, three distinct effects arise: any small reform impacts tax revenues through earnings responses, taxes rich polluters and redistributes to all individuals, and generates environmental benefits. We unpack the substantial meaning of the optimal tax schedules in stages. The first result shows that, quite remarkably, Proposition 1 generalizes to the welfare-optimal Pigouvian income tax:

PROPOSITION 4. *The results from Propositions 1-2 for the constant externality tax extend to the optimal tax system given by formulas (14) and (15) in Theorem 2: the propositions continue to hold with τ replaced by $\tau(y)$ for all y.*

The externality tax schedule lies below or above the welfare-weighted Pigouvian price B^ω, depending only on the primitive vertical and horizontal equity assumptions in 1 and 2 and the joint distribution of n and cost q. When ability n and cost q are independent, a general horizontal equity concern of Assumption 2 dictates if $\tau = \tau(y)$ is below or above B^ω, as in Proposition 1. The same conclusion holds if there is no taste for vertical income redistribution (failure of Assumption 1). In contrast, the second determinant of the deviation from B^ω is related purely to the vertical redistribution motive, as in Proposition 2. Deviations from B^ω allow screening individuals' ability types through their consumption behaviour for better outcomes for vertical equity. The externality tax is distorted to fall below B^ω at *every* ability level when lower incomes are associated with higher costs of reductions in a stochastic dominance sense, and conversely for the reversed association of costs and abilities.

The second result states that if the optimal externality tax is progressive (regressive), then the base income tax is less (more) progressive than the stand-alone income tax schedule:

PROPOSITION 5. *Let T and τ denote the optimal income and externality tax given by formulas (14) and (15) in Theorem 2. Let $\widehat{T}$ denote the optimal stand-alone income tax schedule in Theorem 1. Then:*

(a) *For any $n \in [\underline{n}, \overline{n}]$, if $\tau'(y^0(n)) \geq 0$ then $T'(y^1(n)) \leq \widehat{T}'(y(n))$.*

(b) *For any $n \in [\underline{n}, \overline{n}]$, if $\tau'(y^0(n)) \leq 0$ then $T'(y^1(n)) \geq \widehat{T}'(y(n))$.*

Conditioning the externality tax on income allows utilizing the correlation of cost q between n and ω for gains in both equity and efficiency. Proposition 5 highlights that a progressive externality tax ($\tau'(y) > 0$) is counter-balanced by decreased progressivity of the base income tax, and, conversely, a similar counter-balance in the opposite direction occurs in the case of a regressive externality tax ($\tau'(y) < 0$). The Pigouvian income tax strives to respect the vertical equity preference, which is the primitive preference behind the stand-alone income tax schedule $\widehat{T}(y)$. We can see this precisely by merging the income tax schedules from Theorems 1 and 2, giving

$$(1 - F_{q|n}(\bar{q}(n)|n))\frac{T'(y^0(n)) + \tau'(y^0(n))}{1 - T'(y^0(n)) - \tau'(y^0(n))} + F_{q|n}(\bar{q}(n)|n))\frac{T'(y^1(n))}{1 - T'(y^1(n))} = \frac{\widehat{T}'(y(n))}{1 - \widehat{T}'(y(n))}.$$

The new income tax schedules are thus equal to the standalone schedule on average.

What then determines if the externality tax schedule should be progressive or regressive? The sufficient statistics result in Proposition 3 provided hints of the determinants, but a similar result for the welfare-optimal income tax schedule requires two additional assumptions. The first assumption makes use of the following definition:

$$\eta(n,q) = \frac{f_{q|n}(q|n)}{F_{q|n}(q|n)(1 - F_{q|n}(q|n))}.$$

Intuitively, η measures the change of the pools of polluters and non-polluters due to a marginal increase in q; it is the change in the "odds ratio" familiar from logistic regressions.[24]

ASSUMPTION 3. *$(B^\omega - q)\eta(n,q)$ is non-increasing in q for all $n \in [\underline{n}, \overline{n}]$, $q \in \mathbb{R}_+$.*

The assumption is satisfied, for example, for all $B^\omega > 0$ when q follows an exponential or a logistic distribution.[25] More generally, the assumption is satisfied by any cost distribution where the derivative of $\eta(n,q)$ with respect to q is close enough to zero.

[24] Drop conditioning on n for the sake of illustration and consider a logistic regression model with a constant tax τ predicting the behavioral response of emissions reductions $F_q(\tau) = (1 + e^{-(\beta_0 + \beta_1 \tau)})^{-1}$. We would obtain by differentiating that it holds for the estimated parameters that $\beta_1 = \eta(\tau)$.

[25] For exponential distribution, $F_{q|n}(q|n) = 1 - e^{-\lambda(n)q}$, the derivative of $(B^\omega - q)\eta(n,q)$ w.r.t. q is always negative as $e^{\lambda(n)q} > 1 + \lambda(n)q$ for all $\lambda(n), q > 0$. For logistic distribution, $F_{q|n}(q|n) = (1 + e^{-\frac{q-\mu(n)}{s(n)}})^{-1}$ we have $\eta(n,q) = 1/s(n)$ and therefore the derivative of $(B^\omega - q)\eta(n,q)$ w.r.t. q is -1/$s(n)$. Notice that an exponential distribution of q is supported on a semi-infinite interval while according to our baseline assumptions, the distribution of q is supported on the whole real line.

The next assumption is on the joint distribution of welfare weight ω and cost q. As before, weights ϖ_n^0 and ϖ_n^1 are the mean weights of polluters and non-polluters, respectively, at given n, but to make the cut-off defining the two groups explicit, we write $\varpi_n^0(\tilde{q}) = \mathbb{E}[\omega|n, q' > \tilde{q}]$, and similarly for $\varpi_n^1(\tilde{q}) = \mathbb{E}[\omega|n, q' \leq \tilde{q}]$ for some arbitrary cut-off $\tilde{q}$.

ASSUMPTION 4. $\varpi_n^0(\tilde{q}) - \varpi_n^1(\tilde{q})$ *is non-decreasing in* $\tilde{q}$ *for all* $n \in [\underline{n}, \overline{n}]$, $\tilde{q} \in \mathbb{R}_+$.

Slightly simplifying, the assumption states that at a given ability level n, increasing the externality payment should not *decrease* the difference between the mean weight of polluters and that of non-polluters. That is, raising the externality tax increases the concern for those who pay the tax relative to those who do not.

From the sufficient statistics result, Proposition 3, we know that the preference for caring more about the cost burdens of low-income polluters than those of the high-income polluters is enough for a progressive reform without any "behavioral" considerations. The next proposition captures the same idea for the Pigouvian income tax schedule.

PROPOSITION 6. *Consider an optimal tax system given by formulas (14) and (15) in Theorem 2. Suppose Assumptions 3-4 hold and* $n \perp\!\!\!\perp q$*. Then*

(a) *If* $\varpi_n^0(\tilde{q}) - \varpi_n^1(\tilde{q})$ *is strictly decreasing in* n *for all* $n \in [\underline{n}, \overline{n}]$, $\tilde{q} \in \mathbb{R}_+$, *then* $\tau'(y) \geq 0$ *for all* y.

(b) *If* $\varpi_n^0(\tilde{q}) - \varpi_n^1(\tilde{q})$ *is strictly increasing in* n *for all* $n \in [\underline{n}, \overline{n}]$, $\tilde{q} \in \mathbb{R}_+$, *then* $\tau'(y) \leq 0$ *for all* y.

Independence $n \perp\!\!\!\perp q$ rules out the behavioral considerations, that is, progressivity is not due to pollution responding differently to taxation across incomes. For a positive gap $\varpi_n^0(\tilde{q}) - \varpi_n^1(\tilde{q}) > 0$, the effect captured by condition (a) in Proposition 6 is about seeing the pools of polluters and non-polluters less and less different when incomes increase, which calls for a higher, less distorted tax for the rich. The same logic applies when the externality tax is regressive but now "less distortion for the rich" means a lower tax for them.[26]

The next proposition isolates efficiency reasons for progressive externality taxation due to behavior. To do this, it rules out the variation that is key in Proposition 6, that is, the variation in horizontal equity motive across income levels.

PROPOSITION 7. *Consider an optimal tax system given by formulas (14) and (15) in Theorem 2. Suppose Assumptions 3-4 hold and that* $\varpi_n^0(\tilde{q}) - \varpi_n^1(\tilde{q})$ *is invariant in* n *for all* $n \in [\underline{n}, \overline{n}]$, $\tilde{q} \in \mathbb{R}_+$. *Consider:*

(a) *horizontal equity Assumption 2a holds*

(a') *horizontal equity Assumption 2b holds*

[26]A supermodularity condition of type "if $q' > q$, then $\mathbb{E}[\omega|n, q'] - \mathbb{E}[\omega|n, q]$ decreases in n" is sufficient (but not necessary) for the condition to be satisfied in the proposition when n and q are independent.

(b) *$\eta(n,q)$ is strictly increasing in n* ***(b')*** *$\eta(n,q)$ is strictly decreasing in n.*

If (a)-(b) or (a')-(b'), then it holds for the optimal tax that $\tau'(y) > 0$ in some interval. If (a')-(b) or (a)-(b'), then $\tau'(y) < 0$ in some interval.

For the intuition, consider that the pollution of the rich is highly elastic with respect to the externality tax, a situation of (b).[27] Then, provided there is a general horizontal equity concern so that the welfare weight is increasing in q, situation (a), the tax should be distorted less from B^{ω} for the rich, leading to a progressive externality tax. This intuition echoes the one given for the sufficient statistics result, but here the result is weaker, showing only some progressivity.

5. EXTENSIONS AND ROBUSTNESS

In this section, we briefly present extensions and a robustness analysis of the main model, which is formally detailed in the online appendix.

Continuous choice of emissions. In some applications, pollution reduction is an intensive-margin rather than extensive-margin choice. We can define a continuous variable $x \in [0,1]$ for the proportion of emissions abated. Accordingly, $e = 1 - x$ represents is the level of pollution generated by the individual. The results of Theorem 1 can be generalized to this case, with one important modification: the welfare weight of polluters $\mathbb{E}[\omega|q > \tau]$ is replaced by the emissions-weighted welfare-weight of all individuals, $\frac{\mathbb{E}[\omega e]}{\mathbb{E}[e]}$. We can also extend the results of the simple reform in Section 4.1 to this case. The condition for the optimal progressive (regressive) reform involves the same three effects, with the following adjustments in definitions. First, welfare weights are weighted by individual's emissions. Second, the relative shares of polluters in ξ_1 are replaced by relative shares of emissions in the respective parts of the population. Third, semi-elasticities of average emissions replace the semi-elasticities of polluters' masses in ξ_2. In Section 6, the application to electricity consumption uses the continuous choice version of the sufficient statistics formula.

Heterogeneous emissions. We can extend the model to cover the case of heterogeneous emissions among individuals if we assume that each agent chooses $x \in \{0,1\}$ but pollution emissions differ by income and earning ability according to function $e(y,n)$. This establishes a direct link between pollution externalities and incomes, capturing factors such as commuting as an input to earning income and other determinants that cause varying consumption patterns for a given income. We characterize the optimal separated tax system, defined as $T(\cdot) = T(y) + \tau e(\cdot)$. A marginal increase of $d\tau$ in the externality tax results again in three effects (behavioral, redistributive, environmental). In

[27] If q is logistically distributed on each ability level, that is, $F_{q|n}(q|n) = (1 + e^{-(\beta_0(n)+\beta_1(n)q)})^{-1}$, then the derivative of $\eta(n,q)$ w.r.t. n is $\beta_1'(n)$. If q is exponentially distributed on each ability level, that is, $F_{q|n}(q|n) = 1 - e^{-\lambda(n)q}$, then the sign of the derivative of $\eta(n,q)$ w.r.t. n is equal to the sign of $\lambda'(n)$. Hence, in this case, the assumption about increasingness (decreasingness) of $\eta(n,q)$ in n is equivalent to the cost distributions of the lower-income groups hazard-rate dominating the cost distributions of higher-income groups.

particular, changing τ affects income not only through labor supply response but also because of the link between earnings and emissions. The formula in the online appendix shows that the externality tax is pushed downwards from the welfare-weighted Pigouvian level B^{ω} if the polluters' *emissions-weighted* welfare weight exceeds the average welfare weight in the population, and more so if the externality action is not very responsive and the individuals at the margin of making the externality action do not generate much emissions. Furthermore, if $e_y > 0$, then increasing the externality tax distorts labor market incentives.

Alternative social welfare specifications. Exogenous welfare weights can be linked to a utilitarian social welfare function by considering objective

$$\mathbb{E}[\Omega(u(\theta))] \tag{1'}$$

in which Ω is an increasing and concave function of agent's utility (1). For a given policy, the average social marginal welfare weight of individuals with ability n and cost q is then $\mathbb{E}\big[\omega|n,q\big] = \mathbb{E}\big[\Omega'(u(\theta))/\lambda\big|n,q\big]$ where λ is the shadow price of the budget constraint.[28] To analyze the connection between the social preferences represented by Ω and our assumptions on weights ω, suppose now for simplicity that b is invariant across individuals and that we generate the weight for each individual from Ω. Under the classical assumption that the policy maker has aversion to utility differences ($\Omega'' < 0$), the weights generated by Ω would satisfy the vertical equality concern of Assumption 1 and the horizontal equality concern of Assumption 2a: polluters would always receive a higher weight than non-polluters under these assumptions. The online appendix demonstrates that Assumption 4 can be reconciled with welfare as represented by Ω, showing that relatively broad social objectives can be consistently presented by welfare weights; however, see also Sher (2023).

General cost-of-earnings function. The iso-elastic class for the cost of earning income, $k(y,n) = \frac{\epsilon}{1+\epsilon}(\frac{y}{n})^{\frac{1+\epsilon}{\epsilon}}$, is useful for exposition. It satisfies $k_{yn}(y,n) < 0$ which is the requirement that must be placed on any general $k(y,n)$ to guarantee that earnings schedules $y^0(n)$ and $y^1(n)$ are non-decreasing in n. To illustrate the earnings choice with a general $k(y,n)$, consider separated income and externality choices, as in Theorem 1. For all n such that income $y(n)$ is strictly increasing in n, the optimal mechanism satisfies

$$1 - k_y(y(n),n) + \frac{1-F_n(n)}{f_n(n)}\big(1 - \mathbb{E}[\omega|n \geq n]\big)k_{yn}(y(n),n) = 0. \tag{16}$$

The iso-elastic form gives $k_{yn} = -\big(1+\frac{1}{\epsilon}\big)\frac{1}{n}k_y$ which, when applying $T'(y) = 1 - k_y$, leads to the formula in Theorem 1. While the same simplicity is lost with general $k(y,n)$, there is no material impact on Theorem 1 and the results of Proposition 3 hold under the same conditions. In Theorem 2, the changes can be more substantial because the progression of income taxes, and thus the incentives to work, are linked to emissions choices. Consequently, the elasticity of earnings and emissions can be interconnected.

[28] In our main model with welfare weights, the normalization $\mathbb{E}[\omega] = 1$ implies $\lambda = 1$.

Tagging. The literature has proposed a taxation based on external, easily observable factors like height and age, although this approach raises concerns as it goes against the principle of treating individuals equally when it comes to attributes they cannot control (Akerlof, 1978, Cremer *et al.*, 2010). Such tags are rarely applied in real-life income tax regimes, but "quasi-tags" are used in relation to climate policies: for example, the Canadian Climate Action rebates revenue by family size and the place of residence.[29]

Following earlier literature, consider a model in which each individual belonging to one of two groups, $i \in \{A, B\}$. The policy problem can then be split into two parts: first, we may derive the optimal tax system for each group for a given group-specific budget constraint, and then we can optimize the group-specific budgets so that the aggregate budget constraint is met. If the groups differ in their expected welfare weights, it is welfare-improving to transfer money from the group with a lower welfare weight to the other group. Transferring money from group A to group B improves welfare if $\mathbb{E}^A[\omega] > \mathbb{E}^B[\omega]$. This condition holds, for instance, if the welfare weight is a function of ability n and cost q only, the distribution of ability n is the same in the two groups, Assumption 2a for horizontal equity holds, and the distribution of costs conditional on n in group A first-order stochastically dominates the distribution in group B. If, instead, Assumption 2b holds, the direction of the transfer changes. Even if the groups do not differ in their expected welfare weights, the welfare-maximizing tax system may treat the two groups differently. In a such a case, the optimal separated externality tax in group $i = A, B$ is set at

$$\tau^i = B^\omega + \frac{\mathbb{E}^i[\omega] - \mathbb{E}^i[\omega|q > \tau]}{h_q^i(\tau)},$$

intuitively depending on the concern for polluters relative to the behavioral response.

Public investments. Both private and public choices can reduce externalities. In this case, the sum of actions is $\overline{x} = \mathbb{E}[x(\theta)] + z$, in which z is the public investment with cost $c(z)$ to be funded with taxes. This leads to the adjusted budget constraint

$$\mathbb{E}[t(\theta)] = c(z), \tag{4'}$$

and the optimal public investment is characterized by

$$c'(z) = \mathbb{E}[\omega b].$$

This implies that the Samuelson rule can be modified to include welfare-weighted benefits, which is the only aspect where inequality considerations lead to deviations from the standard Samuelson rule.[30] In contrast, the implementation of policies on private choices are strongly influenced by inequality considerations, as previously discussed.

[29]Family size and the place of residence are *quasi*-tags in the sense that they have been shown to be responsive to financial incentives (Milligan, 2005, Kennan and Walker, 2011).

[30]For an extensive survey on the Samuelson rule literature, see Kreiner and Verdelin (2012).

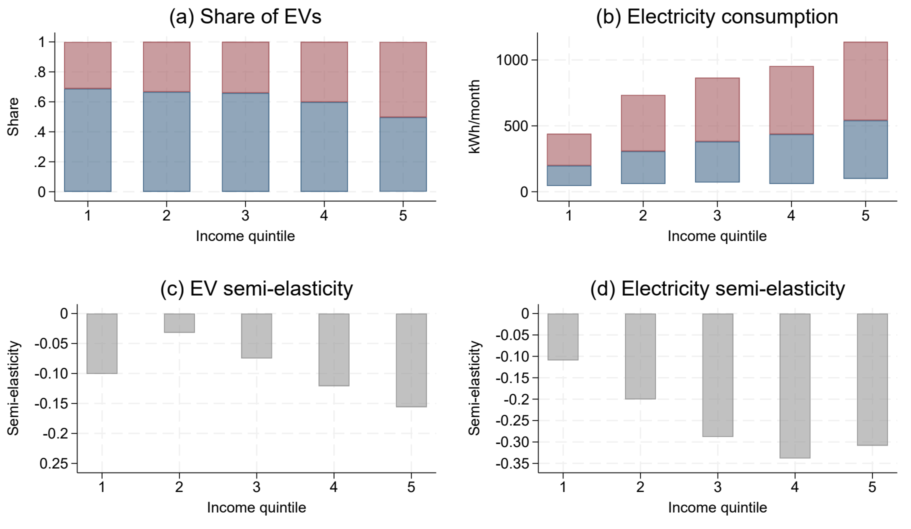

FIGURE 1. Finnish registry data on vehicles and electricity.

Note: All data by households' disposable income quintile. Figure shows (a) the share of electric vehicles (EVs) for new transactions in q1-q3 of 2023 ($N = 44,360$ transactions), (b) electricity consumption for October 2022 - January 2023 for Finnish households ($N = 2,530,334$ households), (c) the semi-elasticity of the number of non-EVs by households, taken from the literature (Halse *et al.*, 2024), (d) the semi-elasticity of electricity consumption by income group estimated based on the Finnish registry data ($N = 2,530,334$ households).

6. TWO APPLICATIONS: VEHICLE CHOICE AND ELECTRICITY CONSUMPTION

This section demonstrates how our theory can be linked to data. We focus on externalities created by carbon emissions in two key sectors, transportation and electricity. The two applications use Finnish individual-level administrative data on incomes, pollution measures, and financial burdens to test income-dependence of externality pricing using Proposition 3. The sufficient statistics in the proposition requires: (i) social welfare weights of polluters and non-polluters, (ii) the semi-elasticity of emission reduction by income group, and (iii) emissions by income groups. Importantly, to analyze a small reform, we do not need the elasticity of earnings or the marginal externality damages assuming that the schedule to be reformed is the optimal separated schedule (see Proposition 3).

The data, divided into five quintiles based on households' disposable income, is summarized in Figure 1. Further details on the quantification are in Appendix B.

Income, consumption, and welfare weights. The unit of observation is a household, i. We observe disposable income y_i for each household. The data also includes the kilometers driven and electricity consumption, which allows us to calculate the expenditure in electricity or gasoline, Δ_i. The net-of-cost income for each household can be calculated as $\mathcal{I}_i = y_i - \Delta_i$. Every agent has n and q which lead to income and consumption decisions, and therefore we denote $\mathcal{I}_i$ by $\mathcal{I}(n, q)$.

We propose a simple and transparent approach to generating an illustrative population-level distribution for welfare weights, based on a constant relative risk aversion (CRRA) welfare function with parameter σ measuring the degree of inequality aversion:

$$\mathbb{E}[\omega|n,q] = \frac{\mathcal{I}(n,q)^{-\sigma}}{\mathbb{E}[\mathcal{I}(n,q)^{-\sigma}]}. \tag{17}$$

The denominator ensures that the normalization $\mathbb{E}[\omega] = 1$ holds. We thus generate the weights at observations prior to potential changes in incomes and burdens, and assume that changes are so small that weights can be taken as exogenous. Equation (17) is a modified version of the formula proposed in draft for U.S. circular A-4 which advocates calculating distributional weights using CRRA-utility function with $\sigma = 1.4$ (White house, 2023). This calculation of weights represents value judgments (of circular A-4), but lessons that we draw from the applications are not sensitive to the precise quantification of inequality aversion.[31]

Application I: vehicle choice. We use data for all new car transactions in Finland for calendar year 2023, quarters q1-q3. Households in the market for new cars make a discrete choice between electric vehicles (EVs) and vehicles with an internal combustion engine, including plug-in hybrids (non-EVs). EVs receive different tax treatments than non-EVs, most importantly they avoid pollution from gasoline and diesel fuels and the related expenditures.[32] The share of EVs by income quintile is shown in Figure 1a. As has been found in other settings (e.g., Bigler and Radulescu, 2022, Halse *et al.*, 2024), the top earners tend to buy most EVs. We get the semi-elasticity of pollution reduction by income quantile from an external source; this hazard rate is from Halse *et al.* (2024). Figure 1c shows that higher income groups have a higher semi-elasticity.

From the population-vehicle data, we produce the distributions of weights separately for polluters and non-polluters, and the three terms entering the sufficient statistics formula (Proposition 3). The results are in Table 1. The mean weight in the population is, by definition, equal to 1.[33] From the first row, we observe a decline in mean weight as we move towards higher income quintiles. In the second row, the average weight of polluters is *higher* than the overall population mean, which follows from the fact that individuals with lower incomes are more likely to purchase more polluting vehicles. The higher mean weight of polluters readily gives our first result from Theorem 1 on setting the optimal tax level: the fuel tax should be lower than in the absence of distributional concerns to reduce a transfer from low-income to high-income individuals.

We further observe that the difference in mean weights declines when conditioning on a higher income quantile, turning the "decreasing-differences" term (i) in Table

[31] Alternative strategies to calculate welfare weights would be to elicit social preferences held by individuals by surveys (Saez and Stantcheva, 2016) or experiments (Capozza and Srinivasan, 2023), or use the "inverse optimal-tax method" in which weights are calculated from existing or proposed policies (Jacobs *et al.*, 2017, Hendren, 2020).

[32] In our model, τ can capture both different tax treatment of new car purchases (car taxes or EV subsidies) and difference treatment of use costs. Finland has a CO_2-based car taxation (see Stitzing, 2016) and a CO_2 component in its gasoline taxation of 62 euros/tCO_2 (government proposal HE 36/2023). The effective CO_2 price for all gasoline taxes is roughly 250 euros/tCO_2 (OECD, 2021).

[33] The population in this illustration is the universe of new car buyers in 2023.

TABLE 1. Simple tax reform for vehicles

		Income quintile				
		All	2	3	4	5
Average weight	$\overline{\omega}_{n' \geq n}$	1	0.876	0.791	0.695	0.583
Average polluter weight	$\overline{\omega}^0_{n' \geq n}$	1.077	0.943	0.848	0.746	0.632
Quantity at $n' \geq n$	$1 - \mathbb{E}[F_{q\|n}(\tau\|n')\|n' \geq n)]$	0.689	0.687	0.681	0.667	0.652
Quantity at $n' = n$	$1 - F_{q\|n}(\tau\|n)$	0.741	0.745	0.743	0.695	0.652
Semi-elasticity at $n' > n$	$h_{q\|n' \geq n}(\tau, n)$	0.010	0.010	0.012	0.015	0.017
Term (i)	$(\overline{\omega}_{n' \geq n} - \overline{\omega}^0_{n' \geq n}) - (\overline{\omega} - \overline{\omega}^0)$	0	0.010	0.020	0.027	0.028
Term (ii)	$(\xi_2 - 1)(\overline{\omega}^0 - 1)$	0	-0.001	0.0016	0.033	0.047
Term (iii)	$(\xi_1 - 1)(\overline{\omega}_{n' \geq n} - 1)$	0	-0.011	-0.019	-0.013	0
Terms (i)+(ii)+(iii)	Net effect	0	-0.001	0.017	0.048	0.750

Note: The first row: mean welfare weight for all car buyers from any income quintile $i = 1, .., 5$ and larger than i. The second row: mean weight for all non-EV buyers from income quintile i and larger. The third row: share of non-EV buyers from income quintile i and larger. The fourth row: share of non-EV buyers from income quintile i. The fifth row: semi-elasticity is the hazard rate for emissions in group i and above. Terms (i)-(iii) refer to the same terms presented in Proposition 3. In term (ii) we use $\xi_2 = h_{q|n' \geq n}(\tau, n)/h_q(\tau)$. In term (iii) we use $\xi_1 = (1 - F_{q|n}(\tau|n))/(1 - \mathbb{E}[F_{q|n}(\tau|n')|n' \geq n])$.

1 positive. This term intuitively captures the idea that as incomes increase, the policymaker is less concerned about the cost burdens to polluters, which has implications for the optimality of a constant tax. Positive term (i) supports the optimality of a progressive reform due to redistribution.

Term (ii) illustrates the role of behavioral elasticities in the sufficient statistics test. We know from the deduced welfare weights that the tax is set below the efficient level due to redistribution considerations. The semi-elasticity of pollution reductions, while small, is increasing in income. Combined, the two observations mean that the efficiency loss from the distorted tax is higher at higher incomes, supporting again a progressive reform. Term (ii) is bigger at higher income quintiles, so its effect is stronger at higher quintiles.

The sign of term (iii) is also intuitive. Term $\overline{\omega}_{n' \geq n} - 1$ is non-positive because the weight does not separate polluters and the rest but conditions only on income. The behavioral earnings response, ξ_1, is captured by the relative sizes of polluter populations of the reform quintile and those above the quintile, obtained by dividing the number in line four by that in three. When the relative size of the population is large in this sense at incomes for which the reform is considered, the reform distorts earnings more than what is the redistributive gain. In fact, term (iii) would support a regressive reform for this reason. Taken together, the three terms support implementing a progressive reform, with the size of the joint effect increasing in the income quintile.

Application II: electricity consumption. We follow the same procedure for quantifying a tax policy on electricity consumption.[34] The data contains household's contract type

[34] As part of the EU Emissions Trading System, emissions from electricity generation face a market price, with a pass-through rate to the electricity wholesale price close to one; see Fabra and Reguant (2014) and Liski and Vehviläinen (2020).

TABLE 2. Simple tax reform for electricity

		Income quintile				
		All	2	3	4	5
Average weight	$\overline{\omega}_{n'\geq n}$	1	0.679	0.505	0.397	0.294
Average polluter weight	$\overline{\omega}^0_{n'\geq n}$	0.652	0.497	0.389	0.316	0.243
Quantity at $n' \geq n$ (kWh)	$1-\mathbb{E}[F_{q\|n}(\tau\|n')\|n'\geq n)]$	374	416	454	490	542
Quantity at $n' = n$ (kWh)	$1-F_{q\|n}(\tau\|n)$	208	310	382	437	542
Semi-elasticity at $n' > n$	$\bar{h}_{q\|n'\geq n}(\tau, n)$	0.249	0.284	0.312	0.324	0.309
Term (i)	$(\overline{\omega}_{n'\geq n}-\overline{\omega}^0_{n'\geq n})-(\overline{\omega}-\overline{\omega}^0)$	0	-0.166	-0.232	-0.267	-0.297
Term (ii)	$(\xi_2-1)(\overline{\omega}^0-1)$	0	-0.049	-0.088	-0.104	-0.083
Term (iii)	$(\xi_1-1)(\overline{\omega}_{n'\geq n}-1)$	0	0.082	0.079	0.065	0
Terms (i)+(ii)+(iii)	Net effect	0	-0.133	-0.241	-0.306	-0.380

Note: The first row: mean welfare weight for all households from any income quintile $i = 1, .., 5$ and larger than i. The second row: consumption-weighted welfare weight for households from income quintile i or above. The third row: the amount of consumption (in kWh, September 2022) of households from income quintile i or above. The fourth row: consumption (in kWh, September 2022) but for income quintile i. Semi-elasticity is the percentage change of demand. The value presented is the mean semi-elasticity for individuals in group i or above derived from Appendix Table B.3. In Term (ii) we use $\xi_2 = \bar{h}_{q|n'\geq n}(\tau, n)/\bar{h}_q(\tau)$ where semi-elasticity is defined as $\bar{h}_q$ and $\bar{h}_{q|n'\geq n}(\tau, n) \equiv -\mathbb{E}\left[\frac{\partial e}{\partial \tau}|n' \geq n\right]/\mathbb{E}[e|n' \geq n]$. In Term (iii) we use $\xi_1 = \mathbb{E}[e|n]/\mathbb{E}[e|n' \geq n]$.

which, based on public contract prices, allows us to determine for each household a monthly electricity expenditure. Figure 1b shows that high-income households consume more electricity. We estimate the semi-elasticity of electricity consumption by income group. Our empirical design exploits variation in household-level exposure to the unanticipated price shock in 2022, resulting from Russia limiting its energy supplies to Europe. We identify semi-elasticities by a difference-in-differences approach, exploiting variation in electricity contract durations.[35] The estimated semi-elasticities are shown in Figure 1.

The results are in Table 2. In a sharp contrast to the vehicle case, polluters now receive a considerably lower mean weight than the non-polluters in the full population.[36] On reflection, electricity consumption correlates strongly with income, implying a negative association between pollution and welfare weights. This observation implies that the tax level should be elevated for a transfer from the wealthier to the poorer (Theorem 1).

Unlike for vehicles, term (i) is now negative and declining in income. This supports a regressive reform that sets a lower tax to individuals with incomes above the reform cutoff, with a logic coming from Proposition 3: the corrective tax should be distorted less when the gap between mean weights of polluters and non-polluters declines in income. This is precisely what happens when incomes increase in Table 2.

[35] The mean contract price for fixed-term contracts lasting throughout the crises was 0.065 Eur/kWh. Contracts that ended in November 2022 faced a higher price in December, on average 0.38 Eur/kWh. The contracts that expired were signed one year or two years before the crises.

[36] Welfare weights are from equation (17) but household-level weights are multiplied by emissions, following Section 5, to account for a continuous consumption choice. Emissions are assumed to be proportional to consumption.

The estimated semi-elasticity of emissions follows the same pattern as for the vehicles, that is, the estimate increases in income, except for the last quintile. This time term (ii) is negative, due to low relative weights given to polluters, and therefore it supports a regressive reform. The reasoning is the same as for vehicles, except now the tax is elevated above the efficient level thereby calling for less elevation at high incomes as the efficiency distortion increases in income.

Last, term (iii) goes against terms (i)-(ii) and supports a progressive reform. The relative quantity of pollution is small at lower incomes supporting a progressive tax for income redistribution. However, the effect is small and the net effect is negative, meaning that the sufficient statistics test is in favor a regressive reform.

7. Concluding remarks

Corrective policies, such as the pricing of climate externalities, are increasingly conditioned on income. What are the general principles guiding the design of such income-dependent policies? Policy makers often have intentions that are good, but the outcomes are not, neither in terms of efficiency nor on distributional grounds (e.g., Borenstein and Davis 2016). Economists have not offered clear principles to help policy makers. To fill this gap, we set out to develop a theory framework that brings together economists' call for efficiency and, on the other hand, policy makers' broad distributional concerns. The approach treats income inequality and issues like energy poverty as distinct factors influencing the policy design. We further assume that policy makers do not have individual-level information about who needs support; there is only general data on how needs are associated with characteristics in the population.

The main results imply that proposals to gain support for climate polices by returning carbon tax revenues back to citizens as a uniform per capita sum are ineffective in solving the critical efficiency-equity tradeoffs.[37] Such rebates and associated carbon taxes should not be isolated from incentives to work and from transfers linked to income taxation. The problem calls for a comprehensive tax reform leading to joint determination of marginal tax rates for earnings and emissions, both depending on individual's income levels.

Entities beyond governments, such as corporations, may also face the challenge of "fair pricing". Corporate policies may impose restrictions on the extent to which companies can leverage consumer data for price discrimination, aiming to ensure equal treatment of customers or to prioritize the welfare of specific consumer groups. The standard model of price discrimination does not weigh consumer surplus (Rochet and Stole 2003, Dubé and Misra 2023), which however seems necessary if broader objectives for pricing are included. The methodology proposed in this paper could offer insights for incorporating these broader objectives into pricing strategies.

[37] In "Economists' Statement on Carbon Dividends", more than 3,600 economists, among them 28 Nobel Laureates, propose that equity concerns should be addressed equal lump-sum rebates of the tax revenues to citizens. The statement is originally published in The Wall Street Journal, 2019.

REFERENCES

AHLVIK, L., ALHOLA, S., BLAUBERG, T., LAASONEN, H., LISKI, M., MALINEN, O., MATTILA, M., NOKSO-KOIVISTO, O., SAHARI, A., SEPPÄ, M. *et al.* (2023). Sähkön hinnannousun ja sähkötukien arvioidut vaikutukset kotitalouksiin. [33]

— and LISKI, M. (2022). Global Externalities, Local Policies, and Firm Selection. *Journal of the European Economic Association*, **20** (3), 1231–1275. [5]

AHONEN, A. (2023). *Tagging for Votes: Can Targeted Transfers Make Carbon Pricing More Acceptable?* Working Papers 2023, Aalto University. [32]

AKBARPOUR, M., BUDISH, E., DWORCZAK, P. and KOMINERS, S. D. (2024a). An economic framework for vaccine prioritization. *The Quarterly Journal of Economics*, **139** (1), 359–417. [4]

—, DWORCZAK, P. and KOMINERS, S. D. (2024b). Redistributive allocation mechanisms. *Journal of Political Economy, accepted.* [4]

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. [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. [5]

ANTHOFF, D., HEPBURN, C. and TOL, R. S. (2009). Equity weighting and the marginal damage costs of climate change. *Ecological Economics*, **68** (3), 836–849. [11]

AUERBACH, A. J. and HASSETT, K. A. (2002). A new measure of horizontal equity. *The American Economic Review*, **92** (4), 1116–1125. [8, 19]

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. [10]

BIERBRAUER, F. (2023). *Is a market-based approach to climate policy desirable?* Working Paper 2023, University of Cologne. [5]

BIERBRAUER, F. J., BOYER, P. C. and PEICHL, A. (2021). Politically feasible reforms of non-linear tax systems. *American Economic Review*, **111** (1), 153–191. [12]

BIGLER, P. and RADULESCU, D. (2022). Environmental, redistributive and revenue effects of policies promoting fuel efficient and electric vehicles. *CESifo Working Paper.* [21, 33]

BORENSTEIN, S. and DAVIS, L. W. (2016). The distributional effects of us clean energy tax credits. *Tax Policy and the Economy*, **30** (1), 191–234. [24]

BOVENBERG, A. and VAN DER PLOEG, F. (1994). Environmental policy, public finance and the labour market in a second-best world. *Journal of Public Economics*, **55** (3), 349–390. [10]

BOVENBERG, A. L. (1999). Green tax reforms and the double dividend: an updated reader's guide. *International tax and public finance*, **6**, 421–443. [10]

CALIFORNIA PUBLIC UTILITIES COMMISSION (2024a). https://www.cpuc.ca.gov. Accessed: April 7, 2024. [3]

CALIFORNIA PUBLIC UTILITIES COMMISSION (2024b). https://www.cpuc.ca.gov/consumer-support/financial-assistance-savings-and-discounts/california-alternate-rates-for-energy. Accessed: April 7, 2024. [3]

CAPOZZA, F. and SRINIVASAN, K. (2023). Who should get money? estimating welfare weights in the us. [21]

CLARKE, E. H. (1971). Multipart pricing of public goods. *Public Choice*, **11**, 17–33. [6]

CREMER, H., GAHVARI, F. and LADOUX, N. (1998). Externalities and optimal taxation. *Journal of Public Economics*, **70** (3), 343–364. [5]

—, — and — (2003). Environmental taxes with heterogeneous consumers: an application to energy consumption in france. *Journal of Public Economics*, **87** (12), 2791–2815. [5]

—, — and LOZACHMEUR, J.-M. (2010). Tagging and income taxation: theory and an application. *American Economic Journal: Economic Policy*, **2** (1), 31–50. [19]

CRONIN, J. A., FULLERTON, D. and SEXTON, S. (2019). Vertical and horizontal redistributions from a carbon tax and rebate. *Journal of the Association of Environmental and Resource Economists*, **6** (S1), S169–S208. [4]

DAVIS, L. W. and KNITTEL, C. R. (2019). Are fuel economy standards regressive? *Journal of the Association of Environmental and Resource Economists*, **6** (S1), S37–S63. [4]

DENNIG, F., BUDOLFSON, M. B., FLEURBAEY, M., SIEBERT, A. and SOCOLOW, R. H. (2015). Inequality, climate impacts on the future poor, and carbon prices. *Proceedings of the National Academy of Sciences*, **112** (52), 15827–15832. [11]

DEPARTMENT OF HEALTH AND HUMAN SERVICES (2024). https://www.acf.hhs.gov/ocs/fact-sheet/liheap-fact-sheet. Accessed: April 7, 2024. [3]

DIAMOND, P. A. (1998). Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review*, pp. 83–95. [3, 10]

DOUENNE, T. (2020). The vertical and horizontal distributive effects of energy taxes: A case study of a french policy. *The Energy Journal*, **41** (3). [4]

—, HUMMEL, A. J. and PEDRONI, M. (2023). Optimal fiscal policy in a climate-economy model with heterogeneous households. *Available at SSRN 4018468*. [5, 10]

DUBÉ, J.-P. and MISRA, S. (2023). Personalized pricing and consumer welfare. *Journal of Political Economy*, **131** (1), 131–189. [24]

DWORCZAK, P., KOMINERS, S. D. and AKBARPOUR, M. (2021). Redistribution through markets. *Econometrica*, **89** (4), 1665–1698. [4, 7]

ERRICKSON, F. C., KELLER, K., COLLINS, W. D., SRIKRISHNAN, V. and ANTHOFF, D. (2021). Equity is more important for the social cost of methane than climate uncertainty. *Nature*, **592** (7855), 564–570. [11]

EUROPEAN COMMISSION (2024). https://alternative-fuels-observatory.ec.europa.eu/transport-mode/road/france/incentives-legislations. Accessed: April 7, 2024. [3]

FABRA, N. and REGUANT, M. (2014). Pass-through of emissions costs in electricity markets. *American Economic Review*, **104** (9), 2872–2899. [22]

FEGER, F. and RADULESCU, D. (2020). When environmental and redistribution concerns collide: The case of electricity pricing. *Energy economics*, **90**, 104828. [5, 12]

FEREY, A., LOCKWOOD, B. and TAUBINSKY, D. (2023). Sufficient statistics for nonlinear tax systems with general across-income heterogeneity. *National Bureau of Economic Research*. [5, 12]

FISCHER, C. and PIZER, W. A. (2019). Equity versus efficiency in energy regulation. *Journal of the Association of Environmental and Resource Economists*. [4]

GOLOSOV, M. and KRASIKOV, I. (2023). *The Optimal Taxation of Couples*. Tech. rep., National Bureau of Economic Research. [5, 17]

GOULDER, L. H. (1995). Environmental taxation and the double dividend: a reader's guide. *International tax and public finance*, **2**, 157–183. [10]

GRAINGER, C. A. and KOLSTAD, C. D. (2010). Who pays a price on carbon? *Environmental and Resource Economics*, **46** (3), 359–376. [4]

GROVES, T. (1973). Incentives in teams. *Econometrica*, **41** (4), 617–631. [6]

HAHN, R. W. and METCALFE, R. D. (2021). Efficiency and equity impacts of energy subsidies. *American Economic Review*, **111** (5), 1658–1688. [4]

HALSE, A., HAUGE, K. E., ISAKSEN, E. T., JOHANSEN, B. and RAAUM, O. (2024). Local incentives and electric vehicle adoption1. *Journal of the Association of Environmental and Resource Economists (forthcoming)*. [20, 21, 32]

HASSETT, K. A., MATHUR, A. and METCALF, G. E. (2009). The incidence of a us carbon tax: A lifetime and regional analysis. *The Energy Journal*, **30** (2). [4]

HENDREN, N. (2020). Measuring economic efficiency using inverse-optimum weights. *Journal of Public Economics*, **187**, 104198. [21]

HÄNSEL, M. C., FRANKS, M., KALKUHL, M. and EDENHOFER, O. (2022). Optimal carbon taxation and horizontal equity: A welfare-theoretic approach with application to german household data. *Journal of Environmental Economics and Management*, **116**, 102730. [5]

INTERNAL REVENUE SERVICE (2024). https://www.irs.gov/credits-deductions/credits-for-new-clean-vehicles-purchased-in-2023-or-after. Accessed: April 7, 2024. [3]

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. [5, 10]

—, JONGEN, E. L. and ZOUTMAN, F. T. (2017). Revealed social preferences of dutch political parties. *Journal of Public Economics*, **156**, 81–100. [21]

— and VAN DER PLOEG, F. (2019). Redistribution and pollution taxes with non-linear engel curves. *Journal of Environmental Economics and Management*, **95**, 198–226. [12]

KAPLOW, B. L. (2012). Optimal control of externalities in the presence of income taxation. *International Economic Review*, **53** (2), 487–509. [5]

KAPLOW, L. (1989). Horizontal equity: measures in search of a principle. *National Tax Journal*, **42** (2), 139–154. [8]

KENNAN, J. and WALKER, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, **79** (1), 211–251. [19]

KLEVEN, H. J., KREINER, C. T. and SAEZ, E. (2009). The optimal income taxation of couples. *Econometrica*, **77** (2), 537–560. [5]

KREINER, C. T. and VERDELIN, N. (2012). Optimal provision of public goods: A synthesis. *The Scandinavian Journal of Economics*, **114** (2), 384–408. [19]

KÄNZIG, D. R. (2023). *The Unequal Economic Consequences of Carbon Pricing*. Working Paper 31221, National Bureau of Economic Research. [4]

LEHMANN, E., SIMULA, L. and TRANNOY, A. (2014). Tax me if you can! optimal nonlinear income tax between competing governments. *The Quarterly Journal of Economics*, **129** (4), 1995–2030. [5]

LISKI, M. and VEHVILÄINEN, I. (2020). Gone with the wind? an empirical analysis of the equilibrium impact of renewable energy. *Journal of the Association of Environmental and Resource Economists*, **7** (5), 873–900. [22]

MILLIGAN, K. (2005). Subsidizing the stork: New evidence on tax incentives and fertility. *Review of Economics and statistics*, **87** (3), 539–555. [19]

MUSGRAVE, R. A. (1990). Horizontal equity, once more. *National Tax Journal*, **43** (2), 113–122. [8]

NURMI, V. and AHTIAINEN, H. (2018). Distributional weights in environmental valuation and cost-benefit analysis: theory and practice. *Ecological economics*, **150**, 217–228. [11]

OECD (2021). Effective carbon rates: Key findings for finland. *Effective Carbon Rates: Key findings for Finland*. [21]

PAI, M. and STRACK, P. (2022). *Taxing Externalities Without Hurting the Poor*. Tech. rep., Cowles Foundation for Research in Economics, Yale University. [4]

PIZER, W. A. and SEXTON, S. (2020). The distributional impacts of energy taxes. *Review of Environmental Economics and Policy*. [4]

ROCHET, J.-C. and STOLE, L. A. (2003). The economics of multidimensional screening. *Econometric Society Monographs*, **35**, 150–197. [5, 8, 24, 17]

SAEZ, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, **68** (1), 205–229. [3]

— (2002). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics*, **83** (2), 217–230. [5, 12]

— and STANTCHEVA, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, **106** (01), 24–45. [21]

SEIERSTAD, A. and SYDSÆTER, K. (1987). *Optimal control theory with economic applications*. Elsevier North-Holland, Inc. [13]

SHER, I. (2023). *Generalized Social Marginal Welfare Weights Imply Inconsistent Comparisons of Tax Policies*. Working Papers 2021-009, Human Capital and Economic Opportunity Working Group. [18]

SPIRITUS, K., LEHMANN, E., RENES, S. and ZOUTMAN, F. (2022). Optimal taxation with multiple incomes and types. *Tinbergen Institute Discussion Paper 2022-006/VI*. [5, 17]

STITZING, R. (2016). Distributional and environmental effects of an emissions-differentiated car sales tax. *Available at SSRN 3050123*. [21]

WEST, S. E. (2004). Distributional effects of alternative vehicle pollution control policies. *Journal of Public Economics*, **88** (3-4), 735–757. [4]

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), 195–213. [4]

Appendix A: Proofs

A.1 *Theorem 1*

Proof. The assumption in the theorem is that the cutoff $\overline{q}$ is the same for all $n \in [\underline{n}, \overline{n}]$, and earnings y depend only on ability n and thus $y^0(n) = y^1(n) =: y(n)$. Under these assumptions, the social welfare expression from the text becomes

$$\mathbb{E}[\omega u(\theta)] = \mathbb{E}\big[y(n) - k(y(n), n) + (\mathbb{E}[\omega b] - q)\mathbb{1}_{q \leq \overline{q}} + (\omega - 1)\mathcal{R}(n, q)\big], \qquad (3'')$$

with the rents term taking the form

$$\mathcal{R}(n, q) = \big[-\int_{\underline{n}}^{n} k_n(y(s), s) ds\big] + \big[\overline{q} - q\big]\mathbb{1}_{q \leq \overline{q}}. \qquad (6')$$

Integrate the first square bracket on the RHS by parts to rewrite the last term in $\mathbb{E}[\omega u(\theta)]$ as

$$\mathbb{E}(\omega - 1)\mathcal{R}(n, q) = (1 - \mathbb{E}[\omega | n' \geq n])\frac{1 - F_n(n)}{f_n(n)} k_n(y(n), n) + \mathbb{E}\big[(\omega - 1)(\overline{q} - q)\mathbb{1}_{q \leq \overline{q}}\big]. \quad \text{(A.1)}$$

We observe that the part of welfare depending on y is

$$\mathbb{E}\Big[y(n) - k(y(n), n) + (1 - \mathbb{E}[w | n' \geq n])\frac{1 - F_n(n)}{f_n(n)} k_n(y(n), n)\Big]. \qquad \text{(A.2)}$$

For each n, $\mathbb{E}[\omega | n' \geq n]$ is exogenous and earnings can be characterized point-wise through the first-order condition

$$1 - k_y(y(n), n) + \big(1 - \mathbb{E}[\omega | n' \geq n]\big)\frac{1 - F_n(n)}{f_n(n)} k_{ny}(y(n), n) = 0. \qquad \text{(A.3)}$$

If term $\big(1 - \mathbb{E}[\omega | n' \geq n]\big)\frac{1 - F_n(n)}{f_n(n)} k_{ny}(y(n), n)$ is strictly increasing in n, the solution for $y(n)$ solving eq. (A.3) is unique and strictly increasing in n. Moreover, the iso-elastic form gives $k_{yn} = -\big(1 + \frac{1}{\epsilon}\big)\frac{1}{n} k_y$ which, when applying $T'(y) = 1 - k_y$, leads to the formula in Theorem 1.

Turn now to the part of the welfare that depends on the choice of $\overline{q}$:

$$\mathbb{E}\big[(\mathbb{E}[\omega b] - q)\mathbb{1}_{q \leq \overline{q}} + (\omega - 1)(\overline{q} - q)\mathbb{1}_{q \leq \overline{q}}\big]. \qquad \text{(A.4)}$$

For a finite $\overline{q} > 0$ the optimality condition is

$$(\mathbb{E}[\omega b] - \overline{q}) f_q(\overline{q}) + (\mathbb{E}\big[\omega | q \leq \overline{q}\big] - 1) F_q(\overline{q}) = 0$$

which can also be written as

$$(\mathbb{E}[\omega b] - \overline{q}) f_q(\overline{q}) + (1 - \mathbb{E}\big[\omega | q > \overline{q}\big])(1 - F_q(\overline{q})) = 0$$

which, after rearranging, gives

$$\overline{q} = \mathbb{E}[\omega b] + (1 - \mathbb{E}\big[\omega | q > \overline{q}\big])\frac{1 - F_q(\overline{q})}{f_q(\overline{q})} = B^{\omega} + \frac{1 - \mathbb{E}\big[\omega | q > \overline{q}\big]}{h_q(\overline{q})}. \qquad \text{(A.5)}$$

Replacing $\overline{q} = \tau$ gives the formula for the externality tax in the theorem. □

A.2 *Theorem 2*

PROOF. We derive here the optimal tax schedules of Theorem 2 by considering a simple tax reform. In the Online Appendix, the theorem is proved using optimal-control theory. To derive formula (14), consider a reform that increases the externality tax above income $y^0(n)$ by $d\tau$ by increasing $\tau'(y)$ marginally in a small band $[y^0(n), y^0(n+dn)]$. Such a reform should not change welfare around the optimal tax schedule.

The reform has three effects. The first effect is the earnings response: around n, polluters face a more progressive tax schedule and consequently decrease their earnings and pay less taxes. Let $d\tau'(y)$ denote the change in marginal tax $\tau'(y)$ in $[y^0(n), y^0(n+dn)]$, and let $dy := y^0(n+dn) - y^0(n)$. We have $d\tau = dy d\tau'(y) = dn \frac{dy^0}{dn} d\tau'(y) = dn$ $(-\frac{k_{ny}(y^0(n),n)}{k_{yy}(y^0(n),n)})d\tau'(y)$, in which the last equation follows from the implicit function theorem applied to the agent's first-order condition $1 - k_y(y^0(n), n) - T'(y^0(n)) - \tau'(y^0(n)) = 0$. The change in the marginal tax in the small band is thus $d\tau'(y) = -d\tau \frac{1}{dn} \frac{k_{yy}(y^0(n),n)}{k_{ny}(y^0(n),n)}$ and the associated change in earnings is $-d\tau \frac{1}{dn} \frac{k_{yy}(y^0(n),n)}{k_{ny}(y^0(n),n)} \cdot \frac{dy^0}{d\tau'(y)} = d\tau \frac{1}{dn} \frac{1}{k_{ny}(y^0(n),n)}$ in which the last equation uses again implicit differentiation of the first-order condition for the agent. The change in tax revenue equals the change in earnings times the marginal tax: $d\tau \frac{1}{dn} \frac{1}{k_{ny}(y^0(n),n)}(T'(y^0) + \tau'(y^0)) = -d\tau \frac{1}{dn} \frac{n}{(1+\frac{1}{\epsilon})k_y(y^0(n),n)}(T'(y^0) + \tau'(y^0))$ $= -d\tau \frac{1}{dn} \frac{n}{(1+\frac{1}{\epsilon})} \frac{T'(y^0)+\tau'(y^0)}{1-T'(y^0)-\tau'(y^0)}$, and when we multiply by the mass $(1 - F_{q|n}(\bar{q}|n)) f_n(n) dn$ of agents affected, we obtain the total earnings-response effect:

$$- d\tau(1 - F_{q|n}(\bar{q}|n)) f_n(n) \frac{n}{1 + \frac{1}{\epsilon}} \frac{T'(y^0) + \tau'(y^0)}{1 - T'(y^0) - \tau'(y^0))}, \tag{A.6}$$

The second effect is a redistributive effect, as the reform mechanically redistributes from rich polluters to all individuals:

$$d\tau \mathbb{E}[(1 - \overline{\omega}^0_{n'})(1 - F_{q|n}(\bar{q}|n'))|n' \geq n](1 - F_n(n)). \tag{A.7}$$

The third effect is an externality effect which arises because some rich polluters stop polluting. The utilities of these individuals are unchanged, but the externality action produces marginal net benefits $B^\omega - \tau^n$ per individual through aggregate externality benefit and tax revenue changes. Multiplying by the mass of affected individuals, this becomes

$$d\tau \mathbb{E}[(B^\omega - \tau^n) f_{q|n}(\bar{q}|n')|n' \geq n](1 - F_n(n)). \tag{A.8}$$

Around the optimal tax schedule, the three effects should cancel out. Hence, by setting the sum of equations (A.6), (A.7) and (A.8) to zero, we obtain optimal tax formula (14).

Tax formula (15) can be derived analogously by considering a small reform that increases the tax burden of rich, clean individuals.

□

APPENDIX B: DETAILS ON THE EMPIRICAL APPLICATIONS

This section provides details of the empirical applications. Both applications make use of the confidential, pseudonymized individual-level data that is achieved through Statistics Finland Fiona remote access system. The data includes information about households' incomes, which we divide into five income quintiles based on households' disposable income. The income groups are created by first dividing the household's income by the consumption units of the household (i.e., equivalent incomes), and then households are divided into five equal groups.

B.1 *Details on vehicles, owners, kilometers, and taxes*

The data comes from the Statistics Finland's 2021 FOLK longitudinal dataset[38] which is linked to the 2023 vehicle register from TRAFICOM (Finnish transportation and communications authority). The vehicle register contains odometer readings from vehicle inspections for the entire fleet; the individual linked to the kilometers must be deduced from ownership spells. For about 75 % the mileage of the vehicles in 2016 is obtained directly from the mileage readings at the inspection, as the ownership has not changed. For 9 % of individuals, data from previous or following years have been used to construct a full year of kilometers. For the remaining 16 %, the kilometers have been estimated by prediction method based on observable characteristics of the individual. We use 2016 kilometers obtained this way for each car user in the subsequent analysis; this construction of kilometers builds on Ahonen (2023).

TRAFICOM vehicle registry contains technical information on cars, including the fuel type and efficiency. This data is used to transform kilometers for each vehicle-user pair to fuel consumption and, further, to total fuel expenditures using the tax inclusive prices. Fuel price data is from Statistics Finland. We use average annual fuel price in the expenditure calculation. We obtain the variables describing the person and the household, such as income data and other socio-economic characteristics, from the FOLK data. Transaction data on new cars is obtained directly from the register of vehicles, as new registrations in TRAFICOM data. Table B.1 provides key descriptive statistics of the vehicle data set; share of EVs in transactions in Fig. 1.

Semi-elasticities of EV adoption from previous literature. We get the semi-elasticity of pollution reduction by income quantile from an external source. Halse *et al.* (2024) estimate the effect of road pricing on electric vehicle adoption. Our assumption is that road pricing has a similar effect on EV adoption that other policies that make driving more expensive, including carbon pricing. We use data from their Figure A.9 (right panel) to get the quintile-level elasticities with respect to the number of internal combustion engine vehicle (ICEVs) for couples (the elasticities for the five income quintiles are: 1st: -0.0088, 2nd: -0.0028, 3rd: -0.0066, 4nd: -0.0107 and 5th: -0.0137). These numbers tell the impact of a one-euro increase in daily road tolls (or other road use fees) on the number of ICEVs. We convert these numbers into quasi-elasticities by using the mean road

[38]This is the most recent available Folk dataset.

TABLE B.1. Descriptive statistics: vehicles

Variable	N	Mean	Sd	p10	p90
Car fleet					
annual fuel exp. (EUR/year)	2,152,443	1569	1593	397	2912
daily kilometers	2,152,443	46	41	11	86
consumption (liter/100 km)	2,152,443	7.2	1.79	5.1	9.3
Disposable income (EUR/year)	2,152,443	27,684	15,609	13,100	43,400
Transaction data 2023, q1-q3					
annual fuel exp. (EUR/year)	44,360	1789	1707	634	3143
daily kilometers	44,360	57	44	20	101
consumption (liter/100 km)	44,360	6.6	1.8	4.7	9
Disposable income (EUR/year)	44,360	37,981	21,834	18,300	63,000

Note: Data from the Statistics Finland's 2021 FOLK longitudinal dataset which is linked to the 2023 vehicle register from TRAFICOM (Finnish transportation and communications authority). Annual fuel expenditure, kilometers, and consumption for each car owner are based on 2016 kilometers (vehicle register) and final fuel prices (Statistic Finland). Disposable income of a car owner is net of taxes and transfers. Transaction data 2023 is from quarters 1-3, new car purchases. Variable definitions are the same as for the car fleet but the sample restricted to new car buyers. Share of EVs in transactions is depicted in Fig. 1.

toll for couples: 9.49 NOK or 0.826 euros. This gives the semi-elasticities that we report in the data.

We can compare these results to those using Swiss data presented by Bigler and Radulescu (2022). Their changes in adoption probabilities for a vehicle tax feebate shows the same pattern for adoption of an ICEV across income quartiles (1st quartile: -.0434, 2nd quartile: -.0433, 3rd quartile: -.0424, 4nd quartile: -.0709). The highest income group is the most elastic, but the effect is monotonic for the other income groups. Note that our quantification uses the relative differences between income groups (Parameter ξ_2 is Proposition 3), and not their absolute values.

B.2 *Electricity use*

We use electricity consumption data for Finnish households from September 2022 to January 2023. The data is from Fingrid Datahub, a centralized information exchange system for the electricity retail market. It includes monthly electricity consumption for all households in Finland. The dataset also includes basic information about contract type (start date, end date and whether the contract is fixed-price). We connect consumption per electricity meter to a dwelling unit, or household for shorthand, based on the social security number of the individual who initiated the contract. Of the 3.2 million observations in the electricity consumption data, about 115,000 cannot be connected to any dwelling units. We find no electricity contract for about 296,000 households in the population register, for instance, a landlord may have procured the electricity contract on behalf of the tenant. Additionally, the data does not include electricity contracts in the autonomous region of Aland. We drop households with income less than 5000 euros per year and electricity consumption over 50 percent of disposable income as outliers. A more detailed description of the dataset is found in Ahlvik *et al.* (2023).

The household-level contract price is not available in the data. Instead of actual prices, we use the mean electricity price per contract type at the time of the contract

TABLE B.2. Descriptive statistics: Electricity use

	N	Mean	Sd	p10	p90
Electricity consumption (kWh)					
September 2022	2,530,334	367.8	408.9	61.6	881.0
October 2022	2,536,155	435.5	487.1	65.5	1077.9
November 2022	2,536,947	542.0	634.5	46.9	1412.9
December 2022	2,536,992	708.4	859.0	46.8	1916.2
January 2023	2,538,713	668.8	807.8 9	68.0	1796.3
Household-level data					
Disposable income (Eur/year)	2,530,334	41848.8	54.863	13993.2	74919.5
Household size	2,501,503	2.100	1.263	1	4

Note: Data from the Statistics Finland's 2020 FOLK longitudinal dataset which is linked to Fingrid Datahub -consumption data based on the person whose name the contract was on. Electricity consumption data is observed monthly and summed over all properties owned by a household. Disposable-income is per household and net of taxes and transfers. Household size is the number of people living in the household.

start date.[39] The background data of households comes from Statistics Finland administrative data. The main variable of interest is disposable income, which consists of wage income, entrepreneurial income, property income, and received income transfers, minus taxes and tax-like payments paid.

Estimating semi-elasticities of electricity consumption. We empirically identify the semi-elasticities by a difference-in-differences approach. Our unit of observation is a household, defined as a combination of individual-id of the person paying the bill and a metering point. We compare electricity consumption of same households between October vs. December, and compare households whose fixed-term contract ended in November to those who had a fixed-term contract throughout the crisis.[40] The assumption is that households could not anticipate the 2022 energy crisis when signing their electricity contracts. The contracts that expired were signed one year or two years earlier.

We define treatment for households whose fixed-term contract ends in November, and use households whose contract ends after the crisis as a control group. We estimate the following OLS regression:

$$\log(Q_{it}) = \beta Price_{it} + \gamma_i + \gamma_{rt} + \epsilon_{it} \tag{B.1}$$

where Q_{it} is the electricity use of household i in month $t = \{October, December\}$. $Price_{it}$ is the average contract price for fixed-term contracts (0.065 Eur/kWh). Contracts that end in November face a higher price in December (0.38 Eur/kWh). In the equation we introduce fixed effects to control for unobservable factors on household- and region-by-month level. In the equation, γ_i is the household-fixed effect capturing e.g. the size of the apartment or heating technology, and γ_{rt} is a region-by-month-fixed effect capturing

[39]The data is from the Finnish energy authority https://energiavirasto.fi/sahkon-hintatilastot

[40]We have monthly consumption data, and therefore the impact of contracts ending in November is visible in December.

TABLE B.3. The impact of electricity prices on consumption

	Treatment (1)	Placebo (2)
Panel A: Average treatment effect		
ATT	-0.2420	-0.0202
	(0.0068)	(0.0070)
N	2,252,016	2,246,030
Panel B: Treatment effect by income group		
1st quintile	-0.1094	-0.0237
	(0.0167)	(0.0177)
2nd quintile	-0.2004	-0.0005
	(0.0147)	(0.0154)
3rd quintile	-0.2884	-0.0180
	(0.0146)	(0.0144)
4th quintile	-0.3386	-0.0569
	(0.0152)	(0.0146)
5th quintile	-0.3085	-0.0332
	(0.0145)	(0.0146)
N	2,252,016	2,246,030
Fixed effects	Month-province Individual-meter	Month-province Individual-meter

Note: The table presents OLS coefficients (β) with log of electricity consumption as the dependent variable. The data is for October and December of year 2022. The treatment variable is the price-level, taking value 0.38 Eur/kWh in December for the treatment group and 0.0065 Eur/kWh for other observations. Column 1 the treatment group is households whose contracts ended in November 2022, during the energy crisis. Column 2 is a "placebo" treatment for households whose contract ended in January 2023, after the crisis. Fixed effects (FE) and controls included as indicated in the bottom rows. Robust standard errors in parentheses clustered on household level.

e.g. local differences in temperature or energy savings campaigns. We define a region to be one of the 19 provinces (*maakunta*) of Finland.

The main results are shown in column (1), Panel A of Table B.3. We find an average treatment effect of -0.2420: a 10 c/kWh increase in electricity price would reduce consumption by 2.42%. Column (2), Panel A of Table B.3 presents a variation, where we define a Placebo-treatment such that variable $Price_{it}$ jumps to a higher price (0.38 Eur/kWh) for households whose contracts end in January 2023. This treatment is "Placebo" in the sense, that these household had no real change in their electricity price through the study period (October to December), and we expect to find no effect in their consumption. Table B.3 shows that the effect on this group is small.

We are mainly interested in the heterogeneous treatment effect by income quintile, which we estimate by:

$$\log(Q_{it}) = \sum_{q=1}^{5} \beta_q Price_{it} \times Quintile_i + \gamma_i + \gamma_{rt} + \epsilon_{it} \tag{B.2}$$

where $Quintile_i$ is an indicator taking value one if the household i's income is in a given quantile bracket (note that household-level fixed effects γ_i also capture $Quintile_i$).

The main results are shown in Panel B of Table B.3. We find a decrease for all income quintiles, but the effect is mainly driven by higher income quintiles. For the lowest income quintile, a 10 c/kWh increase in electricity price would reduce consumption by 1.09%. The elasticity is highest for fourth quintile, for which a 10 c/kWh increase in electricity price reduces consumption by 3.39%. We find no similar effects for the Placebo-group whose contract ends in January, as shown by column (2), Panel B of Table B.3.

Pigouvian Income Taxation

Lassi Ahlvik, Matti Liski, and Mikael Mäkimattila

April 25, 2024

ONLINE APPENDIX

Appendix C: Supplementary online material

C.1 *Optimal-control approach to Theorem 2*

PROOF. We solve the allocation problem in two stages. First, we characterize the optimal allocation when the aggregate externality $\overline{x}$ is taken as a given constraint. Second, we optimize with respect to $\overline{x}$. The tax schedules are then defined by the optimal allocation.

We use definitions $\widehat{v}^0(n) := y^0(n) - k(y^0(n), n) - t^0(n)$ and $\widehat{v}^1(n) := y^1(n) - k(y^1(n), n) - t^1(n)$ where $t^0(n)$ and $t^1(n)$ are transfers associated with actions $x = 0$ and $x = 1$, respectively. We also use notation $\overline{\omega}(n, q) := \mathbb{E}[\omega | n, q]$. Notice that $\overline{q}(n) = \widehat{v}^1(n) - \widehat{v}^0(n)$.

The objective function is $\mathbb{E}[\omega u(\theta)] = \mathbb{E}[\omega\big(v^1(n) - q + b\overline{x}\big)\mathbb{1}_{q\leq\overline{q}(n)} + \omega\big(v^0(n) + b\overline{x}\big)\mathbb{1}_{q>\overline{q}(n)}]$, or $\mathbb{E}[\omega u(\theta)] = \mathbb{E}[\omega b]\overline{x} + \mathbb{E}[(\omega\widehat{v}^1(n) - q)\mathbb{1}_{q\leq\overline{q}(n)} + \omega\widehat{v}^0(n)\mathbb{1}_{q>\overline{q}(n)}]$. Using $B^\omega = \mathbb{E}[\omega b]$ for the aggregate externality benefit and holding $B^\omega\overline{x}$ as given, we can write the objective $\mathbb{E}[\omega u(\theta)]$ in integral form as

$$\int_{\underline{n}}^{\overline{n}} \left(B^\omega \overline{x} + \int_{q\leq\widehat{v}_1(n)-\widehat{v}^0(n)} \overline{\omega}(n,q)(\widehat{v}^1(n) - q) f_{q|n}(q|n)dq \right.$$
$$\left. + \int_{q\geq\widehat{v}_1(n)-\widehat{v}^0(n)} \overline{\omega}(n,q)\widehat{v}^0(n) f_{q|n}(q|n)dq \right) f_n(n)dn$$
$$= \int_{\underline{n}}^{\overline{n}} a_0(n)dn \tag{C.1}$$

where we define the function inside the integral as $a_0(n)$.

The incentive constraints are

$$\widehat{v}^{0\prime}(n) = -k_n(y^0(n), n) \tag{C.2}$$

$$\widehat{v}^{1\prime}(n) = -k_n(y^1(n), n). \tag{C.3}$$

The budget constraint is

$$\int_{\underline{n}}^{\overline{n}} \left(\Big(y^0(n) - k(y^0(n), n) - \widehat{v}^0(n)\Big)(1 - F_{q|n}(\widehat{v}^1(n) - \widehat{v}^0(n)|n)) \right.$$
$$\left. + \Big(y^1(n) - k(y^1(n), n) - \widehat{v}^1(n)\Big) F_{q|n}(\widehat{v}^1(n) - \widehat{v}^0(n)|n) \right) f_n(n)dn = 0$$

$$\implies \int_{\underline{n}}^{\overline{n}} a_1(n)dn = 0 \tag{C.4}$$

where we define the function inside the integral as $a_1(n)$. The constraint on the aggregate externality $\overline{x}$ is

$$\int_{\underline{n}}^{\overline{n}} \left(F_{q|n}(\widehat{v}^1(n) - \widehat{v}^0(n)|n) - \overline{x} \right) f_n(n)dn = 0$$

$$\Longrightarrow \int_{\underline{n}}^{\overline{n}} a_2(n)dn = 0. \tag{C.5}$$

where we define the function inside the integral as $a_2(n)$.

We apply optimal control theory to solve the problem with the objective in (C.1) and the constraints in (C.2)-(C.5). Lemma C.2 shows that there exists a solution to this optimization problem. The optimal policy can then be characterized by the the Pontryagin's maximum principle. Define the Hamiltonian

$$\begin{aligned} &\mathcal{H}(\widehat{v}^0(n), \widehat{v}^1(n), y^0(n), y^1(n), \mu_0(n), \mu_1(n), n) \\ &= a_0(n) - \mu_0(n)k_n(y^0(n), n) - \mu_1(n)k_n(y^1(n), n) \end{aligned} \tag{C.6}$$

and the associated Lagrangian

$$\mathcal{L}(n, \lambda, \gamma) = \mathcal{H}(n) + \lambda \int_{\underline{n}}^{\overline{n}} a_1(n)dn + \gamma \int_{\underline{n}}^{\overline{n}} a_2(n)dn. \tag{C.7}$$

We denoted the multipliers associated with constraints (C.2), (C.3), (C.4) and (C.5) by $\mu_0(n)$, $\mu_1(n)$, λ and γ, respectively (the bunching constraints are left out because Theorem 2 characterizes allocations only for n at which the monotonicity constraints do not bind, i.e., (y^0, y^1) are strictly increasing in n). From $\mathbb{E}[\omega] = 1$ it follows that $\lambda = 1$. Maximizing Hamiltonian $\mathcal{H}$ with respect to controls (y^0, y^1) gives the following necessary conditions for an interior solution:

$$-\mu_0(n)k_{ny}(y^0(n), n) + (1 - k_y(y^0(n), n))(1 - F_{q|n}(\overline{q}(n)|n)))f_n(n) = 0 \tag{C.8}$$

$$-\mu_1(n)k_{ny}(y^1(n), n) + (1 - k_y(y^1(n), n))F_{q|n}(\overline{q}(n)|n)f_n(n) = 0. \tag{C.9}$$

Write (C.8) as $\frac{\mu_0(n)}{1-F_{q|n}(\overline{q}(n)|n)} = \frac{1-k_y(y^0(n),n)}{k_{ny}(y^0(n),n)}f_n(n)$, in which the right-hand side is increasing in y^0 given the functional form of k, and hence $y^0(n)$ satisfying the condition is the global maximizer of the Hamiltonian. Following similar reasoning, we conclude that $y^1(n)$ satisfying (C.9) is the global maximizer.

Using notation $\overline{\omega}_n^0 = \mathbb{E}[\omega|n, q > \overline{q}(n)]$ and $\overline{\omega}_n^1 = \mathbb{E}[\omega|n, q \leq \overline{q}(n)]$, the maximum principle implies that the co-states must satisfy:

$$-\mu_0'(n) = \Big((\overline{\omega}_n^0 - 1)(1 - F_{q|n}(\overline{q}(n)|n)) + (t^0(n) - t^1(n))f_{q|n}(\overline{q}(n)|n) - \gamma f_{q|n}(\overline{q}(n)|n)\Big)f_n(n) \tag{C.10}$$

$$-\mu_1'(n) = \Big((\overline{\omega}_n^1 - 1)F_{q|n}(\overline{q}(n)|n) - (t^0(n) - t^1(n))f_{q|n}(\overline{q}(n)|n) + \gamma f_{q|n}(\overline{q}(n)|n)\Big)f_n(n) \tag{C.11}$$

with transversality conditions $\mu_0(\overline{n}) = \mu_1(\overline{n}) = \mu_0(\underline{n}) = \mu_1(\underline{n}) = 0$.

Integrating equations (C.10) and (C.11) from n to $\overline{n}$ and using the transversality conditions $\mu_0(\overline{n}) = \mu_1(\overline{n}) = 0$ as well as notation $\tau^n(n) = t^0(n) - t^1(n)$, it follows that

$$\mu_0(n) = \int_n^{\overline{n}} \Big((\overline{\omega}_{n'}^0 - \lambda)(1 - F_{q|n}(\overline{q}(n')|n')) - (\gamma - \tau^n(n'))f_{q|n}(\overline{q}(n')|n')\Big)f_n(n')dn \tag{C.12}$$

$$\mu_1(n) = \int_n^{\overline{n}} \Big((\overline{\omega}^1_{n'} - \lambda) F_{q|n}(\overline{q}(n')|n') + (\gamma - \tau^n(n')) f_{q|n}(\overline{q}(n')|n') \Big) f_n(n') dn \tag{C.13}$$

We then plug equations (C.12) and (C.13) into the necessary conditions equations (C.8) and (C.9) to obtain

$$1 - k_y(y^0(n), n) = \frac{k_{ny}(y^0(n), n)}{f_n(n)(1 - F_{q|n}(\overline{q}(n)|n))} \int_n^{\overline{n}} \Big((\overline{\omega}^0_{n'} - 1)(1 - F_{q|n}(\overline{q}(n')|n')) - (\gamma - \tau^n(n')) f_{q|n}(\overline{q}(n')|n') \Big) f_n(n') dn' \tag{C.14}$$

and

$$1 - k_y(y^1(n), n) = \frac{k_{ny}(y^1(n), n)}{f_n(n) F_{q|n}(\overline{q}(n)|n)} \int_n^{\overline{n}} \Big((\overline{\omega}^1_{n'} - 1) F_{q|n}(\overline{q}(n')|n') + (\gamma - \tau^n(n')) f_{q|n}(\overline{q}(n')|n') \Big) f_n(n') dn'. \tag{C.15}$$

Given the iso-elastic class for $k(y,n)$, we have

$$k_{ny}(y^i(n), n) = -\left(1 + \frac{1}{\epsilon}\right) \frac{1}{n} k_y(y^i(n), n), i = 0, 1. \tag{C.16}$$

By using (C.16) in (C.1) and (C.1), it follows that

$$\frac{1 - k_y(y^0(n), n)}{k_y(y^0(n), n)} = \left(1 + \frac{1}{\epsilon}\right) \frac{\mathbb{E}[(1 - \overline{\omega}^0_{n'}(1 - F_{q|n}(\overline{q}(n')|n')) + (\gamma - \tau^n(n')) f_{q|n}(\overline{q}(n')|n')|n' \geq n]}{n(1 - F_{q|n}(\overline{q}(n)|n)) h_n(n)} \tag{C.17}$$

$$\frac{1 - k_y(y^1(n), n)}{k_y(y^1(n), n)} = \left(1 + \frac{1}{\epsilon}\right) \frac{\mathbb{E}[(1 - \overline{\omega}^1_{n'} F_{q|n}(\overline{q}(n')|n') - (\gamma + \tau^n(n')) f_{q|n}(\overline{q}(n')|n')|n' \geq n]}{n F_{q|n}(\overline{q}(n)|n) h_n(n)} \tag{C.18}$$

In the second step, we optimize with respect to the aggregate externality $\overline{x}$. From the social welfare function in (C.1), the marginal benefit from an increase in $\overline{x}$ is $\mathbb{E}[\omega b] = B^\omega$. Hence, in the optimal solution, the aggregate externality $\overline{x}$ is selected so that the shadow cost γ of the aggregate externality constraint (C.5) equals B^ω. By plugging $\gamma = B^\omega$ and using individual optimality conditions (9) and (10) in equations (C.1) and (C.1), we obtain the optimal tax formulas of Theorem 2.

Note that by the transversality conditions, $k_y(y^0(n), n) = 1$ and $k_y(y^1(n), n) = 1$ for $n \in \{\underline{n}, \overline{n}\}$ so we have no distortion at the bottom and the top:

$$T'(y) = \tau'(y) = 0 \text{ for } y \in \{y^i(\underline{n}), y^i(\overline{n})\}, i = \{0, 1\}. \tag{C.19}$$

From this it also follows that $y^0(\underline{n}) = y^1(\underline{n})$ and $y^0(\overline{n}) = y^1(\overline{n})$. □

C.2 *Proposition 1*

PROOF. **Part (a)**: Assume first independence $n \perp\!\!\!\perp q$. Assumption 2a implies that

$$\mathbb{E}[\omega|q > \tau, n = n'] > \mathbb{E}[\omega|n = n'] \tag{C.20}$$

for all $n' \in [\underline{n}, \overline{n}]$. Then we can take expectation on both sides of the inequality to be able to write it as

$$\mathbb{E}_{n' \sim F_n}[\mathbb{E}[\omega|q > \tau, n = n']] > \mathbb{E}_{n' \sim F_n}[\mathbb{E}[\omega|n = n']].$$

By independence, $F_q(\tau) = F_{q|n}(\tau|n')$ for all $n' \in [\underline{n}, \overline{n}]$. Hence we can write the inequality as

$$\frac{\mathbb{E}_{n' \sim F_n}[\mathbb{E}[\omega|q > \tau, n = n'](1 - F_{q|n}(\tau|n'))]}{1 - F_q(\tau)} > \mathbb{E}_{n' \sim F_n}[\mathbb{E}[\omega|n = n']],$$

or simplifying, $\mathbb{E}[\omega|q > \tau] > 1$. From Theorem 1, the result $\tau < B^\omega$ follows.

Now drop the independence assumption but assume $\mathbb{E}[\omega|n] = 1$ for all $n \in [\underline{n}, \overline{n}]$. We may take inequality (C.20) and multiply both sides by $(1 - F_{q|n}(\tau|n'))$ to write that

$$\mathbb{E}[\omega|q > \tau, n = n'](1 - F_{q|n}(\tau|n')) > \mathbb{E}[\omega|n = n'](1 - F_{q|n}(\tau|n'))$$

is true for all $n' \in [\underline{n}, \overline{n}]$. Then plug in $\mathbb{E}[\omega|n = n'] = 1$, take expectation on both sides and divide by $1 - F_q(\tau)$ to write the inequality as

$$\frac{\mathbb{E}_{n' \sim F_n}[\mathbb{E}[\omega|q > \tau, n = n'](1 - F_{q|n}(\tau|n'))]}{1 - F_q(\tau)} > 1$$

or simplifying, $\mathbb{E}[\omega|q > \tau] > 1$. From Theorem 1, the result $\tau < B^\omega$ follows.
Part (b): This part of the proposition can be shown by reversing inequality (C.20) and following similar steps.

□

C.3 *Proposition 2*

PROOF. **Part (a)**: Assumption 1 holds. Then under the FOSD assumption, both $\mathbb{E}[\omega|n = n']$ and $(1 - F_{q|n}(\tau|n = n'))$ increase in n', hence $\mathbb{E}_{n' \sim F_n}[\mathbb{E}[\omega|n = n'](1 - F_{q|n}(\tau|n'))] \geq \mathbb{E}_{n' \sim F_n}[\mathbb{E}[\omega|n = n']]\mathbb{E}_{n' \sim F_n}[1 - F_{q|n}(\tau|n')]$. Divide both sides by $1 - F_q(\tau)$ and simplify to write the inequality as $\mathbb{E}[\omega|q > \tau] \geq 1$. Apply Theorem 1, and the result follows.
Part (b): This can be proved analogously.

□

C.4 *Proposition 3*

PROOF. We develop three effects of the simple reform.

First, the tax increase is phased in over the income bracket $[y^*(n), y^*(n+dn)]$, increasing the effective marginal income tax of polluters in the bracket (but not elsewhere). Thus, these individuals face a stand-alone simple income tax reform, which reduces tax revenue through earnings response by[41]

$$-d\tau f_n(n)\frac{n}{1+\dfrac{1}{\epsilon}}\frac{T'(y^*(n))}{1-T'(y^*(n))}(1-F_{q|n}(\tau|n)).$$

Since T is the optimal stand-alone income tax, by Theorem 1, this is equal to

$$-\,d\tau(1-\mathbb{E}[\omega|n' \geq n])(1-F_n(n))(1-F_{q|n}(\tau|n)) \tag{C.21}$$

which gives the redistributive gain (or loss) from the change in tax revenues due to the behavioral response.

Second, tax revenues change because more will be collected from those who do not change behavior, that is, high-income individuals who continue polluting. This higher revenue leads to the following redistributive gain:

$$d\tau\big(1-\mathbb{E}[\omega|n' \geq n, q>\tau]\big)\big(1-F_n(n)\big)\big(1-\mathbb{E}[F_{q|n}(\tau|n')|n' \geq n]\big). \tag{C.22}$$

Third, the reform produces environmental benefits, measured as marginal net benefits $B^\omega - \tau$ per individual which, when multiplied by the total increase in reductions $d\tau\mathbb{E}[f_{q|n}(\tau|n')|n' \geq n](1-F_n(n))$, gives the total environmental gain. Once again we apply Theorem 1 to re-express the total environmental gain:

$$-\,d\tau\frac{1-\mathbb{E}[\omega|q>\tau]}{h_q(\tau)}\mathbb{E}[f_{q|n}(\tau|n')|n' \geq n](1-F_n(n)). \tag{C.23}$$

A progressive externality tax reform, setting higher externality tax for high-income individuals, is welfare-improving at n, if the net effect of the three effects given in eqs. (C.21)-(C.23) is positive:

$$\begin{aligned} d\tau\Big(&-(1-\mathbb{E}[\omega|n' \geq n])(1-F_n(n))(1-F_{q|n}(\tau|n)) \\ &+\big(1-\mathbb{E}[\omega|n' \geq n, q>\tau]\big)\big(1-F_n(n)\big)\big(1-\mathbb{E}[F_{q|n}(\tau|n')|n' \geq n]\big) \\ &-\frac{1-\mathbb{E}[\omega|q>\tau]}{h_q(\tau)}\mathbb{E}[f_{q|n}(\tau|n')|n' \geq n](1-F_n(n))\Big) > 0. \end{aligned}$$

[41]The behavioral earnings response can be derived in the following way. Let $d\tau'(y)$ denote the change in the marginal tax in $[y^*(n), y^*(n+dn)]$. For given $d\tau$, we have $d\tau = dy d\tau'(y) = dn\frac{dy^*}{dn}d\tau'(y) = dn(-\frac{k_{ny}(y^*(n),n)}{k_{yy}(y^*(n),n)})d\tau'(y)$ in which the last equation follows from the implicit function theorem applied to the agent's first-order condition $1-k_y(y(n),n)-T'(y(n))=0$. The change in the marginal tax in $[y^*(n), y^*(n+dn)]$ is thus $d\tau'(y^*) = -d\tau\frac{1}{dn}\frac{k_{yy}(y^*(n),n)}{k_{ny}(y^*(n),n)}$ and the associated change in earnings is $-d\tau\frac{1}{dn}\frac{k_{yy}(y^*(n),n)}{k_{ny}(y^*(n),n)}\cdot\frac{dy^*}{d\tau'(y^*)} = d\tau\frac{1}{dn}\frac{1}{k_{ny}(y^*(n),n)}$ in which the last equation uses again implicit differentiation of the first-order condition for the agent. The change in tax revenue equals the change in earnings times the marginal tax: $d\tau\frac{1}{dn}\frac{1}{k_{ny}(y^*(n),n)}T'(y^*) = -d\tau\frac{1}{dn}\frac{n}{(1+\frac{1}{\epsilon})k_y(y^*(n),n)}T'(y^*) = -d\tau\frac{1}{dn}\frac{n}{(1+\frac{1}{\epsilon})}\frac{T'(y^*)}{1-T'(y^*)}$, and when we multiply by the mass $(1-F_{q|n}(\tau|n))f_n(n)dn$ of agents affected, we obtain the formula for the effect.

For a regressive externality tax reform, the inequality is reversed. Moreover, for a constant externality tax to be optimal, the condition must be met with equality for all n. Divide both sides by $1 - F_n(n)$, $d\tau$ and $1 - \mathbb{E}[F_{q|n}(\tau, n')|n' \geq n]$ to write:

$$-(1 - \mathbb{E}[\omega|n' \geq n])\frac{1 - F_{q|n}(\tau|n)}{1 - \mathbb{E}[F_{q|n}(\tau|n')|n' \geq n]}$$
$$+(1 - \mathbb{E}[\omega|n' \geq n, q > \tau)]) - (1 - \mathbb{E}[\omega|q > \tau)])\frac{h_{q|n'\geq n}(\tau, n)}{h_q(\tau)} > 0$$

where we defined the hazard-rate for high ability types as $h_{q|n'\geq n}(\tau, n) = \frac{\mathbb{E}[f_{q|n}(\tau|n')|n'\geq n]}{1-\mathbb{E}[F_{q|n}(\tau|n')|n'\geq n]}$. Use definitions $\xi_1 = \frac{1-F_{q|n}(\tau|n)}{1-\mathbb{E}[F_{q|n}(\tau|n')|n'\geq n]}$ and $\xi_2 = \frac{h_{q|n'\geq n}(\tau,n)}{h_q(\tau)}$ and the normalization $\mathbb{E}[\omega] = 1$ to write:

$$\xi_1\mathbb{E}[\omega|n' \geq n] + \xi_2\mathbb{E}[\omega|q > \tau] - (\xi_1 + \xi_2 - 1)\mathbb{E}[\omega] - \mathbb{E}[\omega|n' \geq n, q > \tau] > 0 \tag{C.24}$$

Now use definitions $\overline{\omega} = 1$, $\overline{\omega}^0 = \mathbb{E}[\omega|q > \tau]$, $\overline{\omega}_{n'\geq n} = \mathbb{E}[\omega|n' \geq n]$, and $\overline{\omega}^0_{n'\geq n} = \mathbb{E}[\omega|n' \geq n, q > \tau]$. Restructure the terms in (C.24) to write the condition as in Proposition 3. □

C.5 *Proposition 4*

PROOF. Let $n^* \in \arg\max_{n\in[\underline{n},\overline{n}]} \tau(y^0(n))$ denote the ability level with the highest externality price τ in a tax schedule given by the formulas in Theorem 2. This implies that at $n = n^*$, we have $\tau'(y^0(n^*)) = 0$ and $\tau''(y^0(n^*)) \leq 0$.[42] This is clear if the maximum is attained in interior point $n^* \in (\underline{n}, \overline{n})$ but also true at corner points $n^* \in \{\underline{n}, \overline{n}\}$ as the marginal tax rates are zero there, see equation (C.19).

Based on Lemma C.4,

$$\Big(\overline{\omega}^0_{n^*} - \overline{\omega}^1_{n^*} - (B^\omega - \tau^n)\frac{f_{q|n}(\bar{q}|n^*)}{(1 - F_{q|n}(\bar{q}|n^*))F_{q|n}(\bar{q}|n^*)}\Big)f_n(n^*)$$
$$+\frac{\frac{\partial F_{q|n}(\bar{q}|n^*)}{\partial n}n^*h_n(n^*)}{1 + \frac{1}{\epsilon}}\frac{T'(y^1(n^*))}{1 - T'(y^1(n^*))}\frac{1}{(1 - F_{q|n}(\bar{q}|n^*))F_{q|n}(\bar{q}|n^*)} \leq 0 \tag{C.25}$$

We first prove the analogue of Proposition 1(a) for the optimal tax system of Theorem 2, so assume $n \perp\!\!\!\perp q$ or $\mathbb{E}[\omega|n]$ is invariant in n. Then the term on the left-hand side of the

[42] Theorem 2 characterizes the optimal tax schedules without bunching. In this case state variables $\widehat{v}_0(n)$ and $\widehat{v}_1(n)$ and controls y^0 and y^1 are differentiable (see the proof of Theorem 2 in Appendix A.2), from which it follows that T and τ are twice differentiable.

inequality sign on the second row of (C.25) is 0,[43] so

$$\varpi^0_{n^*} - \varpi^1_{n^*} \leq (B^\omega - \tau^n)\frac{f_{q|n}(\bar{q}|n^*)}{(1-F_{q|n}(\bar{q}|n^*))F_{q|n}(\bar{q}|n^*)} \tag{C.26}$$

where under Assumption 2a, $\varpi^0_{n^*} - \varpi^1_{n^*} > 0$, so the right-hand side of (C.26) must also be positive, and therefore $\tau^n(n^*) < B^\omega$. As $y^0(n^*) = y^1(n^*)$, $\tau^n(n^*) = \tau(y^0(n^*))$, so $\tau(y^0(n^*)) < B^\omega$. Since τ achieves its greatest value at $y^0(n^*)$, we then know that $\tau(y^0(n)) < B^\omega$ for all $n \in [\underline{n}, \overline{n}]$.

We then prove the analogue of Proposition 2(a) for the optimal tax system of Theorem 2, so assume $\mathbb{E}[\omega|n,q]$ is invariant in q. Then $\varpi^0_{n^*} - \varpi^1_{n^*} = 0$ so we can write (C.25) as

$$\begin{aligned}
&\frac{\frac{\partial F_{q|n}(\bar{q}|n^*)}{\partial n} n^* h_n(n^*)}{1+\frac{1}{\epsilon}} \frac{T'(y^1(n^*))}{1-T'(y^1(n^*))} \frac{1}{(1-F_{q|n}(\bar{q}|n^*))F_{q|n}(\bar{q}|n^*)} \\
&\qquad \leq (B^\omega - \tau^n)\frac{f_{q|n}(\bar{q}|n^*)}{(1-F_{q|n}(\bar{q}|n^*))F_{q|n}(\bar{q}|n^*)}
\end{aligned}$$

Under the assumption that $n' > n''$ implies $F_{q|n}(q|n') \geq F_{q|n}(q|n'')$ for all $n', n'' \in [\underline{n}, \overline{n}]$, we have $\frac{\partial F_{q|n}(\bar{q}|n^*)}{\partial n} \geq 0$, so the sign of the term on the left-hand side of (C.5) is non-negative.[44] For the inequality in (C.5) to hold, $\tau^n(n^*) \leq B^\omega$ must be true so consequently $\tau(y^0(n^*)) < B^\omega$. As τ achieves its greatest value at $y^0(n^*)$, we then know that $\tau(y^0(n)) \leq B^\omega$ for all $n \in [\underline{n}, \overline{n}]$.

---

[43] If $n \perp\!\!\!\perp q$ then $\frac{\partial F_{q|n}(\bar{q}|n^*)}{\partial n} = 0$, so if $\mathbb{E}[\omega|n]$ is invariant in n then, as $\tau'(y^0(n^*)) = 0$ and $y^0(n^*) = y^1(n^*)$, we have

$$\begin{aligned}
\frac{T'(y^1(n^*))}{1-T'(y^1(n^*))} &= F_{q|n}(\bar{q}|n^*)\frac{T'(y^1(n^*))}{1-T'(y^1(n^*))} + (1-F_{q|n}(\bar{q}|n^*))\frac{T'(y^0(n^*)) + \tau'(y^0(n^*))}{1-T'(y^0(n^*)) - \tau'(y^0(n^*))} \\
&= \left(1+\frac{1}{\epsilon}\right)(1-\mathbb{E}[\omega|n' \geq n])\frac{1-F_n(n)}{f_n(n)n} \\
&= 0
\end{aligned}$$

where Theorem 2 is used for the second equality and the third equality follows from $\mathbb{E}[\omega|n] = 1$ for all n.

[44] This is true since $\frac{T'(y^1(n^*))}{1-T'(y^1(n^*))} \geq 0$, as we can prove: As $\tau'(y^0(n^*)) = 0$ and $y^0(n^*) = y^1(n^*)$,

$$\begin{aligned}
\frac{T'(y^1(n^*))}{1-T'(y^1(n^*))} &= F_{q|n}(\bar{q}|n^*)\frac{T'(y^1(n^*))}{1-T'(y^1(n^*))} + (1-F_{q|n}(\bar{q}|n^*))\frac{T'(y^0(n^*)) + \tau'(y^0(n^*))}{1-T'(y^0(n^*)) - \tau'(y^0(n^*))} \\
&= \left(1+\frac{1}{\epsilon}\right)(1-\mathbb{E}[\omega|n' \geq n])\frac{1-F_n(n)}{f_n(n)n} \\
&\geq 0
\end{aligned}$$

where Theorem 2 is used for the second equality and the weak inequality follows since $\mathbb{E}[\omega|n = n']$ is by assumption non-increasing in n'.

To prove the analogues of Propositions 1(b) and 2(b) for the optimal tax system of Theorem 2, we take $n^* \in \arg\min_{n\in[\underline{n},\overline{n}]} \tau(y^0(n))$, in which case the inequalities in (C.25), (C.26) and (C.5) are reversed. Then under Assumption 2b, (C.26) implies that $\tau(y^0(n)) > B^\omega$ for all $n \in [\underline{n},\overline{n}]$. Under the assumption that $n' > n''$ implies $F_{q|n}(q|n') \leq F_{q|n}(q|n'')$ for all $n', n'' \in [\underline{n},\overline{n}]$, (C.5) implies that $\tau(y^0(n)) \geq B^\omega$ for all $n \in [\underline{n},\overline{n}]$.

□

C.6 *Proposition 5*

PROOF. To prove part (a) of the proposition, suppose that for some $n \in [\underline{n},\overline{n}]$, $\tau'(y^0(n)) \geq 0$ but by contradiction $T'(y^1(n)) > \widehat{T}'(y(n))$. By Lemma C.3,

$$\frac{T'(y^0(n)) + \tau'(y^0(n))}{1 - T'(y^0(n)) - \tau'(y^0(n))} \geq \frac{T'(y^1(n))}{1 - T'(y^1(n))} > \frac{\widehat{T}'(y(n))}{1 - \widehat{T}'(y(n))}$$

and hence

$$\begin{aligned}(1 - F_{q|n}(\bar{q}(n)|n))\frac{T'(y^0(n)) + \tau'(y^0(n))}{1 - T'(y^0(n)) - \tau'(y^0(n))} + F_{q|n}(\bar{q}(n)|n))\frac{T'(y^1(n))}{1 - T'(y^1(n))} \\ > \frac{\widehat{T}'(y(n))}{1 - \widehat{T}'(y(n))}\end{aligned} \tag{C.27}$$

Use the optimal tax formulas from Theorem 2 to write inequality (C.27) as

$$\left(1 + \frac{1}{\epsilon}\right)\frac{1 - \mathbb{E}[\omega|n' \geq n]}{nh_n(n)} > \frac{\widehat{T}'(y(n))}{1 - \widehat{T}'(y(n))} \tag{C.28}$$

But since $\widehat{T}$ is the optimal stand-alone income tax schedule, by Theorem 1 we know

$$\left(1 + \frac{1}{\epsilon}\right)\frac{1 - \mathbb{E}[\omega|n' \geq n]}{nh_n(n)} = \frac{\widehat{T}'(y(n))}{1 - \widehat{T}'(y(n))}$$

so we obtained a contradiction.

Part (b) can be proved by reversing the inequalities. □

C.7 *Proposition 6*

PROOF. Consider an optimal tax system given by formulas (14) and (15) in Theorem 2 and suppose that Assumptions 3-4 hold and $n \perp\!\!\!\perp q$. We will first prove part (a) of the proposition. To do that, suppose that $\varpi_n^0(\tilde{q}) - \varpi_n^1(\tilde{q})$ is strictly decreasing in n for all $n \in [\underline{n},\overline{n}]$, $\tilde{q} \in \mathbb{R}_+$ but by contradiction it is not true that $\tau'(y) \geq 0$ for all y. Then there must be an interval (n_a, n_b) such that $\tau'(y) < 0$ for all $y \in (y^0(n_a), y^0(n_b))$ and $\tau'(y^0(n)) = 0$ at $n \in \{n_a, n_b\}$, so that $\tau''(y^0(n_a)) \leq 0$ and $\tau''(y^0(n_b)) \geq 0$.[45] Then, by Lemma C.4 (and since

[45] When it is not true that $\tau'(y) \geq 0$ for all y, there must be a non-degenerate interval where $\tau'(y) < 0$ since τ' is continuous; in fact $\tau(y)$ is twice differentiable, see footnote 42. An interval such that $\tau'(y) = 0$ at the end points always exists since $\tau'(y) = 0$ for $y \in \{y^0(\underline{n}), y^0(\overline{n})\}$, see equation (C.19).

$n \perp\!\!\!\perp q$),

$$\overline{\omega}_n^0 - \overline{\omega}_n^1 - (B^\omega - \tau^n)\frac{f_{q|n}(\bar{q}|n)}{(1-F_{q|n}(\bar{q}|n))F_{q|n}(\bar{q}|n)}$$

is non-positive at $n = n_a$ and non-negative at $n = n_b$. This implies, using notation $\tau_a := \tau(y^0(n_a))$ and $\tau_b := \tau(y^0(n_b))$ and assumption $n \perp\!\!\!\perp q$:

$$\overline{\omega}_{n_a}^0(\tau_a) - \overline{\omega}_{n_a}^1(\tau_a) \leq \left(B^\omega - \tau_a\right)\frac{f_q(\tau_a)}{(1-F_q(\tau_a))F_q(\tau_a)} \tag{C.29}$$

$$\overline{\omega}_{n_b}^0(\tau_b) - \overline{\omega}_{n_b}^1(\tau_b) \geq \left(B^\omega - \tau_b\right)\frac{f_q(\tau_b)}{(1-F_q(\tau_b))F_q(\tau_b)} \tag{C.30}$$

Notice that

$$\begin{aligned}\overline{\omega}_{n_a}^0(\tau_a) - \overline{\omega}_{n_a}^1(\tau_a) >& \overline{\omega}_{n_b}^0(\tau_a) - \overline{\omega}_{n_b}^1(\tau_a)\\ \geq& \overline{\omega}_{n_b}^0(\tau_b) - \overline{\omega}_{n_b}^1(\tau_b)\end{aligned} \tag{C.31}$$

where the first inequality follows from the assumption that $\overline{\omega}_n^0(\tilde{q}) - \overline{\omega}_n^1(\tilde{q})$ is strictly decreasing in n and the second inequality follows from Assumption 4 and $\tau_a > \tau_b$.

Combine (C.29), (C.30) and (C.31) to obtain

$$\left(B^\omega - \tau_a\right)\frac{f_q(\tau_a)}{(1-F_q(\tau_a))F_q(\tau_a)} > \left(B^\omega - \tau_b\right)\frac{f_q(\tau_b)}{(1-F_q(\tau_b))F_q(\tau_b)}$$

This implies together with Assumption 3 that $\tau_a < \tau_b$ which is a contradiction. This proves part (a) of the proposition.

Part (b) of the proposition is proved analogously, by reversing the inequalities.

□

C.8 *Proposition 7*

PROOF. Consider an optimal tax system given by formulas (14) and (15) in Theorem 2. Suppose that Assumptions 3-4 hold and that $\overline{\omega}_n^0(\tilde{q}) - \overline{\omega}_n^1(\tilde{q})$ is invariant in n for all $n \in [\underline{n}, \overline{n}]$, $\tilde{q} \in \mathbb{R}_+$.

We will first show that if assumptions (a)-(b) or (a')-(b') in the proposition hold then $\tau'(y) > 0$ in some interval. To do that, suppose that (a)-(b) or (a')-(b') hold but by contradiction $\tau' \leq 0$ everywhere. This implies that $\underline{\tau} \geq \overline{\tau}$ where we use notation $\underline{\tau} := \tau(y^0(\underline{n}))$ and $\overline{\tau} := \tau(y^0(\overline{n}))$. By equation (C.19), $\tau'(y^0(n)) = 0$ at $n = \underline{n}$ and $n = \overline{n}$ and $\tau''(y^0(\underline{n})) \leq 0$ and $\tau''(y^0(\overline{n})) \geq 0$.[46] By Lemma C.4 and since $T'(y^0(n)) = 0$ at $n = \underline{n}$ and $n = \overline{n}$, we have

$$\overline{\omega}_{\underline{n}}^0(\underline{\tau}) - \overline{\omega}_{\underline{n}}^1(\underline{\tau}) \leq \eta(\underline{n}, \underline{\tau})(B^\omega - \underline{\tau}) \tag{C.32}$$

$$\overline{\omega}_{\overline{n}}^0(\overline{\tau}) - \overline{\omega}_{\overline{n}}^1(\overline{\tau}) \geq \eta(\overline{n}, \overline{\tau})(B^\omega - \overline{\tau}). \tag{C.33}$$

[46] Recall that τ is twice differentiable, see footnote 42.

By assumption, $\overline{\omega}_n^0(\tilde{q}) - \overline{\omega}_n^1(\tilde{q})$ is invariant in n and non-decreasing in $\tilde{q}$ (Assumption 4), so we have

$$\overline{\omega}_{\underline{n}}^0(\underline{\tau}) - \overline{\omega}_{\underline{n}}^1(\underline{\tau}) = \overline{\omega}_{\overline{n}}^0(\underline{\tau}) - \overline{\omega}_{\overline{n}}^1(\underline{\tau}) \geq \overline{\omega}_{\overline{n}}^0(\overline{\tau}) - \overline{\omega}_{\overline{n}}^1(\overline{\tau})$$

Use this with inequalities (C.32) and (C.33) to obtain

$$\eta(\overline{n}, \overline{\tau})(B^\omega - \overline{\tau}) \leq \eta(\underline{n}, \underline{\tau})(B^\omega - \underline{\tau}). \tag{C.34}$$

Assumption 3 together with $\underline{\tau} \geq \overline{\tau}$ implies

$$\eta(\overline{n}, \overline{\tau})(B^\omega - \overline{\tau}) \geq \eta(\overline{n}, \underline{\tau})(B^\omega - \underline{\tau}) \tag{C.35}$$

$$\eta(\underline{n}, \underline{\tau})(B^\omega - \underline{\tau}) \leq \eta(\underline{n}, \overline{\tau})(B^\omega - \overline{\tau}). \tag{C.36}$$

Then, combining (C.34) and (C.35) gives

$$\eta(\overline{n}, \underline{\tau})(B^\omega - \underline{\tau}) \leq \eta(\underline{n}, \underline{\tau})(B^\omega - \underline{\tau}) \tag{C.37}$$

and combining (C.34) and (C.36) gives

$$\eta(\underline{n}, \overline{\tau})(B^\omega - \overline{\tau}) \geq \eta(\overline{n}, \overline{\tau})(B^\omega - \overline{\tau}) \tag{C.38}$$

If assumption (a) (i.e. horizontal equity Assumption 2a) holds then $\overline{\omega}_{\underline{n}}^0(\underline{\tau}) - \overline{\omega}_{\underline{n}}^1(\underline{\tau}) > 0$, which together with (C.32) implies $B^\omega - \underline{\tau} > 0$, which together with (C.37) implies

$$\eta(\overline{n}, \underline{\tau}) \leq \eta(\underline{n}, \underline{\tau}) \tag{C.39}$$

so assumption (b) (that $\eta(n, q)$ is strictly increasing in n) cannot hold simultaneously with (a).

If assumption (a') (i.e. horizontal equity Assumption 2b) holds then $\overline{\omega}_{\overline{n}}^0(\overline{\tau}) - \overline{\omega}_{\overline{n}}^1(\overline{\tau}) < 0$, which together with (C.33) implies $B^\omega - \overline{\tau} < 0$, which together with (C.38) implies

$$\eta(\underline{n}, \overline{\tau}) \leq \eta(\overline{n}, \overline{\tau}) \tag{C.40}$$

so assumption (b') (that $\eta(n, q)$ is strictly decreasing in n) cannot hold simultaneously with (a').

Hence we obtained a contradiction, proving that if (a)-(b) or (a')-(b'), then it holds for the optimal tax that $\tau'(y) > 0$ in some interval.

Analogously, we can prove that if (a')-(b) or (a)-(b'), then $\tau'(y) < 0$ in some interval.

□

C.9 *Lemma C.1*

LEMMA C.1. *For any budget feasible, incentive-compatible mechanism $M = (x, y, t)(n, q, b)$, there is another budget feasible, incentive compatible mechanism $\widehat{M} = (\widehat{x}, \widehat{y}, \widehat{t})(n, q)$ that does not depend on b and gives at least as high social welfare as original mechanism M.*

PROOF. Take any direct, budget-balanced incentive-compatible mechanism $M := (x, y, t)(\theta)$ that may condition the allocation on the whole type (n, q, b). By incentive compatibility

$$u(\theta; y(\theta), x(\theta), t(\theta), \overline{x}) = \max_{\theta'} u(\theta; y(\theta'), x(\theta'), t(\theta'), \overline{x})$$

$$\implies v(n, q; y(\theta), x(\theta), t(\theta)) = \max_{\theta'} v(n, q; y(\theta'), x(\theta'), t(\theta'))$$

(where $v(n, q; y, x, t) = u(n, q, b; y, x, t, \overline{x}) - b\overline{x}$ denotes again the private part of the utility) so value $v(n, q; (y, x, t)(n, q, b))$ is invariant with respect to b although allocation $(y, x, t)(n, q, b)$ could depend on b in mechanism M.

Now given mechanism $M = (x, y, t)(\theta)$ choose for every (n, q) some $\widehat{b}(n, q)$ so that $x(n, q, \widehat{b}(n, q))B^{\omega} + t(n, q, \widehat{b}(n, q)) \geq \mathbb{E}[x(n, q, b)B^{\omega} + t(n, q, b)|n, q]$. Define also $\overline{t} = \mathbb{E}_{\theta}[t(n, q, \widehat{b}(n, q)) - t(n, q, b)]$. Now we are ready to propose a new mechanism that does not condition on benefit b. The new mechanism is defined as $\widehat{M} = (\widehat{x}, \widehat{y}, \widehat{t})(n, q) = (x(n, q, \widehat{b}(n, q)), y(n, q, \widehat{b}(n, q)), t(n, q, \widehat{b}(n, q)) + \overline{t})$. In words, mechanism $\widehat{M}$ gives each (n, q) the same allocation and transfer it gives to $(n, q, \widehat{b}(n, q))$ in mechanism M and then redistributes any excess revenue (positive or negative) equally among all agents.

By incentive compatibility of mechanism M,

$$v(n, q; (y, x, t)(n, q, \widehat{b}(n, q)) \in \arg\max_{n', q'} v(n, q; (y, x, t)(n', q', \widehat{b}(n', q'))$$

for all n, q, which implies

$$v(n, q; (\widehat{y}, \widehat{x}, \widehat{t})(n, q, b)) \in \arg\max_{n', q', b'} v(n, q; (\widehat{y}, \widehat{x}, \widehat{t})(n', q', b'))$$

for all n, q, b, so mechanism $\widehat{M}$ is incentive compatible. Budget balance of mechanism $\widehat{M}$ also follows straight-forwardly from budget balance of mechanism M.

Now denote the social welfare resulting from mechanism M by $W(M)$, then the social welfare resulting from mechanism $\widehat{M}$ is

$$\begin{aligned} W(\widehat{M}) &= W(M) + \overline{t} + B^{\omega}\mathbb{E}_{\theta}[\widehat{x}(n, q) - x(n, q, b)] \\ &= W(M) + \mathbb{E}_{\theta}[t(n, q, \widehat{b}(n, q)) + B^{\omega}x(n, q, \widehat{b}(n, q)) - t(n, q, b) - B^{\omega}x(n, q, b)] \end{aligned}$$

The expected value on the second line is non-negative since by the definition of $\widehat{b}$, $t(n, q, \widehat{b}(n, q)) + B^{\omega}x(n, q, \widehat{b}(n, q)) \geq \mathbb{E}[x(n, q, b)B^{\omega} + t(n, q, b)|n, q]$ for all (n, q). Hence, mechanism $\widehat{M}$ results in welfare that is at least equal with welfare resulting from mechanism M. □

C.10 *Lemma C.2*

LEMMA C.2. *There exists a solution to the control problem defined by objective (C.1) and constraints (C.2)-(C.5).*

PROOF. The existence follows by Filippov-Cesari Theorem (e.g., Theorem 8, page 132, Seierstad and Sydsæter, 1987). This type of theorem requires that the control set is well defined. We can take $y^i \in [0, Y]$, $i = 0, 1$, for finite bound $Y > 0$, defined by $1 - k_y(Y, \overline{n}) = 0$, for example. The theorem also requires that functions a_0, a_1, a_2 are sufficiently regular. In our case, they are continuously differentiable and bounded, which guarantees the required regularity. The final requirement is that a set defined by

$$N(\widehat{v}, n, \lambda, \gamma) = \{(a_0 + \lambda a_1 + \gamma a_2 + \kappa, -k_n(y^0, n), -k_n(y^1, n)) : \kappa \leq 0, y \in [0, Y] \times [0, Y]\} \tag{C.41}$$

is convex in which $\widehat{v} = (\widehat{v}^0, \widehat{v}^1)$, $y = (y^0, y^1)$. Thus, N holds the states $\widehat{v}$, type n, λ, and γ as given and defines the set by variations in controls and scalar κ. Note that in $a_0 + \lambda a_1 + \gamma a_2$ only a_1 depends on y. We rewrite the set in terms of variables

$$z = (z^0, z^1) = (-k_n(y^0, n), -k_n(y^1, n)) = \left(\frac{1}{n}\left(\frac{y^0}{n}\right)^{\frac{1+\epsilon}{\epsilon}}, \frac{1}{n}\left(\frac{y^1}{n}\right)^{\frac{1+\epsilon}{\epsilon}}\right)$$

instead of (y^0, y^1) to obtain

$$\begin{aligned} &N(\widehat{v}, n, \lambda, \gamma) \\ &= \left\{(a_0 + \lambda \widehat{a}_1 + \gamma a_2 + \kappa, z^0, z^1) : \kappa \leq 0, z \in \left[0, \frac{1}{n}\left(\frac{Y}{n}\right)^{\frac{1+\epsilon}{\epsilon}}\right] \times \left[0, \frac{1}{n}\left(\frac{Y}{n}\right)^{\frac{1+\epsilon}{\epsilon}}\right]\right\} \end{aligned} \tag{C.42}$$

where

$$\begin{aligned} \widehat{a}_1 = \Bigg(&\left(n^{\frac{1+2\epsilon}{1+\epsilon}}(z^0)^{\frac{\epsilon}{1+\epsilon}} - n\frac{\epsilon}{1+\epsilon}z^0 - \widehat{v}^0(n)\right)(1 - F_{q|n}(\widehat{v}^1(n) - \widehat{v}^0(n)|n)) \\ &+ \left(n^{\frac{1+2\epsilon}{1+\epsilon}}(z^1)^{\frac{\epsilon}{1+\epsilon}} - n\frac{\epsilon}{1+\epsilon}z^1 - \widehat{v}^1(n)\right)F_{q|n}(\widehat{v}^1(n) - \widehat{v}^0(n)|n)\Bigg) f_n(n) \end{aligned} \tag{C.43}$$

Since $a_0 + \lambda\widehat{a}_1 + \gamma a_2 + \kappa$ is a concave function of (z^0, z^1) for all $z^0, z^1 \geq 0$, we can easily see from (C.10) that set N is convex. □

C.11 *Lemma C.3*

LEMMA C.3. *Consider a tax system given by formulas (14) and (15) in Theorem 2. Take any* $n \in [\underline{n}, \overline{n}]$. *Then*

$$\tau'(y^0(n)) \geq 0 \implies \frac{T'(y^0(n)) + \tau'(y^0(n))}{1 - T'(y^0(n)) - \tau'(y^0(n))} - \frac{T'(y^1(n))}{1 - T'(y^1(n))} \geq 0$$

$$\tau'(y^0(n)) \leq 0 \implies \frac{T'(y^0(n)) + \tau'(y^0(n))}{1 - T'(y^0(n)) - \tau'(y^0(n))} - \frac{T'(y^1(n))}{1 - T'(y^1(n))} \leq 0$$

PROOF. Take any $n \in [\underline{n}, \overline{n}]$ such that $\tau'(y^0(n)) \geq 0$. As $y^0(\underline{n}) = y^1(\underline{n})$ and $y^0(\overline{n}) = y^1(\overline{n})$ and y^0 and y^1 are continuous and non-decreasing functions on $[\underline{n}, \overline{n}]$, there must exist $n' \in [\underline{n}, \overline{n}]$ such that $y^1(n') = y^0(n)$. Choose n' so that $y^1(n') = y^0(n)$.

As $\tau'(y^0(n)) \geq 0$, we have $T'(y^0(n)) + \tau'(y^0(n)) \geq T'(y^0(n))$, implying

$$T'(y^0(n)) + \tau'(y^0(n)) \geq T'(y^1(n')) \tag{C.44}$$

The agents' optimality conditions give

$$1 - k_y(y^0(n), n) - T'(y^0(n)) - \tau'(y^0(n)) = 0 \tag{C.45}$$

$$1 - k_y(y^1(n'), n') - T'(y^1(n')) = 0 \tag{C.46}$$

Use (C.44), (C.45), (C.46) and $y^1(n') = y^0(n)$ to obtain

$$k_y(y^0(n), n) \leq k_y(y^0(n), n') \tag{C.47}$$

Since $k_{ny}(y^0(n''), n'') < 0$ for all $n'' \in [\underline{n}, \overline{n}]$, inequality (C.47) implies $n' < n$. Since y^1 is increasing, then $y^1(n') \leq y^1(n)$, so $y^0(n) \leq y^1(n)$. This implies (again using an agent's optimality condition) $T'(y^0(n)) + \tau'(y^0(n)) \geq T'(y^1(n))$. Since $0 \leq T'(y^0(n)) + \tau'(y^0(n)) < 1$ and $0 \leq T'(y^1(n)) < 1$, this implies

$$\tau'(y^0(n)) \geq 0 \implies \frac{T'(y^0(n)) + \tau'(y^0(n))}{1 - T'(y^0(n)) - \tau'(y^0(n))} - \frac{T'(y^1(n))}{1 - T'(y^1(n))} \geq 0$$

The second part of the lemma can be proved analogously.

□

C.12 *Lemma C.4*

LEMMA C.4. *Consider a tax system given by formulas (14) and (15) in Theorem 2. If $\tau'(y^0(n^*)) = 0$ and $\tau''(y^0(n^*)) \leq 0$, then*

$$\begin{aligned}&\left(\overline{\omega}^0_{n^*} - \overline{\omega}^1_{n^*} - (B^\omega - \tau^n)\frac{f_{q|n}(\bar{q}|n^*)}{(1 - F_{q|n}(\bar{q}|n^*))F_{q|n}(\bar{q}|n^*)}\right) f_n(n^*)\\ &+ \frac{\dfrac{\partial F_{q|n}(\bar{q}|n^*)}{\partial n} n^* h_n(n^*)}{1 + \dfrac{1}{\epsilon}} \frac{T'(y^1(n^*))}{1 - T'(y^1(n^*))} \frac{1}{(1 - F_{q|n}(\bar{q}|n^*))F_{q|n}(\bar{q}|n^*)} \leq 0\end{aligned} \tag{C.48}$$

If $\tau'(y^0(n^*)) = 0$ and $\tau''(y^0(n^*)) \geq 0$, then the inequality sign in (C.48) is reversed.

PROOF. Consider a tax system given by formulas (14) and (15) in Theorem 2. Suppose $\tau'(y^0(n^*)) = 0$. We will first prove that

We will first prove that the derivative of

$$\frac{T'(y^0(n)) + \tau'(y^0(n))}{1 - T'(y^0(n)) - \tau'(y^0(n))} - \frac{T'(y^1(n))}{1 - T'(y^1(n))} \tag{C.49}$$

with respect to n at $n = n^*$ is non-positive if $\tau''(y^0(n^*)) \leq 0$ and non-negative if $\tau''(y^0(n^*)) \geq 0$.

To do that, note that (C.49) is equal to

$$\frac{1 - k_y(y^0(n), n)}{k_y(y^0(n))} - \frac{1 - k_y(y^1(n), n)}{k_y(y^1(n))} \tag{C.50}$$

for all n by equations (9) and (10).

Differentiate (C.50) with respect to n to obtain

$$\begin{aligned}
&\Big[- k_{yy}(y^0(n), n)\frac{\partial y^0}{\partial n} - k_{yn}(y^0(n), n)\Big) k_y(y^0(n)) \\
&- \Big(k_{yy}(y^0(n))\frac{\partial y^0}{\partial n} + k_{yn}(y^0(n), n)\Big)(1 - k_y(y^0(n), n))\Big] \frac{1}{k_y(y^0(n))^2} \\
&- \\
&\Big[\Big(- k_{yy}(y^1(n), n)\frac{\partial y^1}{\partial n} - k_{yn}(y^1(n), n)\Big) k_y(y^1(n)) \\
&- \Big(k_{yy}(y^1(n))\frac{\partial y^1}{\partial n} + k_{yn}(y^1(n), n)\Big)(1 - k_y(y^1(n), n))\Big] \frac{1}{k_y(y^1(n))^2}
\end{aligned} \tag{C.51}$$

Since $y^0(n^*) = y^1(n^*)$, expression (C.12) evaluated at n^* is equal to

$$\frac{\Big(\frac{\partial y^1}{\partial n} - \frac{\partial y^0}{\partial n}\Big) k_{yy}(y^0(n^*), n^*)}{k_y(y^0(n^*))^2}. \tag{C.52}$$

By differentiating agents' optimality conditions (9) and (10) with respect to n we can show that

$$\frac{\partial y^1}{\partial n} = \frac{-k_{yn}(y^1(n), n))}{k_{yy}(y^1(n)) + T''(y^1(n))} \tag{C.53}$$

$$\frac{\partial y^0}{\partial n} = \frac{-k_{yn}(y^0(n), n))}{k_{yy}(y^0(n)) + T''(y^0(n)) + \tau''(y^0(n))} \tag{C.54}$$

Since $\frac{\partial y^i}{\partial n} \geq 0$ for $i \in \{0, 1\}$ and $k_{yn} < 0$, the denominators of (C.53) and (C.54) are positive, and then for $n = n^*$, we obtain

$$\tau''(y^0(n^*)) \leq 0 \implies \frac{\partial y^1}{\partial n} - \frac{\partial y^0}{\partial n} \leq 0 \text{ at } n = n^*,$$

$$\tau''(y^0(n^*)) \geq 0 \implies \frac{\partial y^1}{\partial n} - \frac{\partial y^0}{\partial n} \geq 0 \text{ at } n = n^*.$$

since $y^0(n^*) = y^1(n^*)$,

Then, since $k_{yy} > 0$, (C.52) is non-positive if $\tau''(y^0(n^*)) \leq 0$ and non-negative if $\tau''(y^0(n^*)) \geq 0$.

Using the optimal tax formulas of Theorem 2, this implies that the derivative of

$$\left(1+\frac{1}{\epsilon}\right)\frac{\mathbb{E}[(1-\overline{\omega}^0_{n'})(1-F_{q|n}(\bar{q}|n'))+(B^\omega-\tau^n)f_{q|n}(\bar{q}|n')|n'\geq n]}{(1-F_{q|n}(\bar{q}|n))} \\ -\left(1+\frac{1}{\epsilon}\right)\frac{\mathbb{E}[(1-\overline{\omega}^1_{n'})F_{q|n}(\bar{q}|n')-\big(B^\omega-\tau^n\big)f_{q|n}(\bar{q}|n')|n'\geq n]}{F_{q|n}(\bar{q}|n)} \tag{C.55}$$

with respect to n is non-positive at $n=n^*$ if $\tau''(y^0(n^*))\leq 0$ and non-negative if $\tau''(y^0(n^*))\geq 0$, while the value of (C.55) at $n=n^*$ is 0. Differentiate (C.55) with respect to n at $n=n*$, using $\bar{q}'(n^*)=\widehat{v}^{1\prime}(n*)-\widehat{v}^{0\prime}(n*)=0$, to obtain the result of the lemma.

□

C.13 *Extensions and robustness analysis*

C.13.1 *Continuous pollution choices* In some applications, pollution reduction is an intensive-margin rather than extensive-margin choice. We consider now a continuous choice $x\in[0,1]$, reflecting the share of emissions abated, and denote by $e=1-x$ the pollution created by the individual. The cost of abating share x is $s(x,q)$ with $s_x,s_{xx},s_{xq}>0,\lim_{x\to 0}s_x(x,q)=0,\lim_{x\to 1}s_x(x,q)=\infty$. Given type (n,q,b), choices (y,x), transfer t, and aggregate externality $\overline{x}$, the agent's payoff is then

$$y-k(y,n)-s(x,q)+b\overline{x}-t.$$

The results of Theorem 1 can be generalized to cover continuous pollution decisions. Consider a stand-alone income tax and a constant externality tax in emissions, $t(y,x)=T(y)+\tau e$, and the impacts of a marginal increase in τ. The resulting increase in emission reductions creates surplus $(B^\omega-\tau)\mathbb{E}\left[\frac{\partial x}{\partial\tau}\right]$, while the mechanical redistributive effect is $\mathbb{E}[(1-\omega)e]$. For an optimal finite $\tau>0$, the sum of these effects is zero, which gives us the formula

$$\tau=B^\omega+\frac{1-\overline{\omega}^0_e}{\overline{h}_q} \tag{C.56}$$

where $\overline{\omega}^0_e=\frac{\mathbb{E}[\omega e]}{\mathbb{E}[e]}$ denotes the emission-weighted welfare weights and $\overline{h}_q=-\frac{\mathbb{E}[\partial e/\partial\tau]}{\mathbb{E}[e]}=\frac{\mathbb{E}[\partial x/\partial\tau]}{\mathbb{E}[e]}$ is the semi-elasticity of emissions with respect to externality prices. Equation (C.56) shows that Theorem 1 generalizes to continuous choice – we only replace, first, the welfare weight of polluters $\mathbb{E}[\omega|q>\tau]$ with the emissions-weighted welfare-weight of all individuals, $\overline{\omega}^0_e$, and, second, hazard rate $h_q(\tau)$ with the semi-elasticity of emission reduction $\overline{h}_q(\tau)$ which aggregates responses over all types.

We can also extend the results of the simple reform in Section 4.1. The reform increases the externality tax by $d\tau$ for those with labor ability greater than n by increasing $\tau'(y)$ in band $y\in[y(n),y(n+dn)]$. The reform has three effects. First, there is an earnings response at n, which amounts to $-d\tau\frac{n}{1+\frac{1}{\epsilon}}\frac{T'(y(n))}{1-T'(y(n))}\mathbb{E}[e|n]f_n(n)$. Second, the transfer creates a distributional effect $d\tau\mathbb{E}[\big(1-\omega\big)e|n'\geq n](1-F_n(n))$ by raising more tax revenue from high-polluting high-income individuals. Third, emission reduction creates benefits $d\tau\big(B^\omega-\tau\big)\mathbb{E}\Big[\frac{\partial x}{\partial\tau}|n'\geq n\Big](1-F_n(n))$. The small tax reform is welfare-increasing if the sum of the three terms is positive:

$$\Big(1+\frac{1}{\epsilon}\Big)\frac{\mathbb{E}\Big[(1-\omega)e+(B^\omega-\tau)\dfrac{\partial x}{\partial\tau}|n'\geq n\Big]}{nh_n(n)}>\frac{T'(y(n))}{1-T'(y(n))}\mathbb{E}\Big[e|n\Big]$$

If we evaluate the expression at the optimal constant τ, given by (C.56), and use the optimal stand-alone income tax, still given by (12), we arrive at the condition in Proposition 3, with the following changes. First, the polluters' welfare weights $\overline{\omega}^0$ and $\overline{\omega}^0_{n'\geq n}$ are replaced by emissions-weighted welfare weights. Second, the average emissions of type n relative to the average emissions of those above n, $\frac{\mathbb{E}[e|n]}{\mathbb{E}[e|n'\geq n]}$, replaces the share of polluters at n relative to the share of polluters above n. Third, semi-elasticities of average emissions $\overline{h}_q$ and $\overline{h}_{q|n'\geq n}(\tau,n)\equiv-\mathbb{E}\big[\frac{\partial e}{\partial\tau}|n'\geq n\big]/\mathbb{E}\big[e|n'\geq n\big]$ replace the semi-elasticities of polluters' masses. With these changes of definitions, the results on the optimality of progressive or regressive reforms continue to hold.[47]

C.13.2 *Heterogenous emissions* We can extend the model to cover heterogeneous emissions by individuals if we assume each agent chooses $x\in\{0,1\}$ but pollution emissions differ by income and earning ability according to function $e(y,n)$. We consider only the optimal separated tax system $T(\cdot)=T(y)+\tau e(y,n)$. A marginal increase of $d\tau$ in the externality tax results again in three effects (behavioral, redistributive, environmental). We develop next the formulas for the three effects and the optimal linear externality tax that balances these effects.

The behavioral labor market effect through the change in earnings reads in total as

$$d\tau\mathbb{E}\big[\frac{dy}{d\tau}\big(T'(y^0)+e_y(y^0,n)(\tau-B^\omega)\big)(1-F_{q|n}(\overline{q}|n))\big]$$

in which $\frac{dy}{d\tau}$ is determined by e_y. If emissions increase in income, raising externality tax effectively makes the tax schedule of polluters more progressive, worsening polluters' labor market incentives.

There is a mechanical redistributive effect: making polluters pay more taxes redistributes from them to everyone, and this effect is

$$d\tau\mathbb{E}[e(y,n)(1-\overline{\omega}^0_n)(1-F_{q|n}(\overline{q}|n))].$$

There is a direct environmental gain: for polluters at each ability level n, the externality price increases by $d\tau e(y(n),n)$ and hence incentivizes mass $d\tau e(y(n),n)f(\overline{q}|n)$ to cut emissions, affecting social welfare through decreased emissions taxes and total externality. Aggregated over all ability levels, this effect is

$$d\tau\mathbb{E}[f_{q|n}(\overline{q}|n)e(y,n)](B^\omega-\tau).$$

[47]However, extension to the general two-dimensional mechanism with continuous choices is not straightforward. Unlike in in our discrete-choice model, it is difficult to know what incentive compatibility constraints bind, see Rochet and Stole (2003). A fully optimal mechanism is outside the scope of this paper, but we conjecture that it also involves $T(y,e)$ that is nonlinear in e to screen agents in their q-dimension. For recent advances in multidimensional mechanism design and their application to income taxation, see Spiritus *et al.* (2022) and Golosov and Krasikov (2023).

The optimal linear externality tax then balances the three effects and can be written as

$$\tau = \frac{\mathbb{E}[f_{q|n}(\overline{q}|n)e(y,n)]B^{\omega} + \mathbb{E}[\big(e(y,n)(1-\overline{\omega}_n^0) + \frac{dy}{d\tau}\big(T'(y) - e_y(y,n)B^{\omega}\big)\big)(1-F_{q|n}(\widehat{q}|n))]}{\mathbb{E}[f_{q|n}(\overline{q}|n)e(y,n) - \frac{dy}{d\tau}e_y(y,n)(1-F_{q|n}(\overline{q}|n)]} \tag{C.57}$$

and if $e_y = 0$, this simplifies to

$$\tau = B^{\omega} + \frac{\widehat{e}}{\overline{e}}\frac{1-\widehat{\omega}^0}{\widehat{h}_q}$$

where $\overline{e}$ denotes the average emissions of polluters, $\widehat{\omega}^0 \equiv \frac{\mathbb{E}[\overline{\omega}_n^0 e(y,n)(1-F_{q|n}(\overline{q}|n)]}{\mathbb{E}[e(y,n)(1-F_{q|n}(\overline{q}|n))]}$ is polluters' emissions-weighted welfare weight, $\widehat{e} \equiv \frac{\mathbb{E}[f_{q|n}(\overline{q}|n)e(y,n)]}{\mathbb{E}[f_{q|n}(\overline{q}|n)]}$ denotes the average emissions at the cutting margin and $\widehat{h}_q \equiv \frac{\mathbb{E}[f_{q|n}(\overline{q}|n)]}{\mathbb{E}[1-F_{q|n}(\overline{q}|n)]}$ is the hazard rate of the cost distribution. This formula indicates that the externality tax is pushed downwards from the welfare-weighted Pigouvian level B^{ω} if the polluters' *emissions-weighted* welfare weight exceeds the average welfare weight in the population, and more so if the externality action is not very responsive and the individuals at the margin of making the externality action do not generate much emissions. Furthermore, if we drop the assumption $e_y = 0$, equation (C.13.2) shows that new terms emerge: if $e_y > 0$, then increasing the externality tax distorts labor market incentives.

C.13.3 *Alternative social welfare specifications* Consider a utilitarian social welfare function

$$\mathbb{E}[\Omega(u(\theta))] \tag{1''}$$

in which Ω is an increasing and concave function of agent's utility (1). For a given policy, the average social marginal welfare weight of individuals with ability n and cost q is $\mathbb{E}\big[\omega|n,q\big] = \mathbb{E}\big[\Omega'(u(\theta))/\lambda\big|n,q\big]$ where λ is the shadow price of the budget constraint. Suppose further that b is invariant across individuals and that we generate the weight for each individual from Ω. Under the classical assumption that the policy maker has aversion to utility differences ($\Omega'' < 0$), the weights generated by Ω would satisfy the vertical equality concern of Assumption 1 and the horizontal equality concern of Assumption 2a: polluters would always receive a higher weight than non-polluters under these assumptions.

Assumption 4 in the main analysis defines the difference in weighing polluters *vis-à-vis* non-polluters as $\overline{\omega}_n^0(\tilde{q}) - \overline{\omega}_n^1(\tilde{q})$ for a given cut-off $\tilde{q}$ for emission reduction costs. Consider a linear externality tax at $\tau = \tilde{q}$ and assume independently distributed n and q. Does the within-income-group concern for horizontal inequality increase or decrease in n under this tax policy? The answer depends on the third derivative Ω''', often associated with "prudence" of a decision-maker. If $\Omega''' > 0$, the horizontal concern decreases in n, while if $\Omega''' < 0$, the horizontal concern increases in n.[48] For the social welfare representation, the third derivative is therefore an important determinant of whether externality taxes should be progressive or regressive (Proposition 6). For the constant relative risk aversion (CRRA) welfare function $\Omega(u) = \frac{u^{1-\eta}}{1-\eta}$, we have $\Omega''' > 0$. This holds also for a welfare function that is approximately Rawlsian in the CRRA class corresponding to a limit in which inequality aversion becomes very large ($\eta \to \infty$). In sum, we observe that welfare weights consistent with a social welfare function can capture relatively broad preferences.

In our baseline model, q can be interpreted as a difference between benefits from consumption generating and not generating the externality. Alternatively, these benefits (denote by v_1 and v_0, respectively) could be explicit parts of the individual's type and we could write the payoff function, for example, as $y - k(y,n) + v_0(1-x) + v_1 x + \overline{x} - t$. However, we can denote $q \equiv v_0 - v_1$ to write the payoff as $y - k(y,n) + v_0 - qx + b\overline{x} - t$ and then eliminate additive term v_0 to end up with payoff formula (1) as the term does not interact with allocation and hence the elimination does not affect the optimal policy given our baseline social welfare specification. However, this elimination is consequential under the social welfare specification in (1”). It should be noted that the baseline social welfare specification can more flexibly capture different kinds of horizontal and vertical preferences. Under the welfare function presented in this Section, the two preferences cannot be separated from each other (see, e.g., Auerbach and Hassett, 2002).

C.13.4 *General cost-of-effort function* To illustrate the earnings choice with a general $k(y,n)$, consider separated income and externality choices, as in Theorem 1. There is no material impact on Theorem 1 and the results of Proposition 3 hold under the same conditions. The environmental gain is still captured by formula (C.23), the redistributive gain is still captured by formula (C.22), but the behavioral effect (C.21) takes now the form

$$d\tau f_n(n) \frac{1}{k_{ny}(y(n),n)} T'(y(n))(1 - F_{q|n}(\tau,n))$$
$$= -d\tau(1 - \mathbb{E}[\omega|n' \geq n])(1 - F_n(n))(1 - F_{q|n}(\tau|n)) \qquad \text{(C.21}'\text{)}$$

where we use formula (16). Combining the effects shows that the insights regarding the tax level and the simple reform are robust to the functional form assumptions. Following similar steps, the formulas of Theorem 2 for the optimal taxes can be derived in the case of general $k(y,n)$.

C.13.5 *Tagging* In the main text.

C.13.6 *Public investments* In the main text.

[48] Taking τ as given and denoting $\tilde{u}(n) = y(n) - k(y(n),n) + b\overline{x} - T(y(n))$, we can write $\overline{\omega}_n^0(\tau) - \overline{\omega}_n^1(\tau) = (\Omega'(\tilde{u}(n) - \tau) - \mathbb{E}[\Omega'(\tilde{u}(n) - q)|n, q < \tau])/\lambda$ which is positive when $\Omega'' < 0$, and decreases (increases) in n if $\Omega''' > 0$ ($\Omega''' < 0$) given that $\tilde{u}(n)$ increases in n and $n \perp\!\!\!\perp q$.