On the Relative Effectiveness of Voluntary Agreements and Mandatory Regulation, when Consumers are Environmentally Conscious

Christos Constantatos *[1] and Apostolos Ioannis Martis[1]

[1]University of Macedonia, Thessaloniki, Greece

March, 2024

Abstract

We examine the impact of environmental policies in imperfectly competitive markets, when consumers care about the environmental damages of their consumption, and firms respond to consumers' worries using end-of-pipe abatement. We focus on the choice of tax-base between emissions and output, considering a flexible regulator who offers a Voluntary Agreement (VA) in the form of stick & carrot approach: the regulator offers a credible menu of abatement and tax-rates, and the firm decides the tax-rate by choosing its abatement. Except for some low-to-intermediate levels of consumers' social responsibility (SR) taxing emissions tax produces higher welfare than taxing output because the output-tax may lead to over-abatement. Emissions are lower under an emissions (output) tax when SR is low (high). Comparing VA to mandatory regulation with respect to welfare, we find that a VA on emissions tax is better than a mandatory tax-rate on emissions. This does not hold for the output tax, where the welfare comparison between mandatory regulation and VA depends on the level of SR. With respect to environmental protection, a VA provides lower emissions when applied on output, whereas mandatory regulation when applied on emissions.

JEL Classifications: D62, H21, L12
Keywords: Green consumers, Environmental taxes, Monopoly.

*Corresponding Author. Email: cconst@uom.edu.gr

1 Introduction

Consumer sensitivity over environmental issues ties consumption decisions to producers' environmental performance. This creates incentives for firms to ameliorate their environmental performance in order to increase demand. However, due to the externality nature of the problem, consumer awareness may not suffice in order to attain sufficiently clean environment, thus calling for regulatory intervention. We examine two such types of intervention, emissions and output taxes and compare their efficiency when the environmental damage caused by a good's production can be reduced at the expense of a fixed cost, like end-of-pipe abatement (use of filters). Since fixed investments go in pair with imperfect competition, we focus on cases where firms have market power.

In this paper we consider situations where the regulator does not commit to a given tax rate before firms make their long-run abatement decisions. First, we examine the relative efficiency in terms of welfare of two potential tax bases, emissions and output. Second, assuming that the regulator has commitment power, we analyze whether an ex-ante commitment to a tax rate produces higher welfare, compared to the non-commitment case. Crucial parameter in all our analysis is consumers' environmental consciousness and social responsibility.

One may think that, when emissions bear a linear relation to output, if the firms have already made their long-run decision about abatement at the moment the tax rate is imposed the choice of tax-base is irrelevant. Because emissions reductions can only be obtained through output reductions, any feasible result can be obtained using either base and applying the appropriate rate[1]. Hence, only when the tax is imposed before abatement decisions, as in Constantatos et al. (2021) (hereafter, CPS) the choice of the tax-base may be of importance.

The above argument ignores that in many instances, when making their abatement decision firms *anticipate* the upcoming taxation and therefore the expected tax rate affects their decision. Our results show that a crucial factor determining abatement is whether firms expect the tax to be on output or on emissions.

Let us for the moment assume that the regulator announces a tax base but not a tax rate, and ask *why firms may be faster in adopting abatement than the regulator choosing the tax rate?* First because designing and voting a new tax may be a slow process, sufficiently long from its initial inception to the final vote as to provide firms with the necessary time to adjust their abatement level before a tax-rate is finally adopted.[2] If the regulator cares for total welfare rather than only better environment,

[1]For instance, if the emissions-to-output relation is one-to-one, the same applied to either base obtains the same level of welfare. If each unit of output generates h units of emissions, a tax-rate on emissions h times the tax-rate on output leads to the same real outcome (consumption, abatement, emissions, welfare).

[2]Situations where the regulator is slow to react and/or lacks commitment power can also be found in J. Poyago-Theotoky & Teerasuwannajak (2002), Petrakis et al. (2005), Requate (2005), and J. A. Poyago-Theotoky (2007).The first paper deals with the impact of taxation on product differentiation, the second with the importance of the number of firms, while the other two with the impact of tax on environmental R&D in presence of R&D spillovers. None of these papers considers

firms know that, by increasing their abatement they can induce the regulator to impose a lower tax rate. Second, because the regulator *may prefer* to avoid committing into a tax rate before firms do commit in their abatement decision, in order to create higher abatement incentives. While traditional mandatory taxation consists of the regulatory authority unilaterally designing and imposing a tax, voluntary agreements (VA) allow the regulated subjects to have a say on the tax terms.[3] Usually, a VA allows firms to obtain lower tax rates (or even subsidies) in exchange for higher investment in abatement. Such agreements are always exposed to opportunistic behavior by the second mover. By reserving this position to himself, the regulator avoids being deceived, but becomes potential deceiver, since he is able to first make a promise as to induce a given behavior and then adopt a tax rate different than the promised one. Thus unless one assumes commitment power to one or both parties, only VA containing time-consistent promises can be credible.

The above discussion reflects upon modeling choices. Mandatory taxation is modeled as a multi-stage game where at the first stage the regulator announces the regulation terms and at later stages firms make their decisions (see for instance Constantatos et al. (2021)). VA on the other hand, are often modeled as a bargaining process between the regulator and the firm (see Segerson & Miceli (1998)). However, given the sunk-cost nature of end-of-pipe abatement investments and the fact that the regulator has the power to unilaterally modify the tax rules at any moment, both approaches ignore that a regulator's initial decision is most likely time-inconsistent: the regulator may wish to renege on his initial decision after the firm has undertaken the expected action. Only when the regulator possesses substantial commitment power initial decisions and/or agreements can be credible.

In this work we assume that the regulator first announces a tax base, then firms react by choosing abatement (their long-run decision), after observing abatement the regulator determines a tax rate on the announce base and then firms choose their output, and indirectly, their level of emissions. At this point one might ask the question, *why the regulator can commit to a tax base when he is unable to commit to a tax rate?* The answer lies on the argument of the third paragraph above: once end-of-pipe abatement is determined, the regulator is indifferent between tax bases. Since there is no incentive to deceive firms upon this issue, an announcement determining the tax base at the beginning of the political process of tax voting can be considered credible, requiring significantly less commitment power than a corresponding tax-rate announcement (and eventually none).

Our work shows that in an imperfectly competitive environment, as consumer consciousness increases, the abatement incentive provided by an output tax increases, while that provided by an emissions tax decreases. Increases in consumer consciousness generally call for lowering the tax rates, whether on emissions or output, since consumers internalize part of the environmental externality by reducing demand. A reduction of the output-tax rate works like an output subsidy, boosting production. Since abatement is a fixed cost, higher output lowers the cost of abatement per-unit

environmental consciousness and/or output taxation.

[3] "VAs are designed to link private benefits with the voluntary provision of public goods" (Delmas & Montes-Sancho, 2010)

of output, naturally providing a greater abatement incentive[4]. On the other hand, a reduction of the emissions tax-rate due to an increase in consumer consciousness corresponds to making emissions less costly, thus curtailing the firm's incentive for abatement. Thus, the efficiency of an environmental tax based on emissions decreases as consumers become more environmentally conscious.

The aforementioned interaction between environmental consciousness and the efficiency of each tax base is also present in cases where the regulator does commit to a tax-rate, as shown in CPS. However, when the regulator is unable or unwilling to commit, the factors determining this interaction are coupled with an additional abatement incentive, due to the firm's effort to affect/manipulate the tax-rate via its choice of abatement. We show that as consumer consciousness increases and the tax rates are reduced in consequence, the tax menu exacerbates the propensity of the output tax towards higher abatement, while it moderates the emission-tax's tendency towards lower abatement. As a consequence, no matter the level of consumer environmental consciousness, the emissions base yields higher welfare when the regulator offers a tax-menu instead of committing to a single tax rate. For the output base the conclusion is mixed: for low levels of consciousness the tax-menu yields higher welfare, but as consciousness increases the over-abatement effect kicks-in causing welfare to decrease rapidly (this result is true when output subsidies are allowed).

Categorizing environmental taxes according to a) their base, and b) whether there is or not commitment to a given rate, our analysis shows that a menu of emissions tax-rates based on observed abatement (i.e., an emissions tax without commitment to a tax rate) represents for most levels of consumer consciousness the second-best alternative to full regulation (i.e., regulation of both, output and abatement). Even for the small interval of consciousness levels for which an emissions tax without commitment does not represent the second best, it constitutes the third-best alternative producing welfare levels slightly lower than that of the second-best. Thus, if the level of environmental consciousness cannot be accurately measured, offering a tax-rate menu based on observed abatement represents perhaps the safest regulation.

In terms of policy, our comparisons show that when SR is very little developed, a VA on emissions tax is the best instrument for reaching both, high welfare and clean environment. For low-intermediate levels of SR, a VA on output tax is the best instrument for total welfare and the environment. For, intermediate levels of SR, a VA on output tax may produce over-abatement, leading to very clean environment but low welfare; in such cases a VA on emissions tax is again the best instrument. When SR is sufficiently high, mandatory taxation on output is the best instrument both in terms od welfare and environmental protection (see also CPS)

work stands in the junction of three strands of the literature. The first deals with the comparison between emissions and output taxes, the second with the importance of environmental consciousness in affecting behavior and tax efficiency, and the third concerns the commitment power of the regulator to a given tax rate when facing new

[4]The optimal output tax may become negative (output subsidy), while we do not allow for emissions subsidies. This, however, it is not what is crucial, since increases in output are caused by reductions in the output tax rate, whether that rate is positive.

abatement conditions.

The discussion of negative externalities in most textbooks leads to the Pigouvian tax levied either on emission or output, assuming, explicitly or implicitly, that output and emission taxes are equivalent. This equivalence is based on the assumption that the amount of emission produced per unit of output is immutable, ignoring the realistic possibilities of engaging in abatement. Recognizing that a given level of output may yield different levels of emissions, breaks up the equivalence of output and emission taxes. A significant literature has been developed examining the optimal choice of environmental tax instrument in different settings. Schmutzler & Goulder (1997) compare emissions and output taxes in a partial equilibrium framework and in the presence of imperfect monitoring of emissions, while Fullerton et al. (2001) and Cremer & Gahvari (2001) in a general equilibrium framework. Goulder et al. (1997) examine the interactions with pre-existing distortionary taxes. Aoyama & Delfino Silva (2016) compare the effectiveness of output and emission taxation in promoting the adoption of advanced abatement technology without addressing the effect of consumers' environmental awareness. Closest to our work, Constantatos et al. (2021) compare the two tax bases in terms of welfare taking into account environmental awareness, but only considering the case where the regulator can commit to a single tax rate.

The increasing importance of green consumerism has raised the question of the appropriate adjustments to the traditional environmental tax and subsidy policies and furthermore initiated a discussion regarding the effectiveness of information campaigns and advertising, aiming at increase environmental awareness, as an additional policy instrument (see for example Petrakis et al. (2005), Nyborg et al. (2006), Brouhle & Khanna (2007) and Sartzetakis et al. (2012)). The literature has approached the emergence of green consumers using different frameworks. Most of the models assume that green consumers differentiate products based on their environmental attributes inducing some firms to produce a "greener" variety of the product. This differentiation has been examined mainly within a framework of vertical differentiation (Bansal & Gangopadhyay (2003), García-Gallego & Georgantzís (2009), Bansal (2008), Deltas et al. (2013) and Doni & Ricchiuti (2013)), and less within a framework of horizontal differentiation (Conrad (2005)). Alternatively, Gil-Moltó & Varvarigos (2013) examine the case in which environmental consciousness leads consumers to devote resources to reduce pollution (participation in carbon offsetting schemes, donations to NGOs, etc).

In all the aforementioned papers, consumers' involvement is due to *a change in their tastes* that makes environmentally friendly products more highly valued (*hedonic* approach). Our approach is to assume a *change in consumers' behavior*: an environmentally responsible consumer is one that decides her consumption by taking also account of at least a part of the externality she creates. Thus, pro-environment behavior is not the result of a "warm-glow" effect that attributes higher utility to certain behavior; instead, it results from a conscious deviation from strict utility maximization, in order to partially account for a damage the consumer creates without immediately receiving its consequences (we call it *social responsibility* (SR) approach). The major difference between the two approaches is that increases in SR leave the first-best solution unaffected, whereas changes in consumers' tastes impact the ideal outcome. For instance, in the case of SR, full internalisation of the externality simply leaves no role for regulation, whereas according to the hedonic approach, changes in the warm-glow effect that affects consumers' preferences call for constant regulatory intervention (more on this issue furhter down in the text).

The rest of the paper is organized as follows: section 2 presents the model, section 3 derives the equilibrium with each type of tax under the no commitment assumption, section 4 compares the results of section 3 to those obtained in CPS under tax-rate commitment, and section 5 concludes.

2 The Model

2.1 Production, Pollution and Consumption

Assume a monopolist producing Q units of a product X with constant marginal production cost, normalized to 0 for simplicity. In the production process, δ units of harmful pollutant are generated per unit of output. In order to reduce its emissions, the firm invests in end-of-pipe abatement technology, such as filters.[5] Since filters represent fixed costs, the cost function is constant in quantity, but assumed quadratic in abatement:

$$C = kv^2, \quad k \geq 1, \tag{1}$$

where v is the amount of emissions reduction the firm undertakes in order to improve the abating technology. When choosing the level of abatement, the firm generates the following total amount of net emissions:

$$e = \delta Q - v, \quad \delta > 0, \tag{2}$$

where Q is the total output produced by the firm. In order to avoid negative net emissions the level of abatement must be restricted to the $[0, \delta Q]$ interval. The total environmental damage that the firm generates takes the form of:

$$D = ze^2, \quad z > 1, \tag{3}$$

where z shows how convex the damage function is. The parameter z may be interpreted as either a technical, or a social-preference parameter expressing pollution damage in terms of equivalent reductions of the numeŕaire good.
On the consumption side of the model, assume $n \geq 1$ identical consumers, each having the following utility function:

$$U(q_i) = \alpha q_i - \frac{1}{2}q_i^2 + M \quad \alpha > 0, \tag{4}$$

with $q_i \geq 0$ representing individual consumption of X, $\sum_{i=1}^{n} q_i = Q$, and M being the amount of the numéraire good consumed; hereafter, M is dropped.
Social consciousness implies that consumers deviate from strict utility maximization,

[5]There are also other options of abating pollution, including in the "cleaner production" method such as "the re-circulation of materials, the use of environmentally friendly materials and the modification of the combustion chamber design"(Frondel et al. (2007)).

adopting a behavior that is guided by maximization of the following function:

$$\tilde{U}(q_i;\phi,e)=\begin{cases}(\alpha-\phi e)q_i-\frac{1}{2}q_i^2, & \text{if } e>0\\ \\ U(q_i), & \text{if } e\leq 0\end{cases}, \tag{5}$$

where ϕ shows consumers' aversion to pollution. When $\phi=0$, consumers do not care about net emissions generated by the firm and therefore decide their purchase of X by maximizing $U(q)$ in (4). For positive values of ϕ, the representative consumer feels responsible for only part of the externality. In order to avoid over-internalization, we bound ϕ in the $[0,z]$ interval.
After maximizing (5) with respect to quantity, substituting e by $\delta Q-v$ from (2), solving for q_i and aggregating, we obtain aggregate demand function for the case of n consumers:[6]

$$Q(p;n,\delta,\phi,v)=\frac{n(\alpha+\phi v)}{\phi n\delta+1}-\frac{n}{\phi n\delta+1}p$$

Without loss of generality, we assume hereafter a single representative consumer. The inverse demand function that emerges from (5) when $n=1$ is:

$$p(q;n,\delta,\phi,v)=\begin{cases}(\alpha+\phi v)-(1+\phi)q, & \text{if } q\geq v\\ \\ \alpha-q, & q\leq v\end{cases}, \tag{6}$$

and is represented by the line aAq_2 in Figure 1, for given level of abatement, v. The line D_U corresponds to the demand deriving from (4) while the line $D_{\tilde{U}}$ to that deriving from the first part of (5). At point A, net emissions are zero, therefore $\delta q_A=v$.

In our approach the responsible consumer sacrifices utility in order to contribute to environmental quality (a public good). The sacrifice corresponds to the utility loss due to maximizing (5) instead of (4). For instance, at price p_1, strict utility maximization implies the purchase of q_2 units, but the consumer voluntarily restricts her consumption to q_1, depending on her ϕ, which determines the position of $D_{\tilde{U}}$. The presence of ϕ makes the consumer behave as if a corrective per unit tax equal to to BE were imposed to the good, although no tax payment is actually required.[7] Since the quantity restriction q_1,q_2 is not related to a change in the marginal valuation (marginal benefit) function of good X, the total value of the actual purchase is the shaded area $0ABq_1$.

Our *Social Responsibility* (SR) approach differs from *Hedonic* approach models that assume pro-environment action to emanate from *personal satisfaction* considerations. According to those models, the consumer reduces her consumption of X because, once social consciousness makes her see the consequences of such consumption, the good no longer provides her the same utility as before. In those models, total and marginal valuation of X in terms of the numéraire are obtained using $\tilde{U}$ contained in (4). Thus,

[6]The substitution of e must take place *after* deriving the utility function with respect to q_i in order to capture the fact that in the absence of social consciousness, no consumer takes into account the impact of her consumption on total quantity and emissions

[7]The amount BE corresponds exactly to the Pigouvian tax if $\phi=z$

the line $aAEq_2$ in Figure 1, not only shows consumer behavior, but also corresponds to the marginal benefit from each unit consumed. If consumer behavior is guided by *hedonic* considerations, the total value of the q_1 units is $0aEq_1$.[8] In contrast to this "pleasure guided" approach, we maintain that social responsibility modifies behavior without modifying the value of the good.

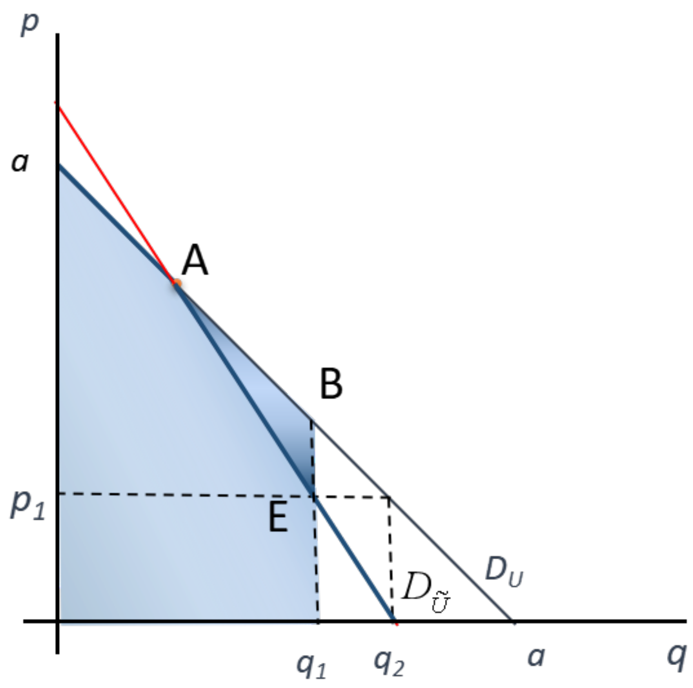

Figure 1: The Demand Function

2.2 Social Welfare and First Best

The Social Welfare function consists of consumers' utility from consumption, minus the sum of all costs, namely, environmental damage, production cost (assumed zero) and the cost of abatement, if any:

$$W = u_i - (D + C)$$

The question here is whether consumption utility must be replaced by (4) or (5). According to the discussion surrounding Figure 1, while consumers' behavior is described by the demand function aAE, which derives from (6), actual consumer surplus for any given quantity is represented by the area below the line aa, which derives from

[8] In most cases, instead of a marginal value reduction due to pollution, hedonic models introduce a "warm-glow" effect from the use of a more environmentally friendly good. In this case, the utility function takes the form $U_H = a + \theta v q_i - (1/2)q_i^2$, where θ is the degree of consciousness, and v an environment-friendly characteristic, here abatement. The corresponding inverse demand is $p = a + \theta v - q_i$, *i.e.*, instead of a negative reaction to emissions, the consumer reacts positively to their restriction (see Lambertini (2017)). Whether the change in consumer behavior is due to emissions-aversion or a "warm-glow" effect from their reduction, the point is that it is related to changes in personal satisfaction, instead of emanating from a sense of duty that calls for a voluntary deviation from strict utility maximization

(4). Hence, while we are using the demand function in (6) in order to determine the quantity at any given price, we use the utility function in (4) in order to determine the value this quantity has for the consumer. The use of (5) in order to compute consumption benefits would imply the inclusion of a psychological damage stemming from emissions aversion *per se* on top of the real environmental damage, which is already captured by the D term in the social welfare function.
Substituting (1), (3) and (4) to the social welfare function, we obtain:

$$W = \alpha q - \frac{1}{2}q^2 - z(q-v)^2 - kv^2. \tag{7}$$

We assume that $z \geq \alpha$, implying that the relative importance of the environmental damage is sufficiently high, and set $\alpha = 1$, a mere normalization. In what follows, we present the first-best solution taking into account all the simplifying assumptions made up to this point, namely that $\delta = 1$, $n = 1$, $a = 1$, and also assuming hereafter $k = 2$[9]. Maximizing (7) with respect to abatement and quantity, we obtain:

$$q^* = \frac{z+2}{5z+2}, \qquad v^* = \frac{z}{5z+2},$$

where a "*" indicates first-best values. It can be easily verified that $\forall z \geq 1$, $q^* > v^*$, so net emissions are never negative. Substituting the above expressions into (3) and (7), we obtain the first-best level of emissions and social welfare:

$$ne^* = \frac{2}{5z+2}, \qquad sw^* = \frac{z+2}{10z+4}.$$

Note that the social responsibility approach implies that none of the first-best expressions contains ϕ, thus the maximum achievable level of welfare is independent of consumer consciousness, since the regulator is able to directly control both, consumption and abatement.

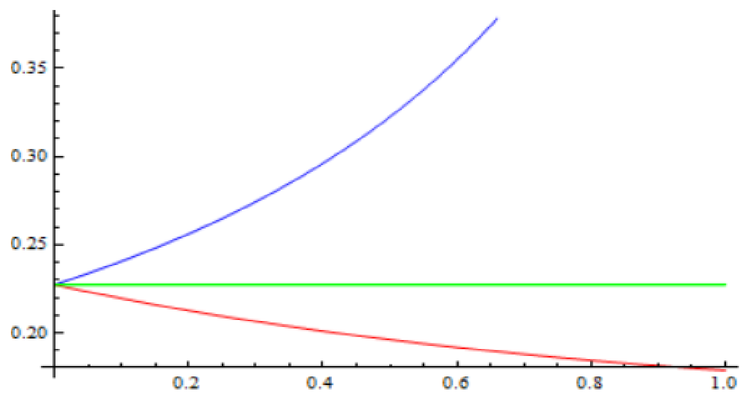

Figure 2: First Best Welfare under Various Approaches of Social Consciousness

All three lines on Figure 2 represent the maximum attainable welfare as function of the social consciousness parameter ϕ and have been drawn for $\alpha = \delta = 1$ and $k = z = 2$.

[9]This value has been chosen to avoid corner solutions when the tax is based on output, as well as to always have 2^{nd} order conditions satisfied. Unless for very small values of k–significantly less than 1, changes is k produce no qualitative results

The cyan horizontal line shows the first best attainable welfare according to the Social Responsibility (SR) approach, where the first-best is independent of ϕ. The magenta line corresponds to first-best welfare if one uses (5) instead of (4), into (7). The decreasing shape of the magenta line shows that as consumers' environmental concerns increase, a regulator in control of both crucial variables is able to generate lower welfare. This is due to the fact that when social consciousness emanates from *hedonic* considerations, the consumption of good X, besides pollution damages, creates also an additional (psychological) burden which increases with consumers' environmental sensitivity. While in some cases consumer behavior may indeed stem from hedonic considerations, we believe that such cases do not truly correspond to SR.[10] Finally, the yellow line shows again first-best welfare under a different type of hedonic considerations, this time assuming that the consumer derives pleasure from consuming goods of higher environmental quality (lower emissions); as a result, first-best welfare *rises* with ϕ. Again, we consider this approach as incompatible with SR in the sense of social duty.[11]

3 Time-consistent Regulator Game

In order to allow for tax-base comparisons, we adopt a four-stage game where at the first stage the regulator chooses a tax-base (either on emissions or on output), followed by the monopolist who chooses abatement at the second stage. At the third stage the regulator decides the tax-rate and at the final stage the monopolist chooses output. After that, consumers purchase the amount of X they wish, and profits and social welfare are determined; we proceed by backwards induction.

Our game structure is almost similar to that in Petrakis & Xepapadeas (2001) and J. Poyago-Theotoky & Teerasuwannajak (2002) with the addition of a first stage where the tax base is announced. The addition of the tax-base choice stage raises the question of whether the regulator, who may not have the power to commit to a tax rate, has the power to commit to a base. Recall, however, that when the emissions-to-output ratio is given, both bases produce the same output and welfare: the choice of tax base matters only to the extent that it affects abatement. Thus, once abatement is undertaken, the regulator has no incentive to renege on its tax-base announcement, unlike his announcement on tax-rate.[12]

The firm's profit functions in case of output and emissions tax are respectively:

[10] The horizontal axis in Figure 2 is limited to $[0, 1]$ since beyond $\phi = 1.0$ the magenta line includes negative net emissions.

[11] The difference between SR hedonic approaches is also important when one considers government's effort to increase consumers' environmental sensitivity. According to SR, the government must invest in order to convince consumers *to change their behavior*, whereas hedonic approaches require investments aiming to *change consumers' preferences*

[12] Since the regulator is indifferent, an ε reputation cost for reneging on his promise suffices to establish credibility in his first-stage tax-base announcement. Note also that while for given abatement the choice of tax-base does not affect total welfare, it may affect its distribution between surplus and profit. Distributional issues lie outside the paper's scope since they require modifying the social welfare function in order to weight profits and consumer surplus differently.

$$\Pi = (p-t)Q - C, \qquad \Pi = pQ - te - C \quad .$$

By substituting (1), (2) and (6) to these profit functions we obtain:

$$\begin{aligned} \Pi_O &= [(1+\phi v) - (1+\phi)q - t]q - 2v^2, \\ \Pi_E &= [(1+\phi v) - (1+\phi)q]q - t(q-v) - 2v^2, \end{aligned} \tag{8}$$

where the subscripts denote the output or emissions tax case. Maximization of either Π_O or Π_E yields the same optimal output expression, namely:

$$q_{3i} = \frac{1-t+v\phi}{2+2\phi}, \quad i = O,\ E. \tag{9}$$

The first digit of the subscript denotes the game-stage (3rd), while the second, the tax regime (O for output and E for emissions tax). Note that for given abatement and tax-rate, the output decision is the same under the two tax regimes[13]. Quantity is obviously increasing in abatement and decreasing in the tax rate. A bit less obvious but straightforward is to show that when output and net emissions are non-negative, third-stage quantity is decreasing in ϕ, as expected.
At the second stage, the social planner maximizes (7) taking the quantity of the firm as given, to obtain:

$$t_{2i} = \frac{-2vz(\phi+2) + (v-2)\phi + 2z - 1}{2z+1}, \quad i = O,\ E \tag{10}$$

For *given* abatement level, as long as $\delta = 1$ the regulator's optimal tax rate is the same, independently of the tax base upon which it applies.[14] Hence, the regulator has no incentive to change an already announced tax-base. It can be easily shown that the optimal tax rate is decreasing in the abatement level.

3.1 Subgame Perfect Nash Equilibrium

At the second stage the firm substitutes t and q given by their expression in (9) and (10), respectively, into (8) and maximizes with respect to abatement. We limit The thus obtained expressions hold only if at the resulting equilibrium net emissions are non-negative:

Lemma 1. *The abatement in the interior solution equilibrium, under output and emissions tax, is respectively:*

$$\begin{aligned} v_O &= \frac{z(1+\phi)}{2z(-z\phi+z+2)+1}, \\ v_E &= \frac{4z(z+1)-1-2\phi}{24z(z+1)+4-2\phi} \end{aligned} \tag{11}$$

[13]The perfect equality is due to the fact that we have assumed a one-to-one relation between output and emissions ($\delta = 1$). If $\delta \neq 1$, in the case of emissions tax, the tax coefficient in the above expression must be multiplied by δ

[14]For $\delta \neq 1$, the output tax must be multiplied by δ

Proof. The derivation of the v_O and v_E expressions is straightforward. First substitute (9) and (10) to (8) and then take the first derivative of it. □

The abatement levels in (1) only hold for interior solutions, yet there are two types of corner solutions that may prevail. First, for a given tax type and SR level, the abatement computed in (1) may not be higher than actual emissions. Since we assumed $\delta = 1$, if abatement is greater than output, net emissions must be considered equal to zero. Second and most important, corner solutions may occur because we rule out subsidies.[15] Thus, for any v_i value, $i = O, E$, computed by (1) to hold, it must be that when inserted into (10) the resulting tax rate is non negative. When this is not the case, the firm must modify its decision away from the rule described in (1). In the appendix it is shown that

Lemma 2. *A necessary and sufficient condition for interior solution in the case of output and emissions tax is $\phi_O \geq 0$ and $\phi_E \geq 0$, respectively, with*

$$\phi_O = \frac{(2z^2 + 3z + 2) - (1 + 2z)\sqrt{4 + z^2}}{2z} \tag{12}$$

and

$$\phi_E = \frac{7}{4} + z(6 + z) - \frac{1}{4}(1 + 2z)\sqrt{81 + 4z(11 + z)} \tag{13}$$

with $\phi_O < \phi_E$.

Proof. See Appendix. □

Hereafter, the analysis is limited in interior solutions, i.e., we only consider values of $\phi \in [0, \phi_i]$.

Proposition 1. *(i)If the announced tax is on output, a) $\forall \phi \leq \phi_O$ abatement is monotonically increasing in ϕ; b) there exists a value of $\phi = \phi^* \in [0, \phi_O]$*

$$\phi^* \triangleq \frac{z - 1}{z + 2} \tag{14}$$

such that the market produces the first-best level of abatement ($v_O = v^*$), while for $\phi > (<) \phi^*$ market abatement is above (below) v^*. [16]
(ii) If the announced tax is on emissions, for all $\phi \in [0, \phi_E]$ abatement is monotonically ***decreasing*** *in ϕ and $v_E < v^*$.*

Proof. See Appendix. □

It is remarkable that, while with an output tax increases in environmental consciousness result in higher abatement (as expected), the opposite holds true with an emissions tax: as ϕ increases, the firm abates less.[17] At any tax rate, more aware consumers internalize a larger part of the externality by demanding a lower quantity at any given price (social responsibility–SR–effect). Hence, increases in ϕ cause optimal tax rates to fall.

[15] While output subsidies are not unthinkable, emissions subsidies represent a politically unpalatable measure.

[16] In the main text, we usually refer to $\phi_O > \phi^*$ as the *over-abatement* effect

[17] A similar result for the case where the regulator can commit to a tax rate is also shown in CPS.

When the tax base is on output, tax-rate reductions boost output,[18] thus allowing for a reduction in the per-unit-of-output cost of abatement,[19] which in turn induces higher abatement (output effect). This output effect works in the same direction with the SR effect and reinforces it. On the other hand, the reduction of the emissions tax due to an increase in ϕ makes emissions less costly and tends to reduce abatement. This effect dominates the SR effect, therefore an increase in ϕ leads to lower abatement.

Figure 3 illustrates the content of proposition 1. It displays abatement under an output tax (cyan line) and emissions tax (magenta line), as functions of social consciousness, assuming $k = z = 2$, and $\delta = n = 1$. The horizontal green line represents the abatement level at the first-best solution. The gray-dotted line shows the point where the output tax becomes zero (ϕ_O) and the black-dashed line displays the level of environmental consciousness where the emissions tax becomes zero (ϕ_E). This color code–green for first-best, cyan for output tax, and magenta for emissions tax–as well as the aforementioned set of specific values for the exogenous parameters are maintained in all subsequent figures.

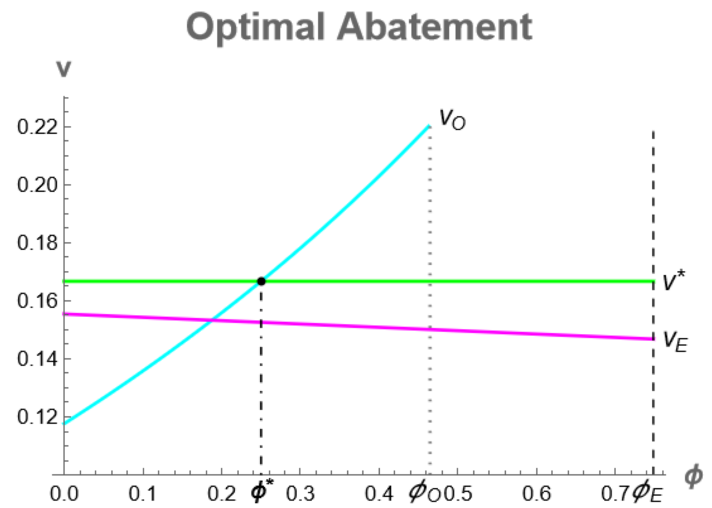

Figure 3: Equilibrium Abatement

Like all first-best values, first-best abatement is independent of ϕ. The cyan line representing v_O is monotonically increasing, eventually reaching and exceeding v^*. On the contrary, the magenta line (v_E) is decreasing and lies everywhere below the green line. For SR increases on the admissible levels of ϕ, abatement reductions are moderate because the monopolist knows that higher abatement leads ceteris paribus to a lower emissions-tax rate. Moreover, the over-provision of environmental quality, even beyond the first-best level, leads to a substantial reduction in the tax rate, thus the output tax becomes zero at a lower level of SR compared to the emissions tax ($\phi_O < \phi_E$). This is a consequence related to the *over-abatement effect.*

Turning to quantities, we substitute optimal abatement from Lemma 1, as well as

[18] Even if the output tax rate is positive, a tax-rate *reduction* corresponds to quantity subsidization.

[19] Recall that end-of-pipe abatement represents a fixed cost with respect to output

the resulting optimal tax rates (see the appendix) into (9). We obtain the optimal quantity under each type of tax as:

$$q_O = \frac{2z+1}{2z(-z\phi+z+2)+1},$$
$$q_E = 1 - \frac{5z(2z+1)}{12z(z+1)-\phi+2} \tag{15}$$

One can easily verify (see the appendix) that when the tax-base is output, higher environmental consciousness leads to higher quantity in equilibrium, whereas when the tax is on emissions, the equilibrium quantity decreases in ϕ. This happens because, *ceteris paribus*, higher values of ϕ reduce demand, but while this reduction is compensated by increased abatement in the case of output tax, the lower abatement in case of emissions tax depresses the demand curve further down.

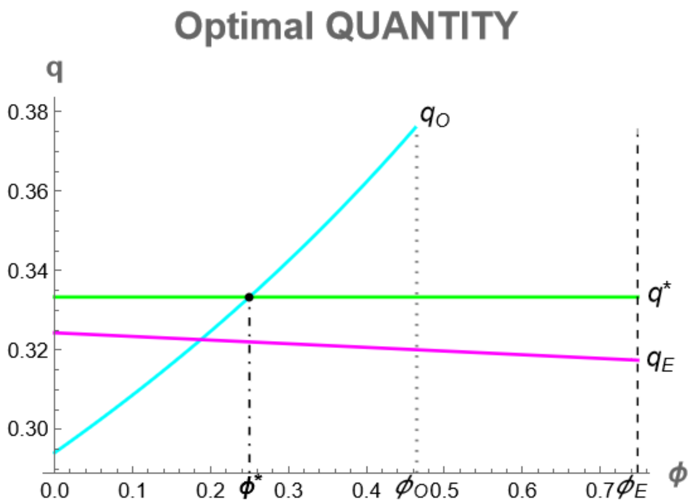

Figure 4: Equilibrium Quantity

The evolution of net emissions as ϕ increases is not obvious since under either tax-base, quantity and abatement move in the same direction, both increasing in the case of output tax, and decreasing in the case of emissions tax. Therefore, it depends on the relative magnitude of the abatement and the output. The following proposition clears this issue. Let $ne_i = ne_i(\phi; z)$, $i = O, E$, describe net-emissions in the interior solution equilibrium under the corresponding tax-base, and an accent denote derivative with respect to ϕ. Then:

Proposition 2. *(i) When the tax is on output, $ne'_O < 0$, with $ne_O(0) > ne^*$ and $ne_O(\phi_O) < ne^*$, i.e., net emissions decrease monotonically with ϕ, starting from a level above that of first-best when SR is zero and ending at a level below it, as ϕ approaches its highest allowed value; at the SR level $\phi = \phi^*$ (see (14)) net emissions are equal to their first-best level.*
(ii) When the tax is on emissions, $ne'_E > 0$, with $ne_E(0) > ne^*$, i.e., net emissions increase monotonically with ϕ and are always above their first-best level.
(iii) For relatively low (high) levels of SR, $ne_O > (<) ne_E$

Proof. See Appendix. □

The content of the proposition can be seen on Figure 7, contained in the next section. In the absence of environmental consciousness, an emissions tax yields a cleaner environment than an output tax. However, as ϕ increases, emissions keep increasing under the optimal emissions tax while they are drastically reduced under the optimal output tax. As a result, in a jurisdiction with relatively developed SR, a VA using output taxation initiates lower emissions compared to a VA on emissions taxation. A VA using emissions taxation is superior only when SR is very low. Also, for even moderate levels of SR, the VA using output taxation results in net emissions that are even lower than those at first-best.[20] It is remarkable that this excellent environmental performance of an output tax without regulator's commitment is not the result of a reduction in the quantity of the regulated good, but on a sharp increase in abatement, resulting from increases in SR. The cost of such performance is paid in terms of the numéraire good, *i.e.,* by withdrawing too many resources from the rest of the economy.

Equilibrium social welfare under each tax-base is obtained by substituting (11) and (13) into (7). The resulting expressions are relegated to the appendix, since they are fairly complex and provide no immediate intuition. At this point we simply denote them as $sw_i(\phi; z)$, $i = O, E$, the equilibrium welfare under each tax-base.

Lemma 3. *(i) Under an output tax,* $\forall \phi \leq (\geq) \phi^*$, $sw'_O \geq (\leq) 0$, *with* $sw_O(\phi^*) = sw^*$. *(ii) Under an emissions tax, the welfare function is monotonically decreasing in* ϕ*; an emissions tax never achieves first-best welfare, i.e.,* $\forall \phi$, $sw_E < sw^*$

Proof. See Appendix. □

Figure 4 shows the evolution of social welfare with increases in SR when an output or emissions tax is in effect (cyan and magenta line, respectively). As explained earlier, increases in ϕ cause reductions of the optimal tax rate due to *ceteris paribus* lower consumption. Reductions in the emissions tax rate make emissions cheaper, resulting in lower abatement and higher net emissions, which leads in turn to further reductions in consumption. As a result, increases in SR lead society to situations with more pollution and lower consumption; despite abatement-cost savings, total welfare decreases as ϕ increases. On the other hand, reductions in the output tax rate act as output boosting subsidies, countering the initial quantity reduction along with increases in abatement, finally resulting in net-emissions reduction. However, the total cost of abatement also increases and for $\phi > \phi^*$ attains sub-optimally high levels. This implies lower-than-optimal consumption of the numéraire good, causing total welfare to decrease with ϕ when SR reaches levels above ϕ^*. The following proposition formally compares welfare under output and emissions taxes.

Proposition 3. *For all* $z \geq 2$*, there exist two critical values of* ϕ*,* $\phi_1 < \phi_2$*, such that* $sw_O(\phi_i) = sw_E(\phi_i)$*,* $i = 1, 2$*. For values of* ϕ *within (outside) the* (ϕ_1, ϕ_2) *interval,* $sw_O > (<) sw_E$.[21]

[20] They may even become zero if the VA includes output subsidies for high environmental performance. Such a complete environmental clean-up is impossible under an emissions tax, which leaves emissions higher than their first-best level for any level of ϕ.

[21] The $z \geq 2$ assumption avoids special cases where $\phi_1 < 0$. The exact expressions of the critical values are shown in the Appendix

Proof. See Appendix. □

Proposition 3 determines the optimal choice of base at the first stage and therefore provides the overall equilibrium of the game.[22] Its main message is that the level of SR influences the optimal choice of tax-base: for low and high levels of SR the emissions tax is more efficient, but there is an intermediate range of SR at which the output tax provides higher welfare.

By using the appropriate tax base and rate, the regulator aims to correct two distortions, a too high level of emissions and a too low level of output, due to market power. In the absence of SR, the emissions tax is more efficient in reducing emissions. This advantage is gradually lost as ϕ increases, since net emissions are continuously decreasing under an output tax and continuously increasing under an emissions tax. At the same time, when the tax base is output, increases in ϕ result in higher output, which helps the regulator deal with the monopoly distortion, while under an emissions tax output is decreasing in ϕ. Hence, the output tax takes full advantage from increases in SR, whereas such increases reduce the welfare-enhancing power of the emissions tax.

However, for SR levels above ϕ^* the the prospect of more favorable taxation induces the monopolist to over-abate, in the sense of engaging in socially wasteful abatement in order to obtain even higher tax-rate reductions (or even subsidies, when allowed). Since the excessive abatement decision cannot be reversed at the moment the tax rate is decided, the regulator has no other choice but to focus on correcting for market-power. To illustrate this point, let us allow for a moment output subsidies, which implies levels of SR beyond ϕ_O, and consider that ϕ is so high as to induce zero net emissions. Now, the only choice of the regulator is to target the optimal quantity, *i.e.,* the one at which marginal benefit equals marginal production cost (here, zero). Such a solution would have been optimal in the absence of environmental externality, but not in our context, where the true marginal cost of the good is positive and equal to environmental damage. As SR increases this distortion becomes increasingly serious, leading to welfare levels well below those that could be obtained using an emissions tax.

Since high levels of SR require negative tax rates to be applied on either base, both tax types cease to be valid policy instruments if no subsidies are allowed. If one sees time-consistent taxation as a kind of VA, this implies that no such agreement can be fully efficient in jurisdictions with high SR, unless it provides stimulus in terms of subsidies. Proposition 3 shows that a VA o tax rates without use of subsidies is viable for a wider range of SR if it applies on emissions, since $\phi_E > \phi_O$.

While we do not have a formal proof, extensive numerical simulations corroborate a result illustrated on figure 5, namely that, when the regulator is not aware of the exact level of SR, choosing mistakenly emissions instead of output as base produces less damage than the opposite; in other words, offering a menu of emissions tax rates is a safer choice for a regulator who possesses imperfect information.

[22]Provided that the regulator can commit to a tax base at the first stage, an assumption that has been justified even if tax-rate commitment is impossible.

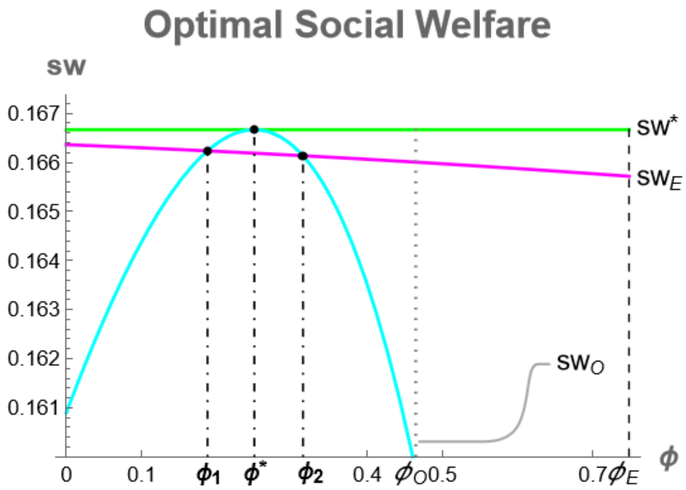

Figure 5: EQUILIBRIUM SOCIAL WELFARE

4 Change in the Order of Moves: Comparisons

In this section we compare the equilibrium with no commitment (time-consistent policy) to the equilibrium in case the regulator is able to commit to a tax-rate from the outset. The latter has been extensively studied in Constantatos et al. (2021) (hereafter CPS), so the analysis in this section consists of comparing the results from the previous section to those contained in that paper. We refer to the commitment model as model C and its equilibrium variables are indicated by a hat (ˆ); whenever it is needed for the comparisons, we refer to time-consistent model analyzed in the previous sections as model N (no commitment). As earlier, the discussion is limited to interior solutions in the optimal choice of the tax rate, nevertheless, in order to enhance comparability *we allow for output, but not emissions-, subsidies.*[23]

Allowing for regulator's commitment eliminates the monopolist's motivation to affect the tax rate and therefore reduces its incentive for abatement (see Petrakis & Xepapadeas (2001)). Thus, it is easy to show that while for each tax base the evolution of abatement as function of ϕ exhibits similar pattern in models N and C, $v_i > \hat{v}_i$, $i = O, E$, *i.e.,* in the C model the regulator's commitment lowers abatement independently of the tax base in use. The situation is depicted in Figure 6 where three new colors are added to our basic color code and remain in use for all the following figures: a blue (red) line indicates the equilibrium value of a variable under output (emissions) tax in the C model and is contrasted to the cyan (magenta) line that still indicates the corresponding value in the N model. Note finally that the cyan v_O line now extends beyond ϕ_O since we allow for output subsidies; we do not show the entire line to avoid affecting the figures scale.[24] Due to the absence of the tax-rate affecting

[23]The argument in Constantatos et al. (2021) is that while output subsidies can be eventually admissible by politicians and the public–especially if one considers time-consistent taxation as a form of VA–emissions subsidies are not acceptable. Note, however, that allowing for any type of subsidy would not affect our previous analysis; we ruled them out up to this point in order to avoid corner solutions that might appear in the N, but not in the C model.

[24]All the figures in CPS have been drawn for $k = 1$, $n = 1$, $\delta = 1$, $\alpha = 1$ and $z = 3/2$. In order to obtain compatibility with our previous figures, we have reproduced the CPS figures assuming $k = 2$ and $z = 2$.

incentive in model C, the blue line lies everywhere below the cyan line, and the same holds true for the red compared to the magenta line.

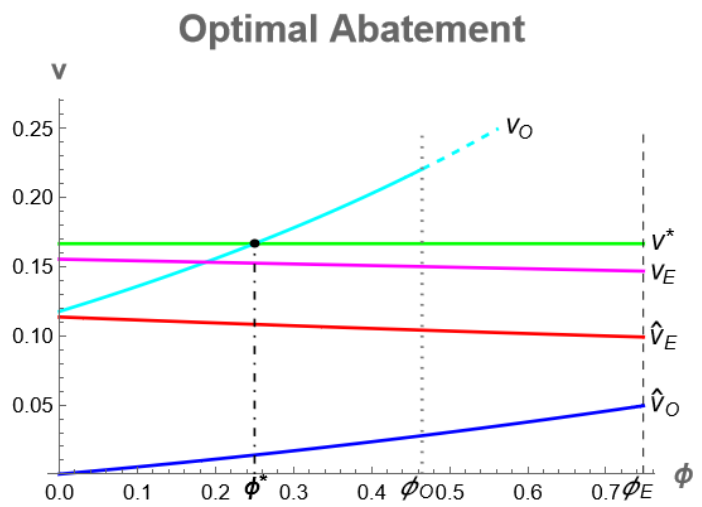

Figure 6: Equilibrium Abatement Comparison

Since abatement under an emissions tax in the N model is always lower than first-best abatement, the incentive reduction caused by a mandatory regulation exacerbates under-abatement. Hence, for any level of social consciousness, the emissions tax yields abatement closer to its first-best level in the absence of regulator's commitment.[25] A similar conclusion does not necessarily hold true for the output tax. As shown in Proposition 1, for some levels of SR, the time-consistent regulator may cause over-abatement, and in such cases the incentive reduction due to regulator's commitment may bring abatement closer to its socially optimal level. While commitment causes equilibrium abatement to increase monotonically with ϕ–exactly as in the N model–it is shown in CPS that abatement never exceeds its first-best level, which is reached only when consumers fully endogenize the externality their consumption creates. Under regulator's inability to commit, on the other hand, abatement reaches its first-best level at relatively low levels of SR, thus leading to over-abatement in a wide range of ϕ values. For instance, at $\phi = .4$ on Figure 6, while consumers internalize only 20% of the externality they create,[26] there is already substantial over-abatement.

Turning to net emissions, we see that the emissions tax in the C model (red line) produces systematically lower than first-best emissions, not because of over-abatement, but because of a too high reduction in output (see CPS). As consumers become more conscious, the quantity reduction in the N model is less pronounced compared to that in the C model, because of the firm's additional abatement incentive, leading to monotonically increasing net emissions that are always above their first-best level. Thus, the impact of increases in ϕ on net emissions is qualitatively very different in the two models when an emissions tax is in use.

Since for any tax base the N-model equilibrium involves higher abatement, it must also involve higher quantity of the polluting good, therefore the comparison of the per-

[25] This is also shown in Petrakis & Xepapadeas (2001), but only for $\phi = 0$.

[26] Recall that the maximum value of $\phi = z = 2$ on this figure.

formance of each tax base with and without commitment cannot be predicted from the outset. Simple inspection of Figure 7 shows that for any tax base, the evolution of net emissions as function of social consciousness is markedly different according to whether the regulator is able or not to commit to a tax rate.

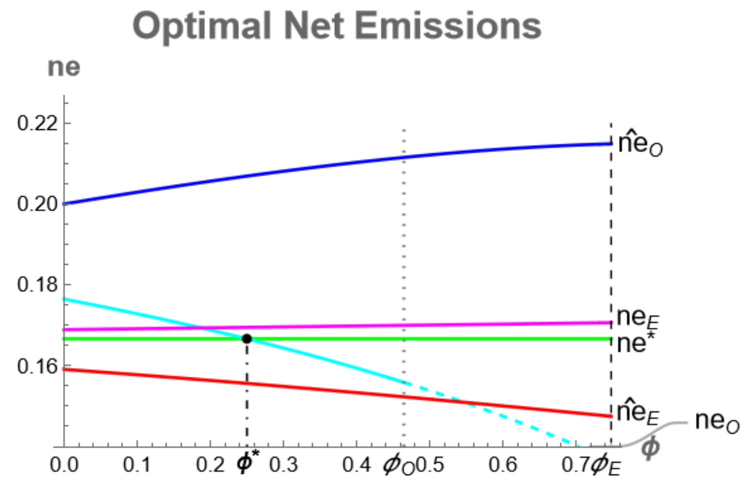

Figure 7: Equilibrium Net Emissions Comparison

The emissions tax produces cleaner environment when the regulator commits to a tax rate. While the time-consistent regulator induces higher abatement, he also induces higher output, yielding overall more pollution. As shown in CPS the *emissions tax* under regulator's commitment produces such a sharp quantity reduction that total net emissions fall below their first-best level for any level of SR (see the red line on figure 7); in the N model net emissions are always higher than their first-best level.

Proposition 2 in this paper states that under an *output tax*, net emissions in the N model decrease rapidly as social consciousness increases, and while at $\phi = 0$ they start from a level above that of first-best, they soon go below it. No such response is observed in the C model, where under an output tax, net emissions are always above their first-best level, initially *increasing* (due to lower tax-rate) but eventually decreasing with social consciousness, reaching their first-best level only when consumers become fully responsible.[27] The already identified incentive to affect the tax rate of an uncommitted regulator and the resulting tendency of the output tax to quickly produce over-abatement in the N model are responsible for the drastic effect that increases in SR have in reducing pollution. Absent this incentive, there is never over-abatement and net emissions never fall below their first-best level. Turning to equilibrium social welfare, we can prove the following proposition.

Proposition 4. *When the tax is on output there is a critical level of SR, call it* ϕ_{sw}*, such that* $\forall \phi \in [0, \phi_{sw}]$*,* $sw_O > \hat{sw}_O$*, while* $\forall \phi \in [\phi_{sw}, \phi_E]$*,* $sw_O < \hat{sw}_O$*. When the tax is on emissions,* $\forall \phi \in [0, \phi_E]$*,* $sw_E > \hat{sw}_E$*, i.e., welfare is always higher in the absence of regulator's commitment.*

Proof. See Appendix. □

The content of proposition 4 is illustrated on figure 8 below. It is shown that the social

[27] See Constantatos et al. (2021). The decreasing part of the blue curve is not shown on figure 7 since it occurs for $\phi > \phi_E$

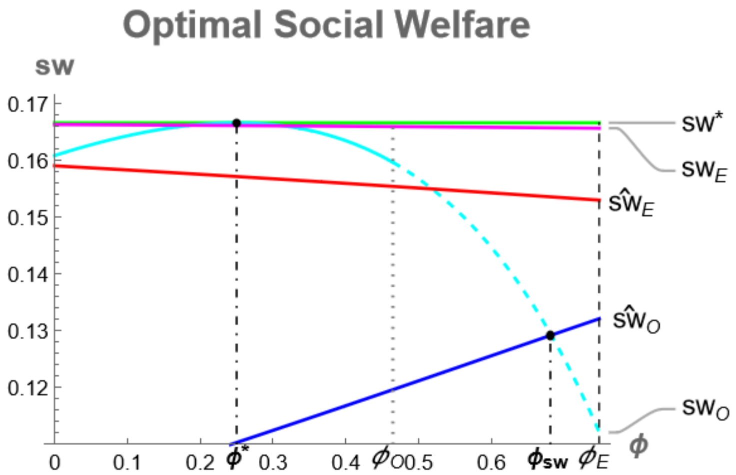

Figure 8: Equilibrium Social Welfare Comparison

welfare resulting from a VA (magenta line) is always higher than that of the mandatory regulation (red line), meaning that when it comes to choosing between offering a VA or using mandatory taxation (*i.e.,* whether committing or not to an emissions tax), by choosing the latter, society enjoys more welfare. The intuition behind this result is straightforward. The time consistent social planner provides a motivation to the firms in order to invest more in abating technology. Since there is no exaggeration in abatement investment (see Figure 5) the resulting social welfare is greater than the one generated by a committed regulator. When it comes to the output tax, there is no clear verdict in favor of a time-consistent schedule, unless negative tax rates are excluded. For sufficiently high levels of SR (above ϕ_{sw}), committing to a tax rate, if feasible, may be welfare superior. The problem with VA allowing the choice of output-tax rate is that they make the firm too eager to provide abatement in exchange of tax-rate reductions. For sufficiently high levels of SR, such that the optimal policy calls for subsidies, and if the latter are allowed, the tendency for excessive abatement tendency is even stronger. While this creates a "happy outcome" with high output and very low emissions in the market in question, this outcome is obtained at the expense of withdrawing too many resources from the rest of the economy; the resulting welfare is thus decreasing in ϕ. Such tendency for over-abatement is not created under mandatory regulation with *a priori* commitment to a given tax rate. As a result, mandatory taxation is in tune with increases in SR, yielding higher welfare as the latter increases.

5 Conclusions

In this work we focus on several issues, assuming that a) consumers are socially responsible (as defined in the paper), and b) the regulator can adjust the tax rate after observing abatement. Following the work in CPS, we base our analysis on a social responsibility concept that relates environmental consciousness to various degrees of internalization of the environmental externality. This concept is different than the usual *ad hoc* assumption of preference for green products.

First, we find that increases in SR lower the abatement incentive under an emissions tax, while the opposite holds true for an output tax: as consumers become more responsible, the firm has an always higher abatement incentive.

Second, social welfare under an emissions tax is decreasing everywhere as SR increases. The behavior of social welfare as function of SR is more complex under an output tax: social welfare increases with SR up to a point, but curbs down sharply, afterwards. When at its peak, welfare under an output tax reaches its first-best level, whereas an emissions tax is under no level of SR able to reach first-best welfare. The reason behind the initially harmonious working between output tax and SR is the same which is responsible for the cacophony between these two variables at higher levels of SR: the strong increases in abatement incentive that the output tax produces as SR increases. While for low levels of SR the increases in abatement incentive are beneficial, beyond the point where abatement reaches its first-best level such increases lead to over-abatement, *i.e.,* levels where the marginal abatement cost exceeds the corresponding marginal benefit.

Third, we show that while for low and high levels of SR a time-consistent *emissions tax* is welfare superior to a time-consistent output tax, there is a middle range of SR parameter values for which the *output tax* performs better. This result is related to the evolution of the abatement incentive following increases in SR, discussed in the previous paragraph.

Forth, we compare the welfare performance of each tax-base between the cases where the regulator is, or is not committed to an initial tax rate announcement. For the emissions tax, it is clear that for any SR level, non-commitment is superior, which corresponds to an extension of the similar result in Petrakis & Xepapadeas (2001) and J. Poyago-Theotoky & Teerasuwannajak (2002), where SR is zero. However, with an output tax, commitment may be superior for high levels of SR. At those levels, the potential of SR to fuel over-abatement makes a time-consistent tax schedule inferior to mandatory imposition of a given tax rate. While increases in SR lead under an output tax to higher abatement levels, the tax-manipulation incentive present when the regulator is unable to commit may easily over-shoot abatement to levels that are uneconomically high.

Time consistent taxation has the characteristics of a "voluntary agreement" in that it allows both parties to agree upon mutually beneficial terms, *i.e.,* greater abatement in exchange for lower taxation, or even subsidization. In that sense, all the above results may highlight benefits and problems potentially related to VA.

Various extensions of this work are possible. First, we limited the analysis to levels of SR that lead to interior solutions. When no such subsidies are allowed there may be levels of SR for which corner solutions occur, with properties other than the aforementioned results; we are currently working on this issue. Also, the analysis must be extended to cases where the abatement investment reduces pollution per unit of output, a topic that features in our research agenda.

References

Aoyama, N., & Delfino Silva, E. (2016, September). Abatement innovation in a cournot oligopoly: Emission versus output tax incentives. In *Cesifo working paper series no. 6094, category 10: Energy and climate economics.*

Bansal, S. (2008). Choice and design of regulatory instruments in the presence of green consumers. *Resource and Energy economics*, *30*(3), 345–368.

Bansal, S., & Gangopadhyay, S. (2003). Tax/subsidy policies in the presence of environmentally aware consumers. *Journal of Environmental Economics and Management*, *45*(2), 333–355.

Brouhle, K., & Khanna, M. (2007). Information and the provision of quality differentiated products. *Economic Inquiry*, *45*(2), 377–394.

Constantatos, C., Pargianas, C., & Sartzetakis, E. S. (2021). Green consumers and environmental policy. *Journal of Public Economic Theory*, *23*(1), 105–140.

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

Delmas, M. A., & Montes-Sancho, M. J. (2010). Voluntary agreements to improve environmental quality: Symbolic and substantive cooperation. *Strategic Management Journal*, *31*(6), 575–601.

Deltas, G., Harrington, D. R., & Khanna, M. (2013). Oligopolies with (somewhat) environmentally conscious consumers: market equilibrium and regulatory intervention. *Journal of Economics & Management Strategy*, *22*(3), 640–667.

Doni, N., & Ricchiuti, G. (2013). Market equilibrium in the presence of green consumers and responsible firms: A comparative statics analysis. *Resource and Energy Economics*, *35*(3), 380–395.

Frondel, M., Horbach, J., & Rennings, K. (2007). End-of-pipe or cleaner production? an empirical comparison of environmental innovation decisions across oecd countries. *Business strategy and the environment*, *16*(8), 571–584.

Fullerton, D., Hong, I., & Metcalf, G. E. (2001). A tax on output of the polluting industry is not a tax on pollution: The importance of hitting the target. In *Behavioral and distributional effects of environmental policy* (pp. 13–44). University of Chicago Press.

García-Gallego, A., & Georgantzís, N. (2009). Market effects of changes in consumers' social responsibility. *Journal of Economics & Management Strategy*, *18*(1), 235–262.

Gil-Moltó, M. J., & Varvarigos, D. (2013). Emission taxes and the adoption of cleaner technologies: The case of environmentally conscious consumers. *Resource and Energy Economics*, *35*(4), 486–504.

Lambertini, L. (2017). Green innovation and market power. *Annual Review of Resource Economics*, *9*, 231–252.

Nyborg, K., Howarth, R. B., & Brekke, K. A. (2006). Green consumers and public policy: On socially contingent moral motivation. *Resource and energy economics*, *28*(4), 351–366.

Petrakis, E., Sartzetakis, E. S., & Xepapadeas, A. (2005). Environmental information provision as a public policy instrument. *Contributions in Economic Analysis & Policy*, *4*(1), 1–31.

Petrakis, E., & Xepapadeas, A. (2001). To commit or not to commit: Environmental policy in imperfectly competitive markets. *Typescript, Department of Economics, University of Crete*.

Poyago-Theotoky, J., & Teerasuwannajak, K. (2002). The timing of environmental policy: A note on the role of product differentiation. *Journal of Regulatory Economics*, *21*(3), 305–316.

Poyago-Theotoky, J. A. (2007). The organization of r&d and environmental policy. *Journal of Economic Behavior & Organization*, *62*(1), 63–75.

Requate, T. (2005). Timing and commitment of environmental policy, adoption of new technology, and repercussions on r&d. *Environmental and resource Economics*, *31*, 175–199.

Sartzetakis, E. S., Xepapadeas, A., & Petrakis, E. (2012). The role of information provision as a policy instrument to supplement environmental taxes. *Environmental and Resource Economics*, *52*, 347–368.

Schmutzler, A., & Goulder, L. H. (1997). The choice between emission taxes and output taxes under imperfect monitoring. *Journal of Environmental Economics and Management*, *32*(1), 51–64.

Segerson, K., & Miceli, T. J. (1998). Voluntary environmental agreements: good or bad news for environmental protection? *Journal of environmental economics and management*, *36*(2), 109–130.

A Appendix

A.1 Proofs for Propositions and Lemmata

Lemma A.2. *A necessary and sufficient condition for interior solution in the case of output and emissions tax is $\phi_O \geq 0$ and $\phi_E \geq 0$, respectively, with*

$$\phi_O = \frac{(2z^2 + 3z + 2) - (1 + 2z)\sqrt{4 + z^2}}{2z}$$

and

$$\phi_E = \frac{7}{4} + z(6 + z) - \frac{1}{4}(1 + 2z)\sqrt{81 + 4z(11 + z)}$$

with $\phi_O < \phi_E$.

Proof. The expressions ϕ_O and ϕ_E arise when we set the output and emissions tax in equilibrium equal to zero (the exact expressions can be found in the A.3 section of the Appendix). First, in order to find ϕ_O, we set $t_O = 0$:

$$-z\phi^2 + (2z^2 + 3z + 2)\phi - 2z^2 + 1 = 0$$

with the discriminant being: $\Delta = (1 + 2z)^2(4 + z^2)$. Therefore:

$$\phi_{1,2} = \frac{(2z^2 + 3z + 2) \pm (1 + 2z)\sqrt{4 + z^2}}{2z}$$

In order to avoid over-internalization of the externality $\phi \leq z$. Therefore, here we will check which of these values satisfies this condition.

$$\frac{(2z^2 + 3z + 2) + (1 + 2z)\sqrt{4 + z^2}}{2z} \leq z \Leftrightarrow$$

$$(2z^2 + 3z + 2) + (1 + 2z)\sqrt{4 + z^2} \leq 2z^2 \Leftrightarrow$$
$$3z^2 + 2 \leq -(1 + 2z)\sqrt{4 + z^2}$$

which is not true. However when instead of the plus sign we use a negative one (i.e., ϕ_2), we have:

$$3z^2 + 2 < (1 + 2z)\sqrt{4 + z^2} \Leftrightarrow$$
$$9z^2 + 12z + 4 \leq (1 + 4z + 4z^2)(4 + z^{2)} \Leftrightarrow$$
$$16z^4 + 16z^3 + 8z^2 + 4z \geq 0$$

which is true, therefore $\phi_O = \dfrac{(2z^2 + 3z + 2) - (1 + 2z)\sqrt{4 + z^2}}{2z}$.
In the case of the emissions tax:

$$t_E = \frac{-4z(z + 6)\phi + 4z(4z - 1) + 2\phi^2 - 7\phi - 4}{24z(z + 1) - 2\phi + 4},$$

The denominator is always positive. It is a decreasing function in ϕ and therefore for

$$\phi = 0 \rightarrow 24z(z + 1) + 4 > 0$$

and for

$$\phi = z \rightarrow 24z(z + 1) - 2z + 4 = 24z^2 + 22z + 4 > 0$$

Focusing on the numerator:

$$2\phi^2 - [7 + 4z(z + 6)]\phi + 4z(4z - 1) - 4 = 0$$

The discriminant of the polynomial above is:

$$\Delta = 81 + 368z + 504z^2 + 192z^3 + 16z^4$$

with the solutions of the polynomial be:

$$\phi_{1,2} = \frac{7 + 4z(z+6) \pm \sqrt{81 + 368z + 504z^2 + 192z^3 + 16z^4}}{4}$$

which are both positive since the coefficient of ϕ^2 and the constant term are positive and by Vieta's formula we know that $P = \phi_1 \cdot \phi_2 = \dfrac{c}{\alpha} > 0$, where c is the constant term and α coefficient of ϕ^2. We will now check under which circumstances these values are $< z$. Suppose,

$$\frac{7 + 4z(z+6) - \sqrt{81 + 368z + 504z^2 + 192z^3 + 16z^4}}{4} \leq z \Leftrightarrow$$

$$7 + 4z(z+6) - \sqrt{81 + 368z + 504z^2 + 192z^3 + 16z^4} \leq 4z \Leftrightarrow$$
$$7 + 4z^2 + 20z \leq \sqrt{81 + 368z + 504z^2 + 192z^3 + 16z^4} \Leftrightarrow$$
$$(7 + 4z^2 + 20z)^2 \leq 81 + 368z + 504z^2 + 192z^3 + 16z^4 \Leftrightarrow$$
$$49 + 280z + 456z^2 + 160z^3 + 16z^2 \leq 81 + 368z + 504z^2 + 192z^3 + 16z^4 \Leftrightarrow$$
$$32z^3 + 48z^2 + 88z + 32 \geq 0$$

which is true without the equality. Therefore, $\phi_1 < z$. We now check for ϕ_2.

$$\frac{7 + 4z(z+6) + \sqrt{81 + 368z + 504z^2 + 192z^3 + 16z^4}}{4} \leq z \Leftrightarrow$$

$$7 + 4z(z+6) + \sqrt{81 + 368z + 504z^2 + 192z^3 + 16z^4} \leq 4z \Leftrightarrow$$
$$\sqrt{81 + 368z + 504z^2 + 192z^3 + 16z^4} \leq -7 - 4z^2 - 20z$$

which cannot be true. Therefore, $\phi_2 > z$ meaning that $t_E \geq 0 \ \forall\ 0 \leq \phi \leq \phi_1,\ z > 1$ and thus we set

$$\phi_E = \frac{7 + 4z(z+6) - \sqrt{81 + 368z + 504z^2 + 192z^3 + 16z^4}}{4}$$
$$= \frac{7}{4} + z(6+z) - \frac{1}{4}(1+2z)\sqrt{81 + 4z(11+z)}$$

Due to the complexity of the calculations, we use *Mathematica* software to prove that $\phi_O < \phi_E$ □

Proposition A.1. *(i)If the announced tax is on output* $\forall \phi \leq \phi_O$ *a) abatement is monotonically increasing in* ϕ*; b) there exists a value of* $\phi = \phi^* \in [0, \phi_O]$

$$\phi^* \triangleq \frac{z-1}{z+2}$$

such that the market produces the first-best level of abatement ($v_O = v^*$)*, while for* $\phi > (<)\phi^*$ *market abatement is above (below)* v^**.*
(ii) If the announced tax is on emissions, for all $\phi \in [0, \phi_E]$ *abatement is monotonically* ***decreasing*** *in* ϕ *and* $v_E < v^*$*.*

Proof. The derivative of

$$v_O = \frac{z(1+\phi)}{2z(-z\phi+z+2)+1}$$

is

$$\frac{\partial v_O}{\partial \phi} = \frac{z[2z(-z\phi+z+2)+1]-(-2z^2)(z+z\phi)}{[2z(-z\phi+z+2)+1]^2}$$

Since the denominator is always positive, we focus on the numerator and expand it to obtain:

$$-2z^3\phi + 2z^3 + 4z^2 + z - (z+z\phi)(-2z^2) =$$
$$-2z^3\phi + 2z^3 + 4z^2 + 2z^2(z+z\phi) =$$
$$4z^3 + 4z^2 + z$$

which is positive $\forall z > 1$, meaning that v_O is monotonically increasing. In what follows we determine the value of ϕ at which v_O is equal to v^*.

$$v_O = v^* \Leftrightarrow$$
$$\frac{z(1+\phi)}{[2z(-z\phi+z+2)+1]} = \frac{z}{2+5z} \Leftrightarrow$$
$$z(1+\phi)(2+5z) = z[2z(-z\phi+z+2)+1] \Leftrightarrow$$
$$z(1+\phi)(2+5z) - 2z^2(-z\phi+z+2) - z = 0$$

Expanding this expression and simplifying it we obtain:

$$z + z^2 - 2z^3 + 2z\phi + 5z^2\phi + 2z^3\phi = 0 \Leftrightarrow$$
$$z + z^2 - 2z^3 + (2z + 5z^2 + 2z^3)\phi = 0 \Leftrightarrow$$
$$(2z + 5z^2 + 2z^3)\phi = 2z^3 - z^2 - z \Leftrightarrow$$
$$\phi = \frac{2z^3 - z^2 - z}{2z + 5z^2 + 2z^3} \Leftrightarrow$$
$$\phi = \frac{2z^2 - z - 1}{2z^2 + 5z + 2} \Leftrightarrow$$
$$\phi = \frac{(z-1)(z+\frac{1}{2})}{(z+2)(z+\frac{1}{2})}$$
$$\therefore \phi^* = \frac{z-1}{z+2}$$

In what follows we show that $\phi^* < \phi_O$. Suppose, that it is not true, *i.e.*, $\phi^* \geq \phi_O$. Then

$$\frac{z-1}{z+2} \geq \frac{2z^2+3z+2-(1+2z)\sqrt{4+z^2}}{2z} \Leftrightarrow$$
$$2z(z-1) \geq (2z^2+3z+2)(z+2) - (1+2z)(z+2)\sqrt{4+z^2} \Leftrightarrow$$
$$2z^3 + 5z^2 + 10z + 4 \leq (z+2)(1+2z)\sqrt{4+z^2}$$

We square both parts of the equation and simplify to obtain: $-16z^4 - 16z^3 - 4z^2 \geq 0$

which cannot true. By contradiction, it holds that $\phi^* < \phi_O$.
Since the first-best case is independent of ϕ and v_O increases in ϕ, for $\phi < \phi^*$, $v_O < v^*$ and for $\phi > \phi^*$, $v_O > v^*$.
The derivative of

$$v_E = \frac{4z(z+1) - 1 - 2\phi}{24z(z+1) + 4 - 2\phi}$$

is

$$\begin{aligned}
\frac{\partial v_E}{\partial \phi} &= \frac{-2\left[24z(z+1)+4-2\phi\right] + 2\left[4z(z+1)-1-2\phi\right]}{\left[24z(z+1)+4-2\phi\right]^2} \\
&= \frac{-2[24z(z+1)+4-2\phi-4z(z+1)+1+2\phi]}{\left[24z(z+1)+4-2\phi\right]^2} \\
&= \frac{20z(z+1)+5}{2[12z(z+1)+2-\phi]^2} \\
&= -\frac{5[4z^2+4z+1]}{2[12z(z+1)+2-\phi]^2} \\
&= -\frac{5(2z-1)^2}{2[12z(z+1)+2-\phi]^2} < 0 \; \forall \; \phi \in [0, z], \; z > 1
\end{aligned}$$

Therefore, v_E decreases in ϕ.

□

Proposition A.2. *(i) When the tax is on output, $ne'_O < 0$, with $ne_O(0) > ne^*$ and $ne_O(\phi_O) < ne^*$, i.e., net emissions decrease monotonically with ϕ, starting from a level above that of first-best when SR is zero and ending at a level below it, as ϕ approaches its highest allowed value; at the SR level $\phi = \phi^*$ (see (14)) net emissions are equal to their first-best level.*
(ii) When the tax is on emissions, $ne'_E > 0$, with $ne_E(0) > ne^*$, i.e., net emissions increase monotonically with ϕ and are always above their first-best level.
(iii) For relatively low (high) levels of SR, $ne_O > (<) ne_E$

Proof.

$$ne_O = \frac{1+z-z\phi}{1+2z(2+z-z\phi)}$$

$$\therefore ne_O' = \frac{\partial ne_O}{\partial \phi} = \frac{-z[1+2z(2+z-z\phi)] - (-2z^2)(1+z-z\phi)}{[1+2z(2+z-z\phi)]^2}$$

Since the denominator is always positive, we focus on the numerator. Let:

$$N_1 = -z[1+2z(2+z-z\phi)]$$

and

$$N_2 = 2z^2(1+z-z\phi)$$

$$N_1 = -z - 4z^2 - 2z^3 + 2z^3\phi$$

and

$$N_2 = 2z^2 + 2z^3 - 2z^3\phi$$

$$\therefore N_1 + N_2 = -z - 2z^2$$

which is negative $\forall\ z > 1$. Therefore, ne_O is a downward sloping function of ϕ. We will now check whether

$$ne_O(0) = \frac{1+z}{1+2z(2+z)}$$

is greater than the first-best case. Suppose

$$ne_O(0) > ne^* \Leftrightarrow$$

$$\frac{1+z}{1+2z(2+z)} > \frac{2}{2+5z} \Leftrightarrow$$

$$(2+5z)(1+z) > 2[1+2z(2+z)] \Leftrightarrow$$

$$2+7z+5z^2 > 2+8z+4z^2 \Leftrightarrow$$

$$7+5z^2 > 8z+4z^2 \Leftrightarrow$$

$$-z+z^2 > 0 \Leftrightarrow$$

$$z(z-1) > 0$$

leading to either $z > 1$ or $z < 0$ in order for the expression to be positive. By construction of the model $z > 1$. Therefore,

$$ne_O(0) > ne^*.$$

Moreover, we know that $ne_O(\phi^*) = ne^*$ and since $\phi^* < \phi_O$ and $ne_O^{'} < 0$ then $ne_O(\phi_O) < ne^*$. Turning to the emissions tax, we have:

$$ne_E = \frac{10z+5}{24z(z+1)-2\phi+4}$$

$$\therefore ne_E^{'} \equiv \frac{\partial ne_E}{\partial \phi} = \frac{-(-2)(10z+5)}{[24z(z+1)-2\phi+4]^2}$$

$$= \frac{5(2z+1)}{2[12z(z+1)-\phi+2]^2}$$

which is positive $\forall\ z > 1$.Suppose,

$$ne_E(0) > ne^* \Leftrightarrow$$

$$\frac{10z+5}{24z(z+1)+4} > \frac{2}{2+5z} \Leftrightarrow$$

$$(10z+5)(2+5z) > 2[24z(z+1)+4] \Leftrightarrow$$

$$10+45z+50z^2 > 8+48z+48z^2 \Leftrightarrow$$

$$2-3z+2z^2 > 0$$

which is always positive, because the discriminant is negative and the coefficient of z^2 is positive. Now suppose,

$$ne_O(0) > ne_E(0) \Leftrightarrow$$

$$\frac{10z+5}{24z(z+1)+4} < \frac{1+z}{1+2z(2+z)} \Leftrightarrow$$

$$(10z+5)[1+2z(2+z)] < (1+z)[24z(z+1)+4] \Leftrightarrow$$
$$1+2z+2z^2-4z3 < 0$$

Let $f(z) = 1+2z+2z^2-4z^3$ with its derivative being $f^{'}(z) = 2+4z-12z^2 = 2(1+2z-6z^2)$. This equation has two solutions, a positive and a negative. The positive one is $\frac{1+\sqrt{7}}{6} \approx 0.607625$ and since $z > 1$, $f(z)$ is downward sloping. At $z = 1 \rightarrow f(1) = 1$ and at $z = 1.15 \rightarrow f(1.15) = -0.1325$. Therefore, by Bolzano's Theorem, there is a $\tau \in (1, 1.15)$ such that $f(\tau) = 0$. For $1 < z < \tau$, $ne_O(0) < ne_E(0)$ and for $z > \tau$, $ne_O(0) > ne_E(0)$. We keep the latter condition since the former generates uninteresting cases. □

Lemma A.3. *(i) Under an output tax,* $\forall\phi \leq (\geq)\phi^*$, $sw_O' \geq (\leq)0$, *with* $sw_O(\phi^*) = sw^*$.
(ii) Under an emissions tax, the welfare function is monotonically decreasing in ϕ*; an emissions tax never achieves first-best welfare, i.e.,* $\forall\phi$, $sw_E < sw^*$

Proof. We expand the numerator of

$$sw_O = \frac{2z\{z[-z(\phi-1)(\phi+3)-2\phi(\phi+2)+4]+3\}+1}{2(2z(-z\phi+z+2)+1)^2}$$

in order to simplify the derivative. Therefore, the expressions above becomes:

$$sw_O = \frac{1+6z+8z^2+6z^3-(8z^2+4z^3)\phi-(4z^2+2z^3)\phi^2}{2[2z(-z\phi+z+2)+1]^2}$$

$$sw_O' = \frac{\partial sw_O}{\partial\phi} = \frac{2[-8z^2-4z^3-(4z^2+2z^3)\phi]-(-8z^2)[1+2z(2+z-z\phi)](\cdot)}{4[2z(-z\phi+z+2)+1]^4}$$

where in $(\cdot)$ we have $1+6z+8z^2+6z^3-(8z^2+4z^3)\phi-(4z^2+2z^3)\phi^2$. By performing some simplifications the derivative becomes:

$$\frac{\partial sw_O}{\partial\phi} = \frac{[-4z^2-2z^3-(4z^2+2z^3)\phi][2z(-z\phi+z+2)+1]+2z^2(\cdot)}{[2z(-z\phi+z+2)+1]^3}$$

The denominator is positive for $\phi < \frac{2z^2+4z+1}{2z^2}$ and since from the proof of Proposition 1 $\phi < \phi_E < \frac{2z^2+4z+1}{2z^2}$, we will focus on the numerator in order to identify the sign of the derivative. Let

$$N_1 = [-4z^2-2z^3-(4z^2+2z^3)\phi][2z(-z\phi+z+2)+1]$$

and

$$N_2 = 2z^2[1+6z+8z^2+6z^3-(8z^2+4z^3)\phi-(4z^2+2z^3)\phi^2]$$

By performing some manipulations and add N_1, N_2, we have:

$$N_1+N_2 = -8\phi z^5+8z^5-24\phi z^4-18\phi z^3-6z^3-4\phi z^2-2z^2$$
$$= -2z^2(2z+1)^2[2\phi+(\phi-1)z+1]$$

$$\therefore \frac{\partial sw_O}{\partial \phi} = \frac{-2z^2(2z+1)^2[2\phi + (\phi-1)z+1]}{[2z(-z\phi+z+2)+1]^3}$$

Under which circumstances is this positive/negative?

$$1 + z(-1+\phi) + 2\phi < 0 \Leftrightarrow$$

$$1 - z + z\phi + 2\phi < 0 \Leftrightarrow$$

$$(z+2)\phi < z-1 \Leftrightarrow$$

$$\phi < \frac{z-1}{z+2} (= \phi^*)$$

meaning the derivative is positive. So for $\phi < (>) \phi^*$, sw_O increases (decreases) in ϕ. Therefore, sw_O picks at $\phi^* = \frac{z-1}{z+2}$ and reaches the first best case since both v_O, q_O (see later) reach the first-best at ϕ^*.
The derivative of

$$sw_E = \frac{56z^4 + 212z^3 - 16\left(z^2+z+1\right)\phi + 218z^2 + 87z - 6\phi^2 + 6}{[-24z(z+1)+2\phi-4]^2}$$

is

$$sw_E' = \frac{\partial sw_E}{\partial \phi} = \frac{[-16(z^2+z+1)][-24z(z+1)+2\phi-4]^2 - 4[-24z(z+1)+2\phi-4](\cdot)}{[-24z(z+1)+2\phi-4]^4},$$

where in $(\cdot)$ we have $56z^4 + 212z^3 - 16\left(z^2+z+1\right)\phi + 218z^2 + 87z - 6\phi^2 + 6$. Let

$$N_1 = [-16(z^2+z+1)][-24z(z+1)+2\phi-4]^2$$

and

$$N_2 = -4[-24z(z+1)+2\phi-4][56z^4+212z^3-16\left(z^2+z+1\right)\phi+218z^2+87z-6\phi^2+6].$$

By adding N_1 and N_2, we get:

$$\begin{aligned} N_1 + N_2 = [-24z(z+1)+2\phi-4]\{&[-16(z^2+z+1)][-24z(z+1)+2\phi-4]- \\ &4[56z^4+212z^3-16\left(z^2+z+1\right)\phi+218z^2+87z-6\phi^2+6]\} \end{aligned}$$

$$\begin{aligned} &= [-24z(z+1)+2\phi-4][64+448z+832z^2+768z^3+384z^4+16\phi+256z\phi+ \\ &256z^2\phi - 24\phi^2 - 24 - 348z - 872z^2 - 848z^3 - 224z^4 + 64\phi + 64z\phi + 64z^2\phi + 24\phi^2] \end{aligned}$$

$$= [-24z(z+1)+2\phi-4](40+100z-40z^2-80z^3+160z^4+80\phi+320z\phi+320z^2\phi)$$

$$= 2[-12z(z+1)+\phi-2]\{20(1+2z)^2[2+z(-3+2z)+4\phi]\}.$$

Therefore,

$$\frac{\partial sw_E}{\partial \phi} = \frac{2[-12z(z+1)+\phi-2]\{20(1+2z)^2[2+z(-3+2z)+4\phi]}{[-24z(z+1)+2\phi-4]^4}$$

$$= \frac{5(1+2z)^2[2+z(-3+2z)+4\phi]}{2[-12z(z+1)+\phi-2]^3}$$

which is negative because the denominator is negative $\forall\ z > 1$. The numerator is positive because $[2+z(-3+2z)+4\phi$ is an increasing function of ϕ and at $\phi = 0$ the expressions becomes $2-3z+2z^2$, which is positive (the discriminator is negative and therefore the sign of the expressions is dictated by the sign of the coefficient of z^2), meaning that $\dfrac{\partial sw_E}{\partial \phi} < 0$.

Since sw_E is a decreasing function of ϕ, it suffices to show that $sw_E(0) < sw^*$ in order to prove that $sw_E < sw^*$. Therefore,

$$\frac{56z^4+212z^3+218z^2+87z+6}{[-24z(z+1)-4]^2} < \frac{z+2}{4+10z} \Leftrightarrow$$

$$(56z^4+212z^3+218z^2+87z+6)(4+10z) < [-24z(z+1)-4]^2(z+2)$$

By expanding and simplifying the resulting expressions, we get:

$$2(1+2z)[2+z(-3+2z)]^2 > 0$$

which is true. Thus, $sw_E < sw^*\ \forall\phi,\ z > 1$ □

Proposition A.3. *For all $z \geq 2$, there exist two critical values of ϕ, $\phi_1 < \phi_2$, such that $sw_O(\phi_i) = sw_E(\phi_i)$, $i = 1,2$. For values of ϕ within (outside) the (ϕ_1,ϕ_2) interval, $sw_O > (<) sw_E$.*

Proof. Let

$$sw_O - sw_E = \frac{N_1+N_2}{2[2z(-z\phi+z+2)+1]^2[-24z(z+1)+2\phi-4]^2}$$

where

$$N_1 = [1+6z+8z^2+6z^3-(8z^2+4z^3)\phi-(4z^2+2z^3)\phi^2][-24z(z+1)+2\phi-4]^2$$

$$N_2 = -2[56z^4+212z^3-16\left(z^2+z+1\right)\phi+218z^2+87z-6\phi^2+6][2z(-z\phi+z+2)+1]^2$$

By expanding N_1 we get:

$$\begin{aligned}N_1 = {} & -1152z^7\phi^2-2304z^7\phi+3456z^7-4608z^6\phi^2-9216z^6\phi+11520z^6+192z^5\phi^3-5760z^5\phi^2\\ & -12864z^5\phi+17280z^5+576z^4\phi^3-2304z^4\phi^2-8256z^4\phi+14784z^4-8z^3\phi^4+400z^3\phi^3\\ & +56z^3\phi^2-3040z^3\phi+7392z^3-16z^2\phi^4+32z^2\phi^3+96z^2\phi^2-928z^2\phi+2048z^2+24z\phi^2\\ & -192z\phi+288z+4\phi^2-16\phi+16\end{aligned}$$

$$\begin{aligned}N_2 = {} & -448z^8\phi^2+896z^8\phi-448z^8-1696z^7\phi^2+5184z^7\phi-3488z^7+128z^6\phi^3-2000z^6\phi^2\\ & +10848z^6\phi-10768z^6+128z^5\phi^3-1464z^5\phi^2+10704z^5\phi-17048z^5+48z^4\phi^4+32z^4\phi^3\\ & -896z^4\phi^2+5904z^4\phi-15056z^4-192z^3\phi^3-448z^3\phi^2+2296z^3\phi-7584z^3-48z^2\phi^3\\ & +112z^2\phi^2+976z^2\phi-2068z^2+96z\phi^2+288z\phi-270z+12\phi^2+32\phi-12\end{aligned}$$

By adding the expressions above we have

$$\begin{aligned}N_1+N_2 = {} & -448z^8\phi^2+896z^8\phi-448z^8-2848z^7\phi^2+2880z^7\phi-32z^7+128z^6\phi^3-6608z^6\phi^2\\ & +1632z^6\phi+752z^6+320z^5\phi^3-7224z^5\phi^2-2160z^5\phi+232z^5+48z^4\phi^4+608z^4\phi^3\\ & -3200z^4\phi^2-2352z^4\phi-272z^4-8z^3\phi^4+208z^3\phi^3-392z^3\phi^2-744z^3\phi-192z^3\\ & -16z^2\phi^4-16z^2\phi^3+208z^2\phi^2+48z^2\phi-20z^2+120z\phi^2+96z\phi+18z\\ & +16\phi^2+16\phi+4\end{aligned}$$

And we simplify the expression to get:

$$N_1 + N_2 = -2(2z+1)\left\{2z\left[2z^2(\phi-1)+7z\phi+z-(\phi-3)\phi+1\right]+2\phi+1\right\}$$
$$\left\{z\left[28z^3(\phi-1)+z^2(66\phi-2)+z\left(6\phi^2+50\phi+26\right)-2\phi(2\phi+5)-1\right]-4\phi-2\right\}$$

Now, we are going to see the signs of the individual expressions in curly brackets. Let

$$P_1 = 2z\left[2z^2(\phi-1)+7z\phi+z-(\phi-3)\phi+1\right]+2\phi+1$$

and

$$P_2 = z\left[28z^3(\phi-1)+z^2(66\phi-2)+z\left(6\phi^2+50\phi+26\right)-2\phi(2\phi+5)-1\right]-4\phi-2$$

We expand P_1 and we get:

$$P_1 = 1+2z+2z^2-4z^3+(2+6z+14z^2+4z^3)\phi-2z\phi^2$$

which has a discriminant of:

$$\Delta_{P_1} = 4(1+8z+27z^2+50z^3+53z^4+28z^5+4z^6)$$

and the resulting solutions are:

$$\phi_{1,2} = \frac{1+3z+7z^2+2z^3 \pm \sqrt{1+8z+27z^2+50z^3+53z^4+28z^5+4z^6}}{2z}$$

We have to check whether the roots are within the $[0, z]$ interval. Suppose:

$$\phi_2 = \frac{1+3z+7z^2+2z^3 + \sqrt{1+8z+27z^2+50z^3+53z^4+28z^5+4z^6}}{2z} \leq z \Leftrightarrow$$

$$1+3z+7z^2+2z^3+\sqrt{1+8z+27z^2+50z^3+53z^4+28z^5+4z^6} \leq 2z^3 \Leftrightarrow$$
$$1+3z+7z^2+\sqrt{1+8z+27z^2+50z^3+53z^4+28z^5+4z^6} \leq 0$$

which is not true. Therefore, $\phi_2 > z$. We now turn to ϕ_1. Suppose:

$$\phi_1 < z \Leftrightarrow$$

$$\frac{1+3z+7z^2+2z^3 - \sqrt{1+8z+27z^2+50z^3+53z^4+28z^5+4z^6}}{2z} < z \Leftrightarrow$$
$$1+3z+7z^2+2z^3-\sqrt{1+8z+27z^2+50z^3+53z^4+28z^5+4z^6} < 2z^3 \Leftrightarrow$$
$$1+3z+7z^2 < \sqrt{1+8z+27z^2+50z^3+53z^4+28z^5+4z^6} \Leftrightarrow$$
$$(1+3z+7z^2)^2 < (\sqrt{1+8z+27z^2+50z^3+53z^4+28z^5+4z^6})^2 \Leftrightarrow$$
$$49z^4+42z^3+23z^2+6z+1 < 1+8z+27z^2+50z^3+53z^4+28z^5+4z^6 \Leftrightarrow$$
$$0 < 2z+4z^2+8z^3+4z^4+28z^5+4z^6$$

which is true. Therefore, $\phi_1 < z$. We also have to check whether $\phi_1 < 0$. To verify that, we take the constant term of P_1 in order to see under which cases it is positive. Let,

$f(z) = 1+2z+2z^2-4z^3$ with the derivative being $f'(z) = 2+4z-12z^2 = 2(1+2z-6z^2)$

The discriminant of $f'(z) = 0$ is 28 and the roots are:

$$z_{1,2} = \begin{cases} \dfrac{1+\sqrt{7}}{4} \approx 0.6 \\ \\ \dfrac{1-\sqrt{7}}{4} \end{cases}$$

Since the coefficient of z^2 is negative, for $z > 1$, $f'(z) < 0$ meaning that $f(z)$ is downward sloping. $f(1) = 1$ and $f(1.14) = -0.046976$, therefore there is a $\mu \in [1, 1.14]$ such that $f(z) < 0 \ \forall \ z > \mu$. Since we consider $z > 2$, we have that $\phi_1 > 0 \ \forall \ z > 2$. Thus,

$$\phi_1 = \frac{1 + 3z + 7z^2 + 2z^3 - \sqrt{1 + 8z + 27z^2 + 50z^3 + 53z^4 + 28z^5 + 4z^6}}{2z}$$

Thus, $P_1 > 0$ for $\phi \in [\phi_1, z]$ and $P_1 < 0$ for $\phi \in [0, \phi_1]$ We will now focus on P_2. Expanding P_2:

$$P_2 = -2 - z + 26z^2 - 2z^3 - 28z^4 + (-4 - 10z + 50z^2 + 66z^3 + 28z^4)\phi + (-4z + 6z^3)\phi^2.$$

The $P_2 = 0$ has a discriminant

$$\Delta_{P_2} = 4(4 + 12z - 67z^2 - 272z^3 + 75z^4 + 1410z^5 + 1957z^6 + 924z^7 + 196z^8)$$

Since the coefficient of ϕ^2 and the constant term have different signs, the $P_2 = 0$ has a positive and a negative root. The positive root is:

$$\phi_3 = \frac{(2 + 5z - 25z^2 - 33z^3 - 14z^4)}{2z(-2 + 3z)}$$

$$+ \frac{\sqrt{4 + 12z - 67z^2 - 272z^3 + 75z^4 + 1410z^5 + 1957z^6 + 924z^7 + 196z^8}}{2z(-2 + 3z)}$$

Suppose,

$$\phi_3 < z \Leftrightarrow$$

$$2 + 5z - 25z^2 - 33z^3 - 14z^4 + \sqrt{4 + 12z - 67z^2 - 272z^3 + 75z^4 + 1410z^5 + 1957z^6 + 924z^7 + 196z^8}$$

$$<$$

$$-4z^2 + 6z^3 \Leftrightarrow$$

$$2 + 5z - 21z^2 - 39z^3 - 14z^4 + \sqrt{4 + 12z - 67z^2 - 272z^3 + 75z^4 + 1410z^5 + 1957z^6 + 924z^7 + 196z^8}$$

$$< 0 \Leftrightarrow$$

$$\left(\sqrt{4 + 12z - 67z^2 - 272z^3 + 75z^4 + 1410z^5 + 1957z^6 + 924z^7 + 196z^8}\right)^2$$

$$< \left(-2 - 5z + 21z^2 + 39z^3 + 14z^4\right)^2$$

which leads to

$$-8z - 8z^2 + 94z^3 + 80z^4 - 88z^5 - 152z^6 - 168z^7 < 0$$

which is true for $z > 1$. Thus $P_2 > 0$ for $\phi \in [\phi_3, z]$ and $P_2 < 0$ for $\phi \in [0, \phi_3]$. In the main text we name ϕ_3 as ϕ_2, thus:

$$\phi_2 = \frac{(2 + 5z - 25z^2 - 33z^3 - 14z^4)}{2z(-2 + 3z)}$$

$$+\frac{\sqrt{4 + 12z - 67z^2 - 272z^3 + 75z^4 + 1410z^5 + 1957z^6 + 924z^7 + 196z^8}}{2z(-2 + 3z)}$$

To conclude, $sw_O - sw_E > 0 \ \forall \ \phi \in (\phi_1, \phi_2)$ and $sw_O - sw_E < 0 \ \forall \ \phi \in [0, \phi_1) \cup (\phi_2, z]$[28] □

Proposition A.4. *When the tax is on output there is a critical level of SR, call it* ϕ_{sw}*, such that* $\forall \phi \in [0, \phi_{sw}]$*,* $sw_O > \hat{sw}_O$*, while* $\forall \phi \in [\phi_{sw}, \phi_E]$*,* $sw_O < \hat{sw}_O$*. When the tax is on emissions,* $\forall \phi \in [0, \phi_E]$*,* $sw_E > \hat{sw}_E$*, i.e., welfare is always higher in the absence of regulator's commitment.*

Proof. We have already shown that sw_E decreases in ϕ. What about $\hat{sw}_E$. The derivative of $\hat{sw}_E$ is:

$$\frac{\partial \hat{sw}_E}{\partial \phi} = \frac{4\phi[32 + 72z + \phi(8 + 5\phi)] - 2(8 + 10\phi)(28 + 14z + \phi^2)}{4[32 + 72z + \phi(8 + 5\phi)]^2}$$

Let

$$N_1 = 4\phi[32 + 72z + \phi(8 + 5\phi)]$$

and

$$N_2 = 2(8 + 10\phi)(28 + 14z + \phi^2)$$

By expanding the expressions, we have:

$$N_1 = 20\phi^3 + 32\phi^2 + 288z\phi + 128\phi$$

$$N_2 = -20\phi^3 - 16\phi^2 - 280z\phi - 560\phi - 224z - 448$$
$$\therefore N_1 + N_2 = 16\phi^2 + (8z - 432)\phi - (224z + 448)$$
$$= 8[2\phi^2 + (z - 54)\phi - (28z + 56)]$$

$N_1 + N_2 = 0$ has two roots, a positive and a negative ones. The positive one is $\phi = 28$. Therefore, for $\phi \in [0, 28]$, $\hat{sw}_E$ is downward sloping. For $\phi > 28$, $\hat{sw}_E$ increases in ϕ. Suppose,

$$28 \leq \frac{7}{4} + z(6 + z) - \frac{1}{4}(1 + 2z)\sqrt{81 + 4z(11 + z)} \Leftrightarrow$$
$$112 \leq 7 + 4z(6 + z) - (1 + 2z)\sqrt{81 + 4z(11 + z)} \Leftrightarrow$$
$$(1 + 2z)\sqrt{81 + 4z(11 + z)} \leq -105 + 4z(6 + z)$$

When is $-105 + 4z(6 + z) > 0$. The aforementioned polynomial has two solutions, a positive and a negative. The positive one is $\frac{\sqrt{141} - 6}{2} \approx 2.93717$. Therefore, for $z \in \left(1, \frac{\sqrt{141} - 6}{2}\right)$ the polynomial is negative, therefore $28 > \phi_E$. For $z > \frac{\sqrt{141} - 6}{2}$

[28] Due to the complexity in the calculations, we use *Mathematica* software to show that $\phi_1, \ \phi_2 < \phi_E$ as well as to show that $\phi_1 < \phi_2$

the polynomial is positive. Therefore, we square the two parts of the inequality to get:

$$[(1+2z)\sqrt{81+4z(11+z)}]^2 \leq [-105+4z(6+z)]^2 \Leftrightarrow$$
$$16z^4+192z^3+504z^2+368z+81 \leq 11025-5040z-264z^2+192z^3+16z^4 \Leftrightarrow$$
$$10944-5408z-768z^2 \geq 0 \Leftrightarrow$$
$$32(-24z^2-169z+342) \geq 0$$

which has a positive and a negative root. The positive one is $\frac{29\sqrt{73}-169}{48} \approx 1.64117$. For $z \in \left(1, \frac{29\sqrt{73}-169}{48}\right)$ the polynomial is positive, else negative. Since we consider $z > 2$ in order to avoid uninteresting cases, we have that $28 > \phi_E$, meaning that $\frac{\partial \hat{sw}_E}{\partial \phi}\ \forall\ \phi \in [0, \phi_E],\ z > 1$. Now, we will compare the two expressions sw_E and $\hat{sw}_E$. Since the two functions monotonically decrease, if there is an intersection between them in the $[0, \phi_E]$ interval, the $sw_E - \hat{sw}_E$ will be positive for some values of ϕ and negative otherwise. So,

$$sw_E(0) - \hat{sw}_E(0) =$$
$$\frac{56z^4+212z^3+218z^2+87z+6}{4[-12z(z+1)-2]^2} - \frac{28+14z}{2(32+72z)} =$$
$$\frac{(56z^4+212z^3+218z^2+87z+6)(32+72z)-2(28+14z)\{[-12z(z+1)-2]^2\}}{4[-12z(z+1)-2]^2(32+72z)}$$

We focus on the numerator, since the denominator is positive. Let:

$$N_1 = (56z^4+212z^3+218z^2+87z+6)(32+72z)$$

and

$$N_2 = -2(28+14z)\{[-12z(z+1)-2]^2\}$$

By expanding these expressions, we get:

$$N_1 = 192+3216z+13240z^2+22480z^3+17056z^4+4032z^5$$

$$N_2 = -224-2800z-12096z^2-21504z^3-16128z^4-4032z^5$$
$$\therefore N_1+N_2 = -32+416z+1144z^2+976z^3+928z^4$$

which is positive for $z > 1$. Therefore, $sw_E(0) - \hat{sw}_E(0) > 0$. By ignoring the corner solutions and keeping the functional form of sw_E and $\hat{sw}_E$ just to see the relative positioning of the two curves, we substitute $\phi = \frac{z+1}{z}$. The reason that we do not substitute ϕ with ϕ_E is that the resulting expressions are complex. After performing some algebraic manipulations:

$$sw_E\left(\frac{z+1}{z}\right) = \frac{-6+z(1+2z)[-4+z(16+7z)]}{4(-1+3z+6z^2)^2}$$

$$\hat{sw}_E\left(\frac{z+1}{z}\right) = \frac{1+z(2+z)(1+14z)}{2\{5+9z[2+z(5+8z)]\}}$$

So,

$$sw_E\left(\frac{z+1}{z}\right)-\hat{sw}_E\left(\frac{z+1}{z}\right)=$$

$$\frac{2\{-6+z(1+2z)[-4+z(16+7z)]\}\{5+9z[2+z(5+8z)]\}-4[1+z(2+z)(1+14z)](-1+3z+6z^2)^2}{8(-1+3z+6z^2)^2\{5+9z[2+z(5+8z)]\}}$$

The denominator is positive, therefore we focus on the numerator. Let:

$$N_1=2\{-6+z(1+2z)[-4+z(16+7z)]\}\{5+9z[2+z(5+8z)]\}$$

and

$$N_2=-4[1+z(2+z)(1+14z)](-1+3z+6z^2)^2$$

By expanding the above-mentioned expressions and adding them, we get:

$$N_1+N_2=-56-260z-536z^2-886z^3+824z^4+4830z^5+6876z^6+2016z^7$$

$$\frac{\partial(N_1+N_2)}{\partial\phi}=-260-1072z-2658z^2+3296z^3+24150z^4+41256z^5+14112z^6$$

$$\frac{\partial^2(N_1+N_2)}{\partial\phi^2}=-1072-5316z+9888z^2+96600z^3206280z^4+84672z^5$$

$$\frac{\partial^3(N_1+N_2)}{\partial\phi^3}=-5316+19776z+289800z^2+825120z^3+423360z^4>0\ \forall\ z>1$$

meaning that $\frac{\partial^2(N_1+N_2)}{\partial\phi^2}$ increases in z. For $z>1$, $\frac{\partial^2(N_1+N_2)}{\partial\phi^2}>0$ meaning that $\frac{\partial(N_1+N_2)}{\partial\phi}$ also increases in z. Moreover, $\frac{\partial(N_1+N_2)}{\partial\phi}>0$ for $z>1$. So, N_1+N_2 increases in z. Finally, $N_1+N_2>0$ for $z>1$. Thus,

$$sw_E\left(\frac{z+1}{z}\right)>\hat{sw}_E\left(\frac{z+1}{z}\right)\ \forall\ \phi\in\left[0,\frac{z+1}{z}\right]$$

We will see later that:

$$\frac{z+1}{z}>\frac{7}{4}+z(6+z)-\frac{1}{4}(1+2z)\sqrt{81+4z(11+z)}$$

for some values of z and

$$\frac{z+1}{z}<\frac{7}{4}+z(6+z)-\frac{1}{4}(1+2z)\sqrt{81+4z(11+z)}$$

for some other values of z.

Finally, we will see the relative positioning of the sw_O and $\hat{sw}_O$. First, the derivative of $\hat{sw}_O$ is:

$$\frac{\partial\hat{sw}_O}{\partial\phi}=\frac{-4[2z(\phi-4)+4\phi]}{[8+z(\phi-4)^2+2\phi^2]^2}$$

We focus on the numerator

$$2z(\phi-4)+4\phi<0\Leftrightarrow$$

$$(2z+4)\phi < 8z \Leftrightarrow$$
$$\phi < \frac{8z}{2z+4} = \frac{4z}{z+2}$$

Therefore, for $\phi < \frac{4z}{z+2}$, $\hat{sw}_O$ increases in ϕ. For $\phi > \frac{4z}{z+2}$, $\hat{sw}_O$ decreases in ϕ. We will now check if the maximum of $\hat{sw}_O$ lies within the allowed bounds of ϕ. Suppose,

$$\frac{4z}{z+2} \leq \frac{z+1}{z} \Leftrightarrow$$
$$4z^2 \leq (z+1)(z+2) \Leftrightarrow$$
$$4z^2 \leq z^2 + 3z + 2 \Leftrightarrow$$
$$3z^2 - 3z - 2 \leq 0$$

. For which z is this true ? The above-mentioned polynomial has a positive and a negative solution. The positive one is $\frac{3+\sqrt{33}}{6} \approx 1.457$. For $z > \frac{3+\sqrt{33}}{6}$, $\frac{4z}{z+2} \leq \frac{z+1}{z}$ and since we consider $z > 2$, this condition is satisfied. Therefore, $\hat{sw}_O$ increases in ϕ for $\phi > \frac{4z}{z+2}$, $z > 2$. For $\phi > \phi^*$, sw_O decreases, whereas $\hat{sw}_O$ increases. For $\phi = \phi^*$, we have:

$$sw_O(\phi^*) = sw^* = \frac{2+z}{4+10z}$$

and

$$\hat{sw}_O(\phi^*) = \frac{4(2+z)}{17+z(46+9z)}$$

Which one is greater? Suppose,

$$\frac{2+z}{4+10z} > \frac{4(2+z)}{17+z(46+9z)} \Leftrightarrow$$
$$17 + z(46+9z) > 4(4+10z) \Leftrightarrow$$
$$17 + 46z + 9z^2 > 16 + 40z \Leftrightarrow$$
$$9z^2 + 6z + 1 > 0$$

which is true. Therefore, $sw_O(\phi^*) > \hat{sw}_O(\phi^*)\ \forall\ z > 2$ We now compare the social welfare functions in $\frac{z+1}{z}$:

$$sw_O\left(\frac{z+1}{z}\right) = \frac{6z^3 - 4(z+1)z^2 + 8z^2 - 2(z+1)^2 z - 8(z+1)z + 6z - 4(z+1)^2 + 1}{2(1+2z)^2}$$
$$= \frac{-3 - 12z - 12z^2}{2(1+2z)^2}$$
$$= -\frac{3}{2}$$

and

$$\hat{sw}_O\left(\frac{z+1}{z}\right) = \frac{4}{8+z(-4+\frac{z+1}{z})^2+2\left(\frac{z+1}{z}\right)^2}$$

$$= \frac{4z^2}{2+5z+4z^2+9z^3}$$

and since this is positive it is greater than $sw_O\left(\frac{z+1}{z}\right)$. Thus, by Bolzano's Theorem there is a $\phi_{sw} \in (0, \phi_c)$ such that $sw_O(\phi_{sw}) = \hat{sw}_O(\phi_{sw})$ and $sw_O > (<)\hat{sw}_O$ for $\phi < (>)\phi_{sw}$.[29]

□

A.2 Movement of the Quantity expressions

When the tax-base is output, higher environmental consciousness leads to higher quantity in equilibrium, whereas when the tax is on emissions, equilibrium quantity is decreasing in ϕ.

Proof. The derivative of $q_O = \frac{1+2z}{1+2z(-z\phi+z+2)}$ is:

$$\frac{\partial q_O}{\partial \phi} = \frac{-[-2z^2(1+2z)]}{[1+2z(-z\phi+z+2)]^2}$$

$$= \frac{2z^2(1+2z)}{[1+2z(-z\phi+z+2)]^2}$$

which is positive $\forall \phi,\ z$. Therefore q_O increases in ϕ

The derivative of $q_E = 1 - \frac{5z(2z+1)}{12z(z+1)-\phi+2}$ is:

$$\frac{\partial q_O}{\partial \phi} = -\frac{-(-1)[5z(2z+1)]}{[12z(z+1)-\phi+2]^2}$$

$$= -\frac{[5z(2z+1)]}{[12z(z+1)-\phi+2]^2}$$

which is positive $\forall \phi,\ z$. Therefore q_E decreases in ϕ. □

A.3 Expressions for tax-rate and social welfare functions

Tax Rates:

$$t_O = \frac{2z^2(\phi-1)-z(\phi-3)\phi+2\phi+1}{2z(z(\phi-1)-2)-1}$$

$$t_E = \frac{-4z(z+6)\phi+4z(4z-1)+2\phi^2-7\phi-4}{24z(z+1)-2\phi+4}$$

[29]It is the case that $\phi_{sw} < \phi_E$ but due to the complexity of the calculations, we use *Mathematica* software to show that this is the case.

Social Welfare functions:

$$sw_O = \frac{2z\{z[-z(\phi-1)(\phi+3)-2\phi(\phi+2)+4]+3\}+1}{2(2z(-z\phi+z+2)+1)^2}$$

$$sw_E = \frac{56z^4+212z^3-16\left(z^2+z+1\right)\phi+218z^2+87z-6\phi^2+6}{4[-12z(z+1)+\phi-2]^2}$$