

**Preprints of the  
Max Planck Institute for  
Research on Collective Goods  
Bonn 2011/17**



**Environmental Taxation  
and Redistribution  
Concerns**

Rafael Aigner



MAX PLANCK SOCIETY



# **Environmental Taxation and Redistribution Concerns**

Rafael Aigner

July 2011, revised June 2013

# Environmental Taxation and Redistribution Concerns

Rafael Aigner\*

June 2013

How is the optimal level of Pigouvian taxation influenced by distributive concerns? With second-best instruments, a higher level of income redistribution calls for a lower level of Pigouvian taxation. More redistribution implies higher distortions from income taxation. Pigouvian tax revenues become more valuable and the optimal level of environmental taxation decreases. With first-best instruments, however, the relation between levels of redistribution and Pigouvian taxation is reversed. So second-best Pigouvian taxes are very different from their first-best counterpart – despite apparently identical first order conditions.

*Keywords:* Pigouvian taxation, comparative statics, externalities, optimal income taxation, second-best

*JEL classification:* H21, H23, D62

---

\*Max Planck Institute for Research on Collective Goods. I thank Felix Bierbrauer, Pierre Boyer, Brian Cooper, Christoph Engel, Martin Hellwig, Marco Kleine, Simon Lange, Johannes Spinnewijn, Dezső Szalay, Christian Traxler and two anonymous referees for very helpful comments, as well as participants of the BGSE Micro Workshop, Bonn, the LSE work in progress seminar, London, the 2011 MMM Workshop, Bonn, the PET 12, Taipei, and the EEA/ESEM 2012, Malaga. The usual disclaimer applies. Contact: Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn, Germany. Phone: ++49 228 91416-35. E-mail: aigner@coll.mpg.de

## 1. Introduction

In June 2008, *The Economist* published an article discussing the pros and cons of a preceding oil price boost. On the one hand, the rise might be considered as “a gigantic carbon tax” that helped fighting global warming. On the other hand, it particularly hurt the poor who spent a considerably higher proportion of their income on fuel than the rich. Financial compensations for the core energy demand could help to solve the issue. However “it seems odd to try to prevent energy use with higher taxes . . . and then to subsidise it” (*The Economist*, 2008). The article thereby raised the question how to design green taxes optimally while accounting for distributive concerns. Rising awareness for global environmental problems under persisting inequality has increased the salience of that question. I propose an answer with a focus on the optimal level of green taxes and the relation to redistribution. Despite a huge theoretical literature on environmental taxation and quite some empirical interest on its impact on poor households, the normative question of the optimal response to inequality concerns in the environmental tax design has not had that much of attention.

I employ a simple Mirrlees (1971) income taxation framework which I extend by consumption externalities as proposed by Cremer et al. (1998). Within this framework, a welfare-optimising government uses non-linear income taxes to redistribute and Pigouvian taxation to reduce negative externalities (Pigou, 1932). I show that the two tax design problems are interconnected. In particular, the higher the level of redistribution, the lower the optimal level of environmental taxation. The optimal level has two determinants. First, the marginal social damage caused by the externality. Second, the cost of public funds, defined as the immediate marginal welfare losses associated with income tax collection.<sup>1</sup> If the government puts more weight on redistribution, it will have to accept a higher cost of public funds. Marginal revenues from the environmental tax are then more valuable from the government’s point of view. Contrary to naive intuition, this calls for a lower environmental tax rate. The

---

<sup>1</sup>In formal terms, the cost of public funds is the Lagrangian multiplier of the resource constraint.

reason is that the tax rate is at its efficient level if the marginal revenues exactly compensate society for the marginal external harm. The more valuable the marginal revenues are, the less one needs to compensate for the marginal externality. To put it another way, consider the Pigouvian tax as a bribe that consumers pay the authorities in order to get allowance for pollution. The government is willing to accept a lower bribe if its utility per dollar is higher. Exactly this is the case if the cost of public funds is higher.

I measure the level of redistribution by a parameter that corresponds to the weight of less productive agents in a social-welfare function. As explained, Pigouvian taxation needs to decrease if the parameter increases. When first-best instruments are available, however, the result reverses. Without distortions, the cost of public funds actually decreases in the parameter, as the disutility of the hard working high productive agents receives less weight in the welfare function. Hence the first-best level of Pigouvian taxation increases with the level of redistribution.

My main contribution with respect to the existing literature is to draw attention to the *level* of Pigouvian taxation. Most of the respective literature focuses on tax *rules* and concludes that the distortions caused by second-best instruments do not alter these rules compared to first-best. I show that, despite the first-best shape of these rules, the second-best level of Pigouvian taxation in fact depends on the distortions and the available income tax instruments.

The paper also contributes to a branch of the literature that uses linear tax schemes to analyse the double-dividend hypothesis. Major insights from the linear model carry over to my setting with incentive constraints and optimal taxes. In particular, the optimal environmental tax is lower in second- than in first-best.

### **1.1. Related literature**

This paper is part of a literature in which Pigouvian taxation meets non-linear income taxes under asymmetric information (Mirrlees, 1971). Cremer et al. (1998) show that under the

separability assumptions from Atkinson and Stiglitz (1976) the optimal Pigouvian tax rate is uniform, i.e., it does not discriminate between agents. Gauthier and Laroque (2009) generalise the insight: a certain part of the second-best problem can be separated such that first-best rules apply for that part of the problem. Examples include Pigouvian taxation and the Samuelson Rule. Hellwig (2010) presents a similar result.

Kopczuk (2003), Pirttilä and Tuomala (1997), Jacobs and de Mooij (2011), and Kaplow (2012) explicitly centre on externality taxation within a general (income) taxation problem. In terms of questions posed their contributions are close to mine. Their answers have a different focus, though.

Kopczuk (2003) proposes to decompose the general taxation problem with externalities into two parts: “First, calculate the appropriate Pigouvian tax necessary to correct the externality. Then, with the externality accounted for, the usual second-best problem can be solved using standard formulae.” (p. 84) His result holds for a variety of specifications (including the model presented here) and generalises the ‘principle of targeting’ (Dixit, 1985). Kopczuk (2003) also points out, though, that actually the two parts are interrelated: the Pigouvian tax rate might only be known after the whole problem is solved. My comparative statics analysis characterises this interrelation.

Kaplow (2012) summarises his findings by stating “that simple first-best rules – unmodified for labor supply distortion or distribution – are correct in the model examined.” My analysis highlights that distribution and distortions have a significant influence on environmental policy with respect to tax *levels*, though.

Jacobs and de Mooij (2011, p. 2) find that the “optimal second-best tax on an externality-generating good should not be corrected for the marginal cost of public funds”. However, they use a non-standard definition for the cost of public funds. Our formal analyses are consistent but focus on different interpretations.

An earlier branch of the literature, dating back to Sandmo (1975), examines environ-

mental taxation as part of linear tax systems. Starting with Bovenberg and de Mooij (1994), the linear-taxation model was a primer workhorse model in the discussion of the double-dividend hypothesis.<sup>2</sup> By a central result of this literature, the second-best environmental tax is below the first-best one (e.g. Orosel and Schöb, 1996). As I show, these insights carry over to the case of optimal/non-linear income taxation. Metcalf (2003) uses the linear model to carry out a comparative static analysis with a focus on environmental quality.

My analysis also relates to the literature on comparative static properties of non-linear taxation, with and without public goods (Weymark, 1987; Brett and Weymark, 2008; Bierbrauer and Boyer, 2010), and to applied analyses of the question how to overcome negative distributional effects of environmental taxes (like Metcalf, 1999; West, 2005; Ekins and Dresner, 2004). Rausch et al. (2011) recently studied the U.S. economy, Kosonen (2012) did so for the European Union. The empirical papers investigate the relationship between household income and emission-heavy consumption like driving or heating in order to check whether environmental taxes are regressive. They also discuss distributional impacts of environmental taxes and policies to support the poorest household. I add insights from normative theory to the discussion. In particular, I show that (a) emission-heavy consumption should not be subsidised for poor households and (b) whether or not environmental taxes are regressive is not per se relevant for their optimal level.

The paper is organised as follows: Section 2 presents the model. Section 3 states the rule for optimal internalisation. Section 4 introduces tax systems. Section 5 analyses optimal environmental taxes and provides the main results. Section 6 concludes. The Appendix holds proofs and formal results. It also characterises optimal allocations and discusses corner solutions.

---

<sup>2</sup>See Goulder (1995), Schöb (1997, 2005), Bovenberg (1999) or Bovenberg and Goulder (2002) for more details and surveys on the double-dividend discussion.

## 2. Model

### 2.1. Production

The model is based on Cremer et al. (1998). It considers three different goods. First, an intermediate good that is referred to as *output* and is denoted by  $Y$ . It serves as the numéraire and may be interpreted as money. Second, a *clean*, completely private consumption good,  $C$ , and third, a *dirty* consumption good,  $D$ . The intermediate good can be transformed into the consumption goods at fixed rates of transformation equal to  $p_C$  and  $p_D$ , respectively. Parameters  $p_C$  and  $p_D$  may be interpreted as the producer prices of  $C$  and  $D$ . The intermediate good itself can be produced with a linear technology using labour as the single input good (but labour is not modelled explicitly). The rate of transformation between labour and the intermediate good mirrors productivity and is denoted by  $w$ . It may be interpreted as the wage rate.

### 2.2. Households and allocations

There is a continuum of measure one of agents. They differ in exactly one dimension, namely their productivity, which can be either low or high. An agent's type is denoted by  $\theta \in \{L, H\}$ . Their respective productivity is  $w_\theta \in \{w_L, w_H\}$ . The fraction of low-type agents is denoted by  $\gamma \in (0, 1)$ . An *allocation*  $A$  specifies levels of  $(C, D, Y)$  for both generic types, i.e.,  $A \equiv (C_L, D_L, Y_L, C_H, D_H, Y_H)$ . For a given allocation the utility of an agent of type  $\theta$  is

$$U_\theta(A) = u(C_\theta, D_\theta) - \frac{Y_\theta}{w_\theta} - (\gamma D_L + (1 - \gamma)D_H)e.$$

Function  $u$  is continuously differentiable three times, strictly increasing, strictly concave, has nonnegative cross derivatives, and satisfies the Inada conditions.<sup>3</sup> It represents private

---

<sup>3</sup>Formally,  $u_{CD} \geq 0$  as well as  $u_K \rightarrow \infty$  as  $K \rightarrow 0$ , and  $u_K \rightarrow 0$  as  $K \rightarrow \infty$ , for  $K \in \{C, D\}$ . The Inada conditions are imposed in order to guarantee strictly positive optimal consumption levels. Strict concavity guarantees unique solutions.



consumption utility. In order to produce  $Y_\theta$  units of output, an agent has to provide  $Y_\theta/w_\theta$  units of labour. This provision is associated with a linear disutility. The last term in the utility function reflects the externality. Independently of his type, every agent suffers from the overall consumption of dirty goods,  $\gamma D_L + (1 - \gamma) D_H$ . The social harm is proportional to total dirty good consumption, and  $e > 0$ . From an agent's point of view, own consumption has no negative effect on own utility as a single contribution is negligible in comparison to the large contribution of others. Individual contributions are in fact zero due to the assumption of a continuum of agents.<sup>4</sup>

Notice that all agents in society have quite similar preferences. In particular, their consumption choice for a given budget is identical. Also, they suffer from the externality in exactly the same way. This is not only a simplification but rather a design choice. If agents had different tastes for environmental protection, then the optimal policy would obviously depend on distributional considerations. The homogeneity in agents' preference allows to isolate the more subtle relations between equity and environmental policy.

### 2.3. Social welfare

This paper takes a normative perspective by examining what a social planner (SP) would do in order to maximise the social welfare function  $W$ , defined as

$$W(A) = \alpha U_L(A) + (1 - \alpha) U_H(A), \quad \alpha \in (0, 1),$$

where  $A$  is the allocation. The welfare function is a weighted sum of the generic types' utilities. The parameter  $\alpha$  measures the weight SP puts on a generic low-type agent. If

---

<sup>4</sup>Externalities of this type were termed "atmospheric" by Meade (1952). A different way to interpret the mechanism is to consider a public good that is provided by nature (like "fresh air" or "nice atmosphere"). Dirty good consumption reduces the level or quality of the public good, whereat only total consumption matters. The presented model would fit this interpretation, with the initial amount of this public good normalised to zero.

$\alpha = \gamma$ , then  $W$  is the utilitarian welfare function. For  $\alpha = 1$ ,  $W$  would be the Rawlsian welfare function.

Overall, the economy cannot consume more than it produces in terms of output. Furthermore, an exogenous revenue requirement  $r$  has to be met. The social planner thus faces a resource constraint given by

$$\gamma(Y_L - p_C C_L - p_D D_L) + (1 - \gamma)(Y_H - p_C C_H - p_D D_H) - r \geq 0. \quad (1)$$

If (1) holds and  $A \geq 0$ , then  $A$  is *feasible*. An allocation that maximises  $W$  among all feasible allocations is a *first-best allocation*.

If the social planner does not observe an agent's type, not all feasible allocations are implementable. If, for instance, an allocation disadvantages the high-type agents, they might have an incentive to pretend to be low-types, making it impossible to implement this allocation. As a consequence, under asymmetric information, SP has to ensure that agents do not want to misrepresent their type. This is the case if the following *incentive-compatibility* constraints hold.

$$u(C_L, D_L) - \frac{Y_L}{w_L} \geq u(C_H, D_H) - \frac{Y_H}{w_L}, \quad (2)$$

$$u(C_H, D_H) - \frac{Y_H}{w_H} \geq u(C_L, D_L) - \frac{Y_L}{w_H}. \quad (3)$$

The underlying idea about the relation between incentive compatibility and decentral implementation, i.e. taxation, is known as the 'Taxation Principle' (Hammond, 1979; Guesnerie, 1998).<sup>5</sup> An allocation that maximises welfare among all feasible, incentive-compatible allocations is a *second-best allocation*.

By means of the following assumption, I restrict the analysis to the cases in which SP likes to redistribute from high-type agents to low-type agents.

---

<sup>5</sup>For a complete formal argument see Aigner (2011).

**Assumption 1.**  $\alpha(1 - \gamma)w_H > (1 - \alpha)\gamma w_L$ .

The assumption generally holds if SP puts a sufficiently high welfare weight on low-type agents. The lower  $w_L$  is relative to  $w_H$ , the lower  $\alpha$  may be, because a large difference in productivity provides an efficiency argument for making high-types work more than low-types. A low population share  $\gamma$  of low-types makes redistribution in their favour very cheap, hence it also allows for a low  $\alpha$ .

Given the shape of  $u$ , it is efficient to produce strictly positive amounts of the consumption goods rather than abstain from economic activity. In turn, agents have to provide output. A look at Assumption 1 and the definition of  $W$  shows that, in terms of welfare, it is always better to let the high- rather than the low-type agents produce an output unit. Consequently, high-type agents should produce *all* output. In first-best, this is indeed the case.<sup>6</sup> In second-best, this might be out of reach, as incentive constraints have to be satisfied. It is then ambiguous whether low-type agents work. My main analysis focuses on the cases in which they do work, i.e.  $Y_L > 0$ . In these cases a reallocation of output provision from low- to high-type agents improves welfare, but is possible only if high-types' incentive constraint (3) is slack. Consequently, at an interior second best allocation, (3) needs to bind. As Assumption 1 favours the low-type agents, their incentive constraint (2) is always slack.<sup>7</sup>

In Appendix C, I discuss existence and comparative statics properties of second-best allocations with  $Y_L = 0$ .

### 3. Optimal Internalisation

This section provides a general property of Pareto-optimal allocations, with respect to the externality. At first sight, the presented rule is identical for first- and second-best allocations. This is a reason why redistribution and distortions are sometimes considered to have no

---

<sup>6</sup>If  $Y_H$  could be negative, (first-best) welfare would be unbounded. Obviously, that is not an option. Accordingly, the nonnegativity constraint for  $Y_L$  binds at the first-best allocation.

<sup>7</sup>See Lemmas 1, 2, 3 in Appendix A for the formal arguments.

structural influence on Pigouvian taxation. In the next step, I show, however, in what way the first- and second-best rules are in fact different.

To shorten exposition, I use the following notation for  $J, K \in \{C, D\}$ .  $u^L := u(C_L, D_L)$ ,  $u_J^L := \partial u(C_L, D_L)/\partial J_L$ ,  $u_{KJ}^L := \partial^2 u(C_L, D_L)/(\partial K_L \partial J_L)$ . Analogous definitions apply to  $u^H := u(C_H, D_H)$ . The Lagrangian multiplier of the resource constraint is denoted by  $\lambda$ . All results in the current section are derived in Appendix A.

### 3.1. A rule for optimal internalisation

Both first- and second-best allocation feature the property that the marginal rates of substitution (MRS) between the two consumption goods are the same for both types of agents. Rather than being equal to the rate of transformation (namely, producer-price ratio), as would be the case in an unregulated market, the MRS is equal to

$$MRS = \frac{u_D^L}{u_C^L} = \frac{u_D^H}{u_C^H} = \frac{p_D}{p_C} + \frac{e}{\lambda p_C}. \quad (4)$$

This is a standard result in the literature. It follows, for instance, from the more general analysis by Hellwig (2010). It is driven by the separability feature of the utility functions. Cremer et al. (1998) point out the relation to the famous result in Atkinson and Stiglitz (1976), namely that, under the given assumptions, commodity prices should not be distorted, and all redistribution can be done within the labour market. The intuition of the Atkinson/Stiglitz result is as follows. By assumption, all agents have the same consumption pattern.<sup>8</sup> Therefore the commodity demand cannot be used to screen types and commodity taxation cannot contribute to relax the equity-efficiency trade-off. Hence there is no point in distorting them.

The intuition carries over partially to the case where an externality is introduced. In fact, as agents are equal in terms of their consumption preferences and their exposure to the

---

<sup>8</sup>More precisely, for a given amount of total consumption spending, all agents consume the same commodity bundle.

externality, there is no point in treating them differently in this respect. Yet, it is no longer true that optimal redistribution only affects the labour market. Optimal consumption now depends on multiplier  $\lambda$ . The multiplier is crucially related to redistribution. Also, while at first sight the above formula is the same for both first- and second-best allocation,  $\lambda$  is different in first- and second-best. This has significant consequences for the relation between the degree of redistribution and the degree of intervention in the commodity market.

### 3.2. The cost of public funds

There is no universal definition for the (marginal) cost of public funds in the literature. Jacobs (2012), for instance, recently suggested a definition which implies a marginal cost of 1 for typical optimal taxation schedules. In this paper I stick to the classical definition, also used in the textbook by Dahlby (2008); the (marginal) cost of public funds measures the loss in welfare associated with raising tax revenues. Being a cost, the concept does not account for potential benefits from the revenues. It just tells how (welfare-)costly it is to raise a (marginal) tax dollar.

As is well known, the so defined (marginal) cost of public funds are equal to  $\lambda$ , the Lagrangian multiplier for the resource constraint (1). Formally,  $\lambda = -\partial\hat{W}/\partial r$ , where  $\hat{W}$  is the optimised value of the welfare function. Throughout the paper I normally drop the explicit “marginal” when referring to the cost of public funds – relying on the fact that concept is per se a marginal one. Also, “multiplier” interchangeably refers to  $\lambda$ , i.e., the cost of public funds.

The quasi-linearity in labour allows for closed-form solutions for the multiplier and plainly reveals the dissimilarity between first- and second-best. It also shows the dependency on the underlying parameters  $\alpha$  and  $\gamma$ .

*Second-best* The value of the multiplier at an interior second-best allocation is

$$\lambda^* := \frac{\alpha}{w_L} + \frac{1 - \alpha}{w_H}. \quad (5)$$

To grasp the intuition, note that agents do not benefit from  $r$ , so an increase is pure burden. A way to finance the additional requirement is to increase output. As the incentive constraint for the high-type agents is binding, their output may only be increased if the low-type's output is increased as well. The weighted welfare loss of such an increase is equal to  $\alpha/w_L$  for the generic low-type and  $(1 - \alpha)/w_H$  for the generic high-type. Notice that the multiplier does not depend on the population shares. The reason is that a higher revenue requirement has to be produced by all agents (independently of their type) in order to sustain incentive compatibility.

*First-best* The multiplier at the first-best allocation is

$$\lambda^F := \frac{1 - \alpha}{w_H(1 - \gamma)}. \quad (6)$$

Because only high-types work in first-best, only parameters related to them matter for  $\lambda^F$ . If SP needs an additional unit of revenue, he will make high-type agents work more. As there are only  $1 - \gamma$  high-type agents, the generic high type has to provide  $1/(1 - \gamma)$  (marginal) units of output and needs to work  $1/(w_H(1 - \gamma))$  additional hours. The incurred marginal disutility is weighted by  $1 - \alpha$ .

The multipliers are not only different in size, but also with respect to their directions of change in the parameters  $\alpha$  and  $\gamma$ . The welfare weight has an impact on the optimal tax design with respect to the externality. If interpreted naively, the optimal rule (4) itself hides this fact.

## 4. Taxation

The current section adapts the interpretation of output being money. In this interpretation,  $Y$  denotes gross income,  $w$  corresponds to the wage rate, and  $p_C, p_D$  are producer prices. A tax system  $\tau = (t_C, t_D, T)$  consists of an income tax function  $T$  and specific commodity taxes  $t_C, t_D \in \mathbb{R}$ . Consumer prices are  $q_k := p_k + t_k$  for  $k \in \{C, D\}$ . Consumption may be subsidised through negative commodity taxes.  $T$  may be negative as well, in which case it is a transfer to the agent.

For any type  $\theta$ , let  $(C_\theta(\tau), D_\theta(\tau), Y_\theta(\tau))$  be the maximisers of individual utility, given  $\tau$ :

$$(C_\theta(\tau), D_\theta(\tau), Y_\theta(\tau)) \in \underset{(C,D,Y)}{\operatorname{argmax}} \left( u(C, D) - \frac{Y}{w_\theta} - (\gamma D_L + (1 - \gamma) D_H) e \right. \\ \left. \text{s. t. } q_C C + q_D D \leq Y - T(\cdot) \right). \quad (7)$$

As before, households take  $D_L$  and  $D_H$  as given, so the externality is not relevant for their decision.

When choosing a tax system, the social planner takes individual optimisation into account and needs to respect the following fiscal budget constraint, which is equivalent to resource constraint (1).

$$\gamma \left( T_L(\tau) + t_C C_L(\tau) + t_D D_L(\tau) \right) + (1 - \gamma) \left( T_H(\tau) + t_C C_H(\tau) + t_D D_H(\tau) \right) \geq r \quad (8)$$

Here,  $T_L(\tau)$  and  $T_H(\tau)$  amount to the respective total income tax payments of low- and high-type agents.

If (8) is satisfied for some tax system  $\tau$ , then  $\tau$  is said to *implement* allocation  $A$  with  $A = (C_L(\tau), D_L(\tau), Y_L(\tau), C_H(\tau), D_H(\tau), Y_H(\tau))$  as defined by (7). The set of available tax systems to choose from depends on the informational constraints. When the social planner can observe an agent's type, the income tax may be contingent on the type. Under

asymmetric information it can only be contingent on observed gross income. In fact, with  $T : (w, Y) \mapsto T(w, Y)$ , it is possible to find a system  $\tau$  that implements the first-best allocation. With  $T : Y \mapsto T(Y)$ , it is possible to find a system  $\tau$  that implements the second-best allocation. This insight allows to restrict attention to the chosen tax structure albeit the linearity in commodity taxation.<sup>9</sup>

#### 4.1. Normalisation

As usual in these type of models, there is a degree of freedom in the taxation choice. A common way to deal with this is to normalise the tax system and often it is innocuous to do so. Yet, when properties of the tax system, like a particular tax level, are the object of interest rather than the real allocation, one has to be careful with normalisations.

This was a major issue in the double-dividend discussion between Bovenberg and de Mooij (1994, 1997), Fullerton (1997) and others. The discussion centres on the comparison of the second-best pollution tax and the first-best Pigouvian tax (the marginal social harm). The actual tax level obviously depends on the chosen normalisation and a priori it is unclear which normalisation is “correct”. In a related contribution, Schöb (1997) focuses on the normalisation choice and shows that also “the difference of the first-best and second-best optimal tax on the polluting good depends on the normalization chosen.” (p. 174) He concludes that “such a comparison provides an inappropriate indicator for the existence of a second dividend.” (ibid.)

To obtain valid results on comparative static properties of environmental taxation and the relation between first- and second-best level, it is important to avoid the “normalization trap” (Schöb, 1997). Orosel and Schöb (1996) propose to study an object called the *second-best internalization tax*. Unlike an actual tax rate, it is a “real” variable, derived directly from

---

<sup>9</sup>The underlying arguments are standard. For a rigorous application to the current model see Aigner (2011, Appendix B).



the underlying allocation, and independent of the normalisation.<sup>10</sup> Using their concept, the authors find a particular normalisation to be correct for doing the comparison of actual first- vs. second-best tax rates.

The aforementioned contributions feature linear labour and commodity taxation, and do not model distributive issues.<sup>11</sup> Their insights on normalisations carry over to my model, though. Here the “real” object of interest is the so called *greenness*  $g$ . For any tax system  $\tau = (t_C, t_D, T)$ , the greenness is defined by

$$g := t_D - t_C \frac{q_D}{q_C}.$$

Similar to the second-best internalisation tax proposed by Orosel and Schöb (1996), the greenness is a real variable independent of the normalisation. More precisely,  $g$  is unique in the sense it is the same for any tax system which implements the second-best allocation. The same is true with respect to the first-best allocation.

For an intuitive understanding of the greenness consider an agent who faces some tax system  $\tau$  and decides to purchase an additional (marginal) unit of  $D$ , while keeping total spending constant. Then the greenness is the change in the agent’s total tax payment. Thereby it quantifies the tax system’s inherent incentives to shift consumption from  $D$  to  $C$ . Plainly put, it tells how “green” the system is. Notice that the a tax system could be green due to high  $t_D$  or due to low (potentially negative)  $t_C$ . The greenness covers both cases.

It turns out that the greenness equals the tax rate  $t_D$  on the dirty good iff the tax rate  $t_C$  on the clean good is normalised to zero: precisely the normalisation identified as “correct” by Orosel and Schöb (1996) for the respective purpose. For this reason it is save to proceed the analysis with  $t_C \equiv 0$ . For further reference, I call such a tax system *normalised*.

---

<sup>10</sup>The definition of the second-best internalisation tax uses the observation that private marginal utility should equal social marginal welfare – a property of an allocation rather than a tax system.

<sup>11</sup>The papers on the double dividend normally have identical/representative consumers. Distributional concerns appear only indirectly as a motive for the unavailability of lump sum taxation.

For a detailed discussion of the greenness and the respective proofs see Aigner (2011).

## 5. The optimal Pigouvian tax

Given  $t_C \equiv 0$ , how high should  $t_D$  be? An optimum is characterised by the fact that a marginal reallocation does not change welfare. In particular, keeping private spending constant, a marginal change in consumption levels must not change welfare. Consider a marginal shift from  $C$  to  $D$  (for all agents, taking account of differences in prices). This has three effects: (1) Consumption utility  $u$  is unchanged as agents are at their individual optimum. (2) External harm increases at rate  $e$ . (3) Tax revenues increase at rate  $t_D$  and relax the budget constraint of the social planner. Multiplier  $\lambda$  tells how welfare is affected from relaxing the public budget. Thus the marginal effect of tax revenues on welfare amounts to  $t_D\lambda$ . The overall marginal change in welfare is  $-e + t_D\lambda$ . For this change to be zero,  $t_D$  needs to be

$$t_D = \frac{e}{\lambda}. \quad (9)$$

### 5.1. The comparative statics of Pigouvian taxation

The following Propositions essentially combine equation (9) with the findings from Section 3.2. They state the main result of the paper: comparative static properties of those tax systems that implement the first- and second-best allocation, respectively.

**Proposition 1** (First-best Pigouvian taxation). *If a normalised tax system  $\tau^F = (0, t_D^F, T^F)$  implements the first-best allocation  $A^F$ , then*

$$t_D^F = \frac{e}{\lambda^F}.$$

Furthermore,

$$\frac{\partial t_D^F}{\partial \alpha} > 0, \quad \frac{\partial t_D^F}{\partial \gamma} < 0.$$

Despite the lack of distortions, distributive concerns influence the environmental tax; more redistribution calls for a higher first-best Pigouvian tax  $t_D^F$ . The relation reverses completely if first-best instruments are not available and the labour market is distorted.

**Proposition 2** (Second-best Pigouvian taxation). *If a normalised tax system  $\tau^* = (0, t_D^*, T^*)$  implements an interior second-best allocation  $A^*$ , then*

$$t_D^* = \frac{e}{\lambda^*}.$$

Furthermore,

$$\frac{\partial t_D^*}{\partial \alpha} < 0, \quad \frac{\partial t_D^*}{\partial \gamma} = 0.$$

In a nutshell, higher labour market distortions coming from increased redistribution imply a lower optimal Pigouvian tax level. To develop a detailed intuition for the results, decompose the comparative statics into two aspects. (1) In first- as well as in second-best  $t_D$  is inversely proportional to  $\lambda$ . (2) The reaction of  $\lambda$  differs for first- and second-best. The first aspect is not new. It is already well established for models of linear labour/commodity taxation. As shown, it carries over to a world with incentive constraints. The second aspect has not drawn that much of attention in the literature but is crucial as it drives the reversed results. I discuss the two aspects in turn.

### 5.1.1. The inverse relation of environmental taxes and the cost of public funds

To grasp the intuition behind the inverse relation, consider the purpose of Pigouvian taxation: its (only) goal is to restore the efficient level of dirty-good consumption. From a welfare perspective, a unit of the dirty-good should be consumed if and only if consumption is not only individually optimal, but private benefits also outweigh social harm. Consequently, dirty-good consumption is at its socially optimal level only if marginal private (net) benefits exactly equal marginal social harm. To measure and compare these two objects, it is useful to quantify them in terms of money.

(1) The optimising agent is willing to pay  $t_D$  units of additional taxes for the right to consume her last unit of  $D$  rather than spending the respective money on  $C$ . So  $t_D$  is a good measure of (net) private benefits of the marginal unit of dirty-good consumption.

(2) Now consider the social planner. If  $D$  increases by one unit, welfare decreases by  $e$ . If SP receives exactly  $e/\lambda$  units of money to relax the budget constraint, welfare increases by  $(e/\lambda)\lambda = e$ . Thus the marginal social harm measured in money is equal to  $e/\lambda$ . It is the exact amount of money that society needs as a compensation for additional dirty-good consumption. The amount is lower if the received money is more useful in the sense that the cost of public funds is higher. Putting together (1) and (2) shows that if  $t_D = e/\lambda$ , then individual maximisation leads to an allocation in which, at the margin, private (net) benefits equal social harm.

A more naive view, which evaluates Pigouvian taxes in a partial or isolated manner rather than viewing it as part of a whole tax system, could reason that 'Pigouvian taxes do two things: reduce pollution and create revenue. So they should be high if pollution is severe or if revenues are very valuable to the state.' Naive intuition would thus suggest that higher cost of public funds (associated with marginal tax revenue being more valuable) should lead to higher Pigouvian tax rates. In fact, this 'rationale' would provide a straightforward argument for the double dividend hypothesis, which by now has been mostly falsified (e.g.,

Fullerton and Metcalf, 1998). The strong form of the double-dividend hypothesis states that a revenue-neutral introduction of green taxes is desirable even if environmental benefits are not taken into account (Goulder, 1995). In the model that I propose this fails clearly: Pigouvian taxation, namely  $t_D > 0$ , is optimal only if an externality is present, i.e., if  $e > 0$ . Among others, Bovenberg (1999) gives the same argument, albeit for a model with linear taxation. Empirical investigations by Goulder (1995) tend to reject the hypothesis as well.

The intuition that rejects the double dividend hypothesis is also central to the comparative static analysis. The more valuable the marginal tax revenues, the less is needed to compensate for the marginal externality, and — because the one and only purpose of Pigouvian taxation is to induce alignment of private benefits and social harm at the margin — a *lower* Pigouvian tax rate is asked for. Various authors have noticed the underlying rationale in their respective settings, so it applies quite universally (e.g., Schöb, 1997). As Bovenberg and de Mooij (1994) put it, “each unit of pollution does not have to yield as much public revenue to offset the environmental damage if this revenue becomes more valuable” (p. 361).

In a recent contribution, Jacobs and de Mooij (2011) make the seemingly contradictory statement that the optimal second-best environmental tax is not sensitive to the cost of public funds at all. Their conclusion follows from a their newly proposed definition of the cost of public funds. So the difference in conclusion is one of interpretations rather than formal results. Their interpretation suggests that tax distortions do not play a role for optimal environmental taxes, which clearly is at odds with my analysis. Indeed, Jacobs and de Mooij (2011, p. 13) qualify their interpretation themselves: “The optimal second-best environmental tax does require a correction for distributional concerns and interactions with labor supply, but not for pre-existing tax distortions.” The comparative statics results fill the gap of specifying the “correction for distributional concerns” but also broaden the existing insights by highlighting that even without distortions, distribution concerns influence the optimal environmental tax level.

I should highlight that the preceding discussion is about *marginal* rather than *total* revenues. The difference is crucial: total revenues from Pigouvian taxation do *not* compensate for the overall external harm. Although the two figures coincide in the linear specification, they generally differ. More to the point, Pigouvian revenues should not be used to compensate the harmed people; it is not its purpose, and it might reduce incentives to avoid exposure to an externality in the first place (Oates, 1995).

### 5.1.2. How the cost of public funds changes in parameters

The changes of  $t_D^F$  and  $t_D^*$  with respect to welfare weight  $\alpha$  have different signs. This point is worth stressing again as previous contributions with non-linear income taxes tend to highlight the similarities rather than the differences of first- and second-best Pigouvian taxes. This focus comes naturally when examining the optimal rules, which are – almost – identical for first- and second-best.

For the version of their model that resembles the one of this paper, Cremer et al. (1998) conclude that “the optimal tax on the externality generating good is strictly Pigouvian” (p. 345; Proposition 1), where the term ‘Pigouvian’ is based on the first-best tax on the dirty good (Definition 1). Likewise Gauthier and Laroque (2009) show that first-best rules quite often hold also at second-best allocations if utility is separable. With respect to externalities they find that “a non-satiated second best allocation can be supported with a first best Pigouvian tax” (Remark 4).<sup>12</sup> Kopczuk (2003) and Kaplow (2012) make similar observations.

While all of these findings are correct, they suggest (quite explicitly in some cases) that distortions are not that relevant for the second-best tax. Proposition 2 highlights the opposite. Also, these results might distract from the considerable differences between first- and second-best when it comes to tax *levels* rather than tax *rules*. In fact, the optimal rule for

---

<sup>12</sup>Gauthier and Laroque (2009) do point out, though, that the whole second-best problem must be solved to obtain the actual Pigouvian tax.

the model at hand is given in (4) – for both first- and second-best. Only an inspection of the respective multipliers reveals the differences between them.

*First-best multiplier* Recall that only high types work at a first-best allocation and that  $\lambda^F$  is derived from an output increase of high-types. Now, if  $\alpha$  increases, SP cares less about high-type agents working more, thus the cost of public fund *decrease*. Tax revenues generated by  $t_D$  are less valuable per unit so more (marginal) revenues need to be collected at the optimum. If  $\gamma$  is increased, the generic high-type has to work more for an higher overall output requirement and the cost of public funds increases. Marginal revenues generated by  $t_D$  now have higher value per unit and less is needed to satisfy optimality condition (9).

*Second-best multiplier* From (5) the cost of public funds at a second-best allocation is  $\lambda^* = \alpha/w_L + (1 - \alpha)/w_H$ . As  $\lambda^*$  does not depend on  $\gamma$ , neither does  $t_D^*$ . The higher the welfare weight of low-type agents, the more redistribution is asked for and the more distortions are accepted. Higher distortions imply higher excess burden of taxation and thereby *higher* cost of public funds. Marginal revenues from Pigouvian taxation are then more valuable and less marginal revenue is needed to satisfy optimality condition (9).

*First- vs. second-best* As argued, the different directions of change of the Pigouvian tax with respect to the welfare weight derive from the different reactions of the cost of public funds. In the one case, the social planner cares less about those who work, i.e., environmental revenues decrease in value, in the other case, income tax distortion increase and environmental revenues increase in value.

*Distribution and taxes* To sum up, this section highlights a link between Pigouvian taxation and the degree of redistribution as measured by welfare weight  $\alpha$ . Welfare optimising societies with different opinions about equity need to have different levels of Pigouvian taxation, even if first-best instruments are feasible. This is not entirely obvious because in basic

partial equilibrium models, the level of Pigouvian taxation is typically pinned down solely by Pareto efficiency. Asymmetric information proves to be a crucial determinant of the link between environmental taxation and redistribution: the sign of the dependence changes when going from first- to second-best.

### Corner solutions

So far, the second-best comparative statics assumed an interior solution. Corollary 2 in Appendix C shows that for corner solution  $\partial t_D / \partial \alpha = 0$ . In fact, if  $Y_L^* = 0$  for some  $\alpha$ , then a further increase in  $\alpha$  cannot change the optimal allocation at all: it is neither possible to decrease  $Y_L$  nor to increase low-type agents' consumption without violating incentive constraints. Consequently, the optimal tax remains constant as well.<sup>13</sup>

### 5.2. First- vs second-best Pigouvian tax level

Bovenberg and de Mooij (1994) examine the double dividend hypothesis by comparing first- and second-best environmental tax. Their contribution led to quite some follow-up papers on the subject. The workhorse model of this literature is a representative household model with linear income and commodity taxes (e.g., Bovenberg, 1999). Lump sum taxes are allowed or disallowed for exogenous reasons. In these settings, the second-best environmental tax falls short of the first-best one.<sup>14</sup> The result carries over to my setting with incentive-constraint redistribution.

**Proposition 3.** *Fix parameters and consider two normalised tax systems,  $\tau^F$  and  $\tau^*$ , which implement the first- and second-best allocations, respectively. Then*

$$t_D^* < t_D^F.$$

<sup>13</sup>The result is shown to hold for  $\alpha \geq \gamma$ . I expect it to hold for a broader range of parameters, though.

<sup>14</sup>See also Schöb (2005). As detailed in Section 4.1 and the references mentioned there, the comparison hinges on the “correct” normalisation choice, which by now is well-understood (Orosel and Schöb, 1996).



*Proof.* (6) and (5) imply  $\lambda^* - \lambda^F > 0$  due to Assumption 1.  $t_D^* < t_D^F$  follows from (9).  $\square$

Following the discussion of the comparative statics, the intuition for the result should be clear. The second-best tax system distorts the labour market which increases the cost of public funds compared to a first-best system. The difference in the cost of public funds causes the difference in the tax levels.

### 5.3. The role of regressivity

Environmental taxes are *regressive* if total tax payments in proportion to total consumption spending decrease in income levels. If so, these taxes impose a disproportionate burden on low-income households. Energy-intensive goods like electricity and heating are often considered to feature regressive consumption pattern, so taxes on these goods might indeed be regressive. The Economist (2008), for instance, is concerned about this possibility in the article I cited in the Introduction. The presumption also gave rise to applied studies on the impact of green tax reforms on low-income households like Metcalf (1999), West (2005) or Ekins and Dresner (2004). In a recent empirical study focusing on the European Union, Kosonen (2012) finds that electricity and heating tend to be regressive. For transport fuel and vehicles it is the other way around, though; they seem to be progressive. Also, there are considerable differences between countries. Overall, the actual evidence for regressive spending patterns is quite mixed.

The analysis in this paper contributes to these considerations by showing that the question of regressivity might not be that important after all. By choosing consumption utility  $u$ , one could easily induce regressive as well as progressive consumption patterns for the dirty good. Yet, none of this would have any influence on the optimal level of the environmental tax. The income tax/transfer system might adapt but not so the optimal incentive to reduce pollution.

To be concrete consider a normalised second-best tax system. Now change the underlying sub-utility  $u$  in a way to make dirty good spending patterns more/less regressive. How does the optimal tax system change? First,  $t_D$  does not change. The income tax  $T$  generally does change, though. If marginal income taxes changed, the incentive to pollute would change despite constant  $t_D$ . An analysis of the optimal allocation shows, however, that this is not the case; the distortions induced by redistribution do not hinge on the particularities of  $u$  (see Appendix A, Proposition 5). So, the optimal response to a “sudden regressivity” requires no action with respect to the environmental tax level. It could trigger a higher transfer amount to support the poor, though.

Admittedly, this neutrality observation would be diluted without the linearity and separability assumptions. However, regressivity apparently is not relevant for the Pigouvian tax per se – otherwise this should appear in the model presented here, which allows for a pure regressivity effect.

## 6. Conclusion

This paper looks at the interdependence of distributive and environmental policies from a normative perspective. It reveals a qualitative difference between first- and second-best. Distributive goals and environmental policies are linked by the cost of public funds. On the one hand they influence the optimal environmental tax level, on the other hand they are a function of distribution policies. I find that if society wants more redistribution, the second-best environmental tax is lower, whereas the first-best environmental tax is higher.

The results also clarify some aspects of the literature on Pigouvian taxation. First, it is important to distinguish optimal rules from optimal levels. Former contributions on second-best environmental taxation with non-linear income taxes tend to focus on the optimal tax rule and point out their “first best flavor”<sup>15</sup>, emphasising the similarity of first- and second-

---

<sup>15</sup>Gauthier and Laroque (2009, p. 1168)

best with respect to environmental taxes. My focus on tax level shows significant differences in the level and the parameter dependence. Income tax distortions do play a substantial role for optimal environmental taxes.

Second, insights gained from models with linear income/commodity taxation carry over to settings with non-linear income taxation and incentive constraints. This holds true for the role of distortions as well as the result that the second-best environmental tax falls short of the first-best one.

What can be learned in terms of policy implications? First, the view that the two goals of redistribution and environmental protection can be addressed independently by means of two different instruments (income tax and Pigouvian taxation) needs to be reconsidered. In particular, the designer of environmental taxes has to account for the value in terms of welfare that is created by the tax revenues. This value is a function of the income tax schedule and depends on the set of available instruments as well as on informational constraints. The optimal tax level then derives from the trade off between external harm and useful tax revenues. Importantly, it is the marginal effect that counts. Total revenues are irrelevant for the optimal level of Pigouvian taxation. So are the total environmental taxes paid by the households.

Second, the intricate empirical question of regressivity is not too relevant for tax designers. In fact, whether or not environmental taxes are regressive should not influence their level. In particular, tax rates should not be reduced for poor households in an attempt to compensate for any disproportionate burden from environmental taxes. This would reduce incentives and provide an inefficient means of redistribution. Instead, one might raise the transfers to those households.

## A. First- and Second-Best Allocation

**Lemma 1.** *If  $A^F$  is a first-best allocation under Assumption 1, then  $Y_L^F = 0$  and  $Y_H^F > 0$ .*

*Proof.* Suppose  $Y_L^F > 0$ . If the total output of all low-type agents is lowered by  $\Delta \in (0, \gamma Y_L^F)$ , every low-type individual may reduce his own output by  $\Delta/\gamma$ . The immediate welfare gain is  $\alpha\Delta/(\gamma w_L)$ . To finance the output reduction high types have to increase their total output by  $\Delta$ , resp. their individual output by  $\Delta/(1 - \gamma)$ . The immediate welfare loss is  $(1 - \alpha)\Delta/((1 - \gamma)w_H)$ . The net effect of the alteration is strictly positive given Assumption 1, a contradiction.

Hence  $Y_L^F = 0$ .  $Y_H^F > 0$  needs to hold given the Inada-conditions on  $u$ .  $\square$

**Lemma 2.** *If  $A$  is a second-best allocation, then*

1. *At most one incentive compatibility constraint is binding.*
2.  *$Y_H > Y_L$  and  $u^H > u^L$ .*

*Proof.* 1. Suppose the contrary. Summation of both ICs yields  $Y_L = Y_H$  and  $u(C_L, D_L) = u(C_H, D_H)$ . Due to the shape of  $u$ , this can be optimal only if  $(C_L, D_L) = (C_H, D_H)$ . To complete the argument, it suffices to show that such a bunching allocation is dominated by a *constrained laissez-faire* allocation. Fix any feasible bunching allocation  $A^b = (C^b, D^b, Y^b, C^b, D^b, Y^b)$  and define for any type  $\theta$ ,  $(C_\theta^{lf}, Y_\theta^{lf}) := \operatorname{argmax}_{C, Y} \{u(C, D^b) - Y/w_\theta \text{ s.t. } Y \geq p_C C + p_D D^b + r\}$ . Then, in particular,  $u(C_\theta^{lf}, D^b) - Y^{lf}/w_\theta \geq u(C^b, D^b) - Y^b/w_\theta$ . Furthermore, maximisers are unique and  $(C_L^{lf}, Y_L^{lf}) \neq (C_H^{lf}, Y_H^{lf})$ . Thus there exist  $\theta$  such that  $u(C_\theta^{lf}, D^b) - Y^{lf}/w_\theta > u(C^b, D^b) - Y^b/w_\theta$ . The constraint laissez-faire allocation  $A^{lf}$  thereby Pareto-dominates the bunching allocation  $A^b$ .  $A^{lf}$  is also incentive

compatible and feasible. Hence,  $A^b$  cannot be a solution to the second-best problem and the contradiction is completed. (The argument builds on Bierbrauer and Boyer, 2010, Lemma 1)

2. Add both ICs to obtain  $Y_H \geq Y_L$ . Equality would imply a bunching allocation which is not optimal as shown above. Hence  $Y_H > Y_L$ .  $IC_H$  then implies  $u^H > u^L$ .  $\square$

**Lemma 3.** *If  $A^*$  is an interior second-best allocation under Assumption 1, then high types' incentive constraint (3) is binding, low types' incentive constraint (2) is slack.*

*Proof.* Suppose by contradiction that (3) was slack, i.e.,  $u^H - Y_H/w_H > u^L - Y_L/w_H$ . Then there exists an  $\epsilon > 0$  such that also  $u_H - (Y_H + \epsilon)/w_H > u_L - (Y_L - \epsilon(1 - \gamma)/\gamma)/w_H$ . The  $\epsilon$ -perturbed allocation is constructed in a way to keep total output constant. Incentive compatibility is sustained, too. The welfare effect of the perturbation is

$$dW = \alpha \frac{1 - \gamma}{\gamma w_L} \epsilon - \frac{(1 - \alpha)\epsilon}{w_H} > 0 \Leftrightarrow \frac{\alpha}{\gamma w_L} > \frac{(1 - \alpha)}{(1 - \gamma)w_H}$$

$dW$  is strictly positive precisely under Assumption 1, hence a contradiction.

If  $IC_H$  is binding then  $IC_L$  must be slack by Lemma 2.  $\square$

### First-order conditions

Considering the lemmas, an appropriate Lagrangian for an optimal allocation is

$$\begin{aligned} \mathcal{L} = & \alpha [u(C_L, D_L) - Y_L/w_L - (\gamma D_L + (1 - \gamma)D_H)e] \\ & + (1 - \alpha) [u(C_H, D_H) - Y_H/w_H - (\gamma D_L + (1 - \gamma)D_H)e] \\ & + \lambda(\gamma(Y_L - p_C C_L - p_D D_L) + (1 - \gamma)(Y_H - p_C C_H - p_D D_H) - r) \quad (10) \\ & + \mu(u(C_H, D_H) - Y_H/w_H - u(C_L, D_L) + Y_L/w_H) \\ & + \delta Y_L. \end{aligned}$$

Next, set the partial derivatives to zero.

$$\alpha u_C^L - \gamma \lambda p_C - \mu u_C^L = 0 \Leftrightarrow (\alpha - \mu) u_C^L = \lambda \gamma p_C \quad (11)$$

$$\alpha u_D^L - \gamma \lambda p_D - \mu u_D^L - \gamma e = 0 \Leftrightarrow (\alpha - \mu) u_D^L = \lambda \gamma p_D + \gamma e \quad (12)$$

$$-\alpha/w_L + \gamma \lambda + \mu/w_H + \delta = 0 \Leftrightarrow \gamma \lambda = \alpha/w_L - \mu/w_H - \delta \quad (13)$$

$$(1 - \alpha) u_C^H - (1 - \gamma) \lambda p_C + \mu u_C^H = 0 \Leftrightarrow (1 - \alpha + \mu) u_C^H = (1 - \gamma) \lambda p_C \quad (14)$$

$$(1 - \alpha) u_D^H - (1 - \gamma) \lambda p_D + \mu u_D^H - (1 - \gamma) e = 0 \Leftrightarrow (1 - \alpha + \mu) u_D^H = (1 - \gamma) (\lambda p_D + e) \quad (15)$$

$$-(1 - \alpha)/w_H + (1 - \gamma) \lambda - \mu/w_H = 0 \Leftrightarrow (1 - \alpha + \mu)/w_H = (1 - \gamma) \lambda \quad (16)$$

It follows that

$$\begin{aligned} \mu &= w_H \left( \frac{(1 - \gamma) \alpha}{w_L} - \frac{\gamma(1 - \alpha)}{w_H} \right) - w_H (1 - \gamma) \delta \\ \lambda &= \frac{\alpha}{w_L} + \frac{1 - \alpha}{w_H} - \delta \end{aligned} \quad (17)$$

For a first-best allocation, set  $\mu = 0$ , for an interior second-best allocation, set  $\delta = 0$ .

**Proposition 4** (First-best allocation). *Given Assumption 1, allocation  $A^F$  is a first-best allocation if and only if it satisfies the following system of equations.*

$$\begin{aligned} u_C^L &= \frac{p_C}{w_H} \frac{1 - \alpha}{\alpha} \frac{\gamma}{1 - \gamma}, & u_D^L &= \frac{p_D + e/\lambda^F}{w_H} \frac{1 - \alpha}{\alpha} \frac{\gamma}{1 - \gamma}, \\ u_C^H &= \frac{p_C}{w_H}, & u_D^H &= \frac{p_D + e/\lambda^F}{w_H}, \end{aligned}$$

$$Y_L^F = 0, \quad Y_H^F = \frac{\gamma}{1 - \gamma} (p_C C_L^F + p_D D_L^F) + p_C C_H^F + p_D D_H^F + \frac{r}{1 - \gamma},$$

$$\lambda^F = \frac{1 - \alpha}{w_H(1 - \gamma)}.$$

*Proof.* With  $\mu = 0$ , the Lagrange function (10) is concave and the first order conditions are necessary and sufficient for a solution. Consider conditions (13) and (16) with  $\mu = 0$ . Then

$$\lambda^F = \frac{1 - \alpha}{w_H(1 - \gamma)}, \quad \delta^F = \frac{\alpha}{w_L} - \gamma\lambda > 0 \Rightarrow Y_L^F = 0.$$

Notice that the inequality is satisfied if and only if Assumption 1 holds. The statement of the Proposition now follows from conditions (11), (12), (14), (15), and the binding resource constraint (1).  $\square$

Low-type agents do not work at all. Due to linear disutility from working, Assumption 1 implies that any given amount of output requirement fosters lower aggregated disutility if it is provided solely by high types rather than low types. If  $Y_L$  could be negative, welfare would be unbounded.

For a moment, ignore the Lagrangian multiplier of the resource constraint  $\lambda^F$ . Then consumption of high types is independent of the welfare weight and the population shares, and is just determined by efficiency considerations. It departs from standard results only through a corrective element that takes care of the external effects of dirty-good consumption. The consumption levels of the low-type agents, though, heavily depend on welfare weights as well as the population shares. The underlying trade-off lies between consumption utility of low-types and disutility of high types, who have to work for the provision of low-type consumption. Low-type productivity  $w_L$  is irrelevant for the allocation given that they do not work.

**Lemma 4.** *Let  $u(C, D)$  be strictly concave and continuously differentiable and let  $k^C, k^D$  be two constants such that the system  $u_C(C, D) = k^C$ ,  $u_D(C, D) = k^D$  has a solution. Then the solution is unique.*

*Proof.* Consider the three-dimensional space. Let  $s = (s^C, s^D)$  be a solution. The tangential plane at  $S = (s^C, s^D, u(s))$  is spanned by the directions of the two partial derivatives at  $S$ .

As  $u$  is strictly concave, the whole range of  $u$  – except  $u(s)$  – lies below that plane. Now consider a point  $s'$  that also solves the above system but is different from  $s$ . The tangential plane at  $s'$  is parallel to the one at  $s$ , yet one of the planes is higher than the other. But than it is no longer possible that the whole range of  $u$  lies below the lower plane. This creates a contradiction.  $\square$

**Proposition 5** (Interior second-best allocation). *If  $A^*$  is an interior second-best allocation under Assumption 1, then it is unique and solves the following system of equations with  $d = \mu^*(1 - w_L/w_H)/(\alpha - \mu^*)$ .*

$$\begin{aligned} u_C^L &= \frac{p_C}{w_L}(1 + d) & u_D^L &= \frac{p_D + e/\lambda^*}{w_L}(1 + d) \\ u_C^H &= \frac{p_C}{w_H} & u_D^H &= \frac{p_D + e/\lambda^*}{w_H} \end{aligned}$$

$$\begin{aligned} Y_L^* &= r + \gamma(p_C C_L^* + p_D D_L^*) + (1 - \gamma)(p_C C_H^* + p_D D_H^*) - w_H(u^H - u^L)(1 - \gamma) \\ Y_H^* &= r + \gamma(p_C C_L^* + p_D D_L^*) + (1 - \gamma)(p_C C_H^* + p_D D_H^*) + w_H(u^H - u^L)\gamma \end{aligned}$$

$$\lambda^* = \frac{\alpha}{w_L} + \frac{1 - \alpha}{w_H}, \quad \mu^* = \alpha(1 - \gamma)\frac{w_H}{w_L} - (1 - \alpha)\gamma. \quad (19)$$

*Proof.* If  $A^*$  is an interior second-best allocation, then it satisfies conditions (13) and (16) with  $\delta$  set to zero. Then  $\lambda^*$  and  $\mu^*$  are uniquely determined and strictly positive. For given values of  $\lambda^*$  and  $\mu^*$ , (11), (14), (12), and (15) uniquely determine the consumption levels (uniqueness is established by Lemma 4). Output requirements follow from the binding resource constraint (1) combined with the binding incentive constraint (3).  $\square$

The conditions for high-type consumption levels are almost identical to the corresponding first-best conditions (18). The subtle but important difference lies in the Lagrangian multi-



plier  $\lambda$ , which is different in first- and second-best and, most importantly, features different comparative statics properties.

The consumption levels of low types are distorted downwards, i.e., the labour choice is distorted in favour of leisure. The distortion is captured by  $d$  and is higher if  $\mu^*$  is higher or the difference in productivities is larger.

## B. Proofs of the main results

*Proof of Proposition 1.* Combining  $t_C \equiv 0$  with Proposition 4 and the fact that

$$\frac{q_D}{q_C} = \frac{p_D + t_D}{t_C} = MRS$$

yields  $MRS = p_D/p_C + e/(\lambda^F p_C) = (p_D + t_D)/p_C \Leftrightarrow t_D = e/\lambda^F$ . From (6),  $\lambda^F = (1 - \alpha)/(w_H(1 - \gamma))$ . The comparative statics immediatly follow.  $\square$

*Proof of Proposition 2.* Combining  $t_C \equiv 0$  with Proposition 5 and the fact that

$$\frac{q_D}{q_C} = \frac{p_D + t_D}{t_C} = MRS$$

yields  $MRS = p_D/p_C + e/(\lambda^* p_C) = (p_D + t_D)/p_C \Leftrightarrow t_D = e/\lambda^*$ . From (19),  $\lambda^* = \alpha/w_L + (1 - \alpha)/w_H$ . The comparative statics immediatly follow.  $\square$

## C. Second-Best Corner Solutions

A *second-best corner solution* is a second-best allocation with  $Y_L = 0$ . In this section I show that for some parameters this is the relevant case. I then claim that a corner solution does not change at all if  $\alpha$  is increased (Proposition 6). Thereby I extend the comparative statics properties of Pigouvian taxation to instances of corner solutions (Corollary 2). Proposition

6 builds on a conjecture that generalises Lemma 3. Unfortunately, I can only partially verify that conjecture (Lemma 7).

**Lemma 5.** *If  $A$  is a second-best allocation and*

$$\alpha \geq \frac{\gamma}{1 - \gamma} \frac{1}{\frac{w_H}{w_L} - 1}, \quad (20)$$

*then  $A$  is a corner solution, i.e.  $Y_L = 0$ .*

*Proof.* By contradiction, assume  $A$  is an interior solution. Then it satisfies conditions (11) to (16) with  $\delta = 0$ . Hence  $\mu = (1 - \gamma)\alpha w_H/w_L - \gamma(1 - \alpha)$  by (17). But then the condition on parameters stated in the Lemma implies  $\alpha - \mu \leq 0$ . This, however, contradicts (11).  $\square$

**Remark 1.** Bierbrauer and Boyer (2010) exclude corner solutions in their comparative statics analysis by assuming  $1 > (1 - \gamma)w_H/w_L$ . Their inequality always holds if (20) is not satisfied, but the converse is not true. Hence, I do not expect (20) to be a necessary condition for a corner solution.

To proceed, let me introduce some convenient notation.

**Definition 1.** Define  $A(\alpha)$  to be a second-best allocation, in which the welfare weight is given by  $\alpha$  and all other parameters are fixed. For  $\alpha' < \alpha''$  and  $\theta \in \{L, H\}$  define  $U'_\theta := U_\theta(A(\alpha'))$ ,  $U''_\theta := U_\theta(A(\alpha''))$ , and  $dU_\theta := U''_\theta - U'_\theta$ .

Quite intuitively, if the taste for redistribution increases, low-type agents receive higher utility. At the same time, high-type agents have to receive lower utility because someone has to pay for the increase in  $U_L$ . The following lemma formalises this intuition. Notice that also a zero-change in utility is possible.

**Lemma 6.** *If  $\alpha$  increases from  $\alpha'$  to  $\alpha''$ , then*

$$dU_L \geq 0 \geq dU_H.$$

*Proof.* From Definition 1 it follows, in particular,

$$\alpha'U'_L + (1 - \alpha')U'_H \geq \alpha'U''_L + (1 - \alpha')U''_H \Rightarrow 0 \geq \alpha'dU_L + (1 - \alpha')dU_H \quad (21)$$

$$\alpha''U''_L + (1 - \alpha'')U''_H \geq \alpha''U'_L + (1 - \alpha'')U'_H \Rightarrow \alpha''dU_L + (1 - \alpha'')dU_H \geq 0 \quad (22)$$

Summing up (21) and (22) yields  $dU_L \geq dU_H$ . Next, suppose by contradiction that  $dU_H > 0$ , then  $dU_L > 0$ , but that contradicts (21). It follows similarly that  $dU_L \geq 0$ .  $\square$

Lemma 3 shows that the incentive constraint of high-type agents binds at interior second-best allocations, given Assumption 1. Its proof does not work for corner solutions, though. By contrast, the following lemma does hold for corner solutions, albeit under more restrictive conditions on parameters.

**Lemma 7.** *Suppose  $\alpha \geq \gamma$ . If  $A$  is a second-best allocation, then the incentive constraint (3) for the high-type agents is binding at  $A$ .*

*Proof.* I first show that marginal utility is lower for high- than for low-type agents. Then I show that a marginal redistribution of  $C$  from high- to low-type agents increases welfare and hence needs to be ruled out by a binding incentive constraint. Otherwise the allocation cannot be second-best. In term of notation, recall that  $u^\theta = u(C_\theta, D_\theta)$ .

Claim:  $u_C(C_L, D_L) > u_C(C_H, D_H)$

Case 1,  $D_H \leq D_L$ : then  $C_H > C_L$ , because, by Lemma 2,  $u(C_L, D_L) < u(C_H, D_H)$ . Decreasing marginal utility and a positive cross derivative  $u_{CD} \geq 0$  then imply  $u_C(C_L, D_L) > u_C(C_H, D_L) \geq u_C(C_H, D_H)$ , hence the claim holds.

Case 2A,  $D_H > D_L, C_H \leq C_L$ : then, similar to Case 1,  $u_D(C_L, D_L) > u_D(C_L, D_H) \geq u_D(C_H, D_H)$ . At an optimal allocation,  $u_D^L/u_C^L = u_D^H/u_C^H$ , hence  $u_D^L > u_D^H$  implies  $u_C^L > u_C^H$ , as claimed.

Case 2B,  $D_H > D_L, C_H > C_L$ : as  $u$  is strictly concave,

$$\begin{aligned} u^H - u^L &< (C_H - C_L)u_C^L + (D_H - D_L)u_D^L, \text{ and} \\ u^L - u^H &< (C_L - C_H)u_C^H + (D_L - D_H)u_D^H \end{aligned}$$

need to hold.<sup>16</sup> Rearranging the second inequality gives, in combination with the first,

$$u_D^H(D_H - D_L) + u_C^H(C_H - C_L) < u^H - u^L < (C_H - C_L)u_C^L + (D_H - D_L)u_D^L.$$

$(C_H - C_L), (D_H - D_L) > 0$ , thus  $u_D^H < u_D^L$  or  $u_C^H < u_C^L$ . If one of these two inequalities holds, the other one must hold as well, otherwise  $u_D^L/u_C^L = u_D^H/u_C^H$  cannot be true. This completes the proof of the claim.

Now, suppose that the lemma is false, then  $u(C_H, D_H) - Y_H/w_H > u(C_L, D_L) - Y_L/w_H$  at a second-best allocation. Then there exist  $dC_H, dC_L$  with  $dD_H = -dC_L\gamma/(1-\gamma) < 0$ , such that the incentive constraint still holds, i.e. that  $u(C_H + dC_H, D_H) - Y_H/w_H > u(C_L + dC_L, D_L) - Y_L/w_H$ . The modified allocation  $(C_L + dC_L, D_L, Y_L, C_H + dC_H, D_H, Y_H)$  is also feasible by construction (and still satisfies low-type agents' incentive constraint). For  $dD_L \rightarrow 0$ , the change in welfare is approximately

$$dW \approx \alpha u_C^L dC_L - (1-\alpha)u_C^H dC_L\gamma/(1-\gamma) = dC_L (\alpha(1-\gamma)u_C^L - (1-\alpha)\gamma u_C^H). \quad (23)$$

If  $\alpha \geq \gamma$ , and  $u_C^L > u_C^H$  as claimed, then  $\alpha(1-\gamma)u_C^L - (1-\alpha)\gamma u_C^H > 0$  and welfare increases.

Hence, a contradiction.  $\square$

If the welfare function is utilitarian ( $\alpha = \gamma$ ) or exhibits an even stronger tendency to redistribute in favour of the low-type agents, high-type agents incentive constraint must

<sup>16</sup>In general, if a continuously differentiable function  $f$  is strictly concave over an open, convex subset of  $\mathbb{R}^n$ , then  $f(x) - f(x^0) < \sum_i f_{x_i}(x^0)(x_i - x_i^0)$ , for all  $x, x^0$  from that subset. See Sydsæter et al. (2008, Theorem 2.4.1) for a textbook reference.

be binding. This is not a necessary condition, though. Equation (23) shows that even with  $\alpha < \gamma$  a slack incentive constraint would be impossible, provided that  $u_C^L$  is sufficiently greater than  $u_C^H$ . In fact, I believe that the constraint is binding whenever Assumption 1 is satisfied.

**Conjecture 1.** *If  $A$  is a second-best allocation under Assumption 1, then high types' incentive constraint is binding at  $A$ .*

The following proposition is the main result of the current section. If conjecture 1 holds, the proposition and its two corollaries extend to all parameters satisfying Assumption 1.

**Proposition 6.** *Suppose  $\alpha' \geq \gamma$ . Let  $A(\alpha')$  be a second-best corner solution. Then  $A(\alpha'') = A(\alpha')$  for all  $\alpha'' > \alpha'$ .*

*Proof. Claim 1:* At a corner solution all agents have the same utility level.

By Lemma 7, high-type agents incentive constraint is binding. Adding  $-(\gamma D_L + (1-\gamma)D_H)e$  to the binding incentive constraint gives  $U_H(A(\alpha')) = u(C_H, D_H) - Y_H/w_H - (\gamma D_L + (1-\gamma)D_H)e = u(C_L, D_L) - (\gamma D_L + (1-\gamma)D_H)e = U_L(A(\alpha'))$ .

*Claim 2:*  $U_L(A(\alpha'')) = U_H(A(\alpha''))$ .

From Lemma 6,  $U_L(A(\alpha'')) - U_L(A(\alpha')) \geq U_H(A(\alpha'')) - U_H(A(\alpha'))$ . Given Claim 1 this reduces to  $U_L(A(\alpha'')) \geq U_H(A(\alpha''))$ . The incentive constraint of high types implies, though, that  $U_H(A(\alpha'')) \geq U_L(A(\alpha''))$ . Hence  $U_L(A(\alpha'')) = U_H(A(\alpha''))$ .

*Claim 3:*  $A(\alpha'') = A(\alpha')$ .

Suppose the opposite, then  $U_L(A(\alpha'')) > U_L(A(\alpha'))$  by Lemma 6 and the fact that solutions are unique (if they exist). But then Claims 1 and 2 imply that also  $U_H(A(\alpha'')) > U_H(A(\alpha'))$ . This contradicts Lemma 6.  $\square$

Increasing the welfare weight of low-type agents does not change the allocation if low-type agents already provide zero output. The only way to increase their utility is to increase their consumption. But then high-types incentive constraint can no longer be satisfied.

Thus the limits of redistribution (under information constraint) are met, once all output is produced by high-type agents:

**Corollary 1.** *Suppose  $\alpha' \geq \gamma$ . Let  $A(\alpha')$  be a second-best corner solution. Then the Rawlsian allocation  $A^R = \lim_{\alpha \rightarrow 1} A(\alpha)$  is equal to  $A(\alpha')$ . Also, the second-best Pareto-frontier has a kink at  $[U_L(A^R), U_H(A^R)]$  if  $\alpha' < 1$ .*

Notice that it is possible that the Rawlsian allocation is not a corner solution. Put differently, (second-best) redistribution can hit its very limit well before low-type agents provide zero output.

Yet, if for some  $\alpha < 1$ , low-type agents' output does equal zero, then, consequently, the comparative statics of Pigouvian taxation are also zero:

**Corollary 2.** *Suppose  $\alpha \geq \gamma$ . Let  $A(\alpha)$  be a second best allocation with  $Y_L = 0$ . Let  $t_D(\alpha)$  be the dirty good tax of a normalised tax system that implements  $A(\alpha)$ . Then  $t_D(\alpha') = t_D(\alpha)$  for all  $\alpha' \geq \alpha$ .*

## References

- AIGNER, R. (2011): "Environmental Taxation and Redistribution Concerns," *Preprints of the Max Planck Institute for Research on Collective Goods*, 2011/17.
- ATKINSON, A. B. AND J. E. STIGLITZ (1976): "The design of tax structure: Direct versus indirect taxation," *Journal of Public Economics*, 6, 55–75.
- BIERBRAUER, F. AND P. BOYER (2010): "The Pareto-Frontier in a simple Mirrleesian model of income taxation," *Preprints of the Max Planck Institute for Research on Collective Goods*, 2010/16.
- BOVENBERG, A. (1999): "Green tax reforms and the double dividend: an updated reader's guide," *International Tax and Public Finance*, 6, 421–443.

- BOVENBERG, A. AND R. DE MOOIJ (1994): "Environmental levies and distortionary taxation," *The American Economic Review*, 84, 1085–1089.
- (1997): "Environmental Levies and Distortionary Taxation: Reply," *The American Economic Review*, 87, pp. 252–253.
- BOVENBERG, A. AND L. GOULDER (2002): "Environmental taxation and regulation," in *Handbook of Public Economics*, ed. by A. J. Auerbach and M. Feldstein, Amsterdam: Elsevier, vol. 3, 1471 – 1545.
- BRETT, C. AND J. A. WEYMARK (2008): "Public Good Provision And The Comparative Statics Of Optimal Nonlinear Income Taxation," *International Economic Review*, 49, 255–290.
- CREMER, H., F. GAHVARI, AND N. LADOUX (1998): "Externalities and optimal taxation," *Journal of Public Economics*, 70, 343 – 364.
- DAHLBY, B. (2008): *The Marginal Cost of Public Funds: Theory and Applications*, The MIT Press.
- DIXIT, A. (1985): "Tax Policy in Open Economies," in *Handbook of Public Economics*, ed. by A. Auerbach and M. Feldstein, North-Holland, Amsterdam; New York: Elsevier, vol. 1, 313–374.
- EKINS, P. AND S. DRESNER (2004): *Green taxes and charges: reducing their impact on low-income households*, Joseph Rowntree Foundation.
- FULLERTON, D. (1997): "Environmental levies and distortionary taxation: Comment," *The American Economic Review*, 87, 245–251.
- FULLERTON, D. AND G. METCALF (1998): "Environmental taxes and the double-dividend hypothesis: Did you really expect something for nothing?" *Chicago-Kent Law Review*.

- GAUTHIER, S. AND G. LAROQUE (2009): "Separability and public finance," *Journal of Public Economics*, 93, 1168–1174.
- GOULDER, L. (1995): "Environmental taxation and the double dividend: a reader's guide," *International Tax and Public Finance*, 2, 157–183.
- GUESNERIE, R. (1998): *A contribution to the pure theory of taxation*, 25, Cambridge University Press.
- HAMMOND, P. (1979): "Straightforward individual incentive compatibility in large economies," *The Review of Economic Studies*, 46, 263–282.
- HELLWIG, M. (2010): "A generalization of the Atkinson-Stiglitz (1976) theorem on the undesirability of nonuniform excise taxation," *Economics Letters*, 108, 156?–158.
- JACOBS, B. (2012): "A Contribution to the Theory of the Marginal Cost of Public Funds," *working paper*.
- JACOBS, B. AND R. DE MOOIJ (2011): "Pigou Meets Mirrlees: On the Irrelevance of Tax Distortions for the Second-Best Pigouvian Tax," *CESifo Working Paper Series*, 3342.
- KAPLOW, L. (2012): "Optimal control of externalities in the presence of income taxation," *International Economic Review*, 53, 487–509.
- KOPCZUK, W. (2003): "A note on optimal taxation in the presence of externalities," *Economics Letters*, 80, 81–86.
- KOSONEN, K. (2012): "Regressivity of environmental taxation: myth or reality?" Taxation Papers 32, Directorate General Taxation and Customs Union, European Commission.
- MEADE, J. (1952): "External economies and diseconomies in a competitive situation," *The economic journal*, 62, 54–67.



- METCALF, G. (1999): "A distributional analysis of green tax reforms," *National Tax Journal*, 52, 655–681.
- (2003): "Environmental levies and distortionary taxation::: Pigou, taxation and pollution," *Journal of Public Economics*, 87, 313–322.
- MIRRLEES, J. (1971): "An exploration in the theory of optimum income taxation," *The Review of Economic Studies*, 175–208.
- OATES, W. (1995): "Green taxes: Can we protect the environment and improve the tax system at the same time?" *Southern Economic Journal*, 61, 915–922.
- OROSL, G. O. AND R. SCHÖB (1996): "Internalizing Externalities in Second-Best Tax Systems," *Public Finance*, 51, 242–57.
- PIGOU, A. (1932): *Economics of welfare.*, Macmillan. London.
- PIRTTILÄ, J. AND M. TUOMALA (1997): "Income tax, commodity tax and environmental policy," *International Tax and Public Finance*, 4, 379–393.
- RAUSCH, S., G. METCALF, AND J. REILLY (2011): "Distributional impacts of carbon pricing: A general equilibrium approach with micro-data for households," *Energy Economics*, 33, 20–33.
- SANDMO, A. (1975): "Optimal taxation in the presence of externalities," *The Swedish Journal of Economics*, 77, 86–98.
- SCHÖB, R. (1997): "Environmental taxes and pre-existing distortions: The normalization trap," *International Tax and Public Finance*, 4, 167–176.
- (2005): "The Double Dividend Hypothesis of Environmental Taxes: A Survey," in *The International Yearbook Of Environmental And Resource Economics 2005/2006*, ed. by H. Folmer and T. Tietenberg, Cheltenham: Edgar Elgar, 223–279.

SYDSAETER, K., P. HAMMOND, A. SEIERSTAD, AND A. STROM (2008): *Further mathematics for economic analysis*, Financial Times Prentice Hall.

THE ECONOMIST (2008): “Let them heat coke – How green taxes hurt the poor,” *The Economist*, June 12th, 2008.

WEST, S. (2005): “Equity implications of vehicle emissions taxes,” *Journal of Transport Economics and Policy*, 39, 1–24.

WEYMARK, J. (1987): “Comparative static properties of optimal nonlinear income taxes,” *Econometrica*, 55, 1165–1185.