CREATE

Centre de Recherche en économie de l'Environnement, de l'Agroalimentaire, des Transports et de l'Énergie

Center for Research on the economics of the Environment, Agri-food, Transports and Energy

Certification as a Rationale for Voluntary Agreements

Patrick González

Cahier de recherche/Working Paper **2011-2**

Novembre/November 2011

González : CREATE, GREEN and Département d'économique, Université Laval
pgon@ecn.ulaval.ca

Abstract: I model the participation of firms in a voluntary agreement as a costly certification process whereby a firm informs the Regulator of its pollution intensity. Without this knowledge, the Regulator imposes the same tax on all firms in a heterogeneous industry, unduly hurting the clean ones with the lowest intensity. Certification allows clean firms to get a tax rebate. It also entails an informational externality as the dispersion of types decreases within the pool of non-participating firms, following an unraveling process. Because participation is a firm's private decision, there is such a thing as a bad voluntary agreement.

Résumé: Je modélise la participation des entreprises à un accord volontaire comme un processus coûteux de certification suivant lequel une entreprise informe le Régulateur de l'intensité de sa pollution. Le Régulateur impose la même taxe pigouvienne à l'ensemble des entreprises de l'industrie dont il ignore l'intensité, en affectant inopinément les plus propres. La certification permet à ces dernières d'obtenir un congé de taxes. Elle génère en outre une externalité informationnelle puisque la disparité des différents types d'entreprise au sein des entreprises non-participantes diminue, selon une cascade informationnelle. Puisque la décision de participer ou non à l'accord revient à l'entreprise, il se peut qu'un accord volontaire soit néfaste.

Environmental regulation requires information. To adequately command or incite firms to mitigate the impact of their production on the environment, a Regulator must learn about their technology. That learning phase is an integral part of the regulation process even if it proceeds at a time when the firms are not actually submitted to any regulation. Firms may welcome or lobby against a particular regulation; in any event, they understand that once it is in place, they will have to cope with it. It is at the time of its conception that they have more clout to influence its design through their exchanges with the Regulator.

A voluntary agreement whereby the firms and the Regulator address an industry wide externality is a valuable forum for these exchanges within the learning phase alluded to above. In support of this view, Price and al. (2003, p.2) note that

> *Supporting programs and policies [...], such as entreprise audits, assessments, benchmarking, monitoring, information dissemination, and financial incentives all play an important role in assisting the participants in meeting the target goals* [of a voluntary agreement].

In this paper, I show that there is an inherent logic that compels firms to share knowledge with the Regulator on a voluntary basis. To focus on the issue of information, I consider an abstract idealization of a voluntary agreement: a certification process whereby a firm informs the Regulator about its pollution intensity.

Participation in a voluntary agreement entails a loss in current profits but firms may prefer a small reduction of profits today to avoid a larger one in the future (Segerson and Miceli 1998). As for the Regulator, it might prefer the volun-

tary approach to traditional regulation because it economizes on transaction costs (Lévêque 1999, Segerson and Miceli 1999).

Transaction costs account for any factor likely to impact the efficiency of the regulatory process. For example, Schmelzer (1999) analyzes a case where firms can observe each others' emissions at a lower cost than the Regulator. The Regulator then economizes on monitoring cost by delegating this function to the industry. A voluntary agreement is then an economic way to regulate. But, it is not clear that the industry can regulate itself at a lower cost. Contrary to the Regulator, the industry must reconcile the conflicting incentives of individual firms (Ashby and als, 2004; Dawson and Segerson, 2008).

In this paper, a firm participates to ensure a lighter regulation not for the industry but for itself. The Regulator fosters a voluntary agreement to gather information in the short-run while no coercive regulation is in place. My approach combines transaction costs with incomplete information about the firms polluting intensities. Firms differ in their pollution intensity, but the Regulator cannot perceive these different types. To reduce the emissions, the Regulator imposes a uniform tax upon the firms' outputs regardless of their types (the second-best taxation scheme since a firm's pollution intensity is not related to any of its other observable attributes). Hence, the cleanest firms support a disproportionate share of the tax burden. To avoid the impact of an undiscriminating regulation, clean firms have an incentive to join a voluntary agreement and inform the Regulator about their type.

Heyes (2005) makes a similar argument with abatement standards instead of

tax rates. More recently, Denicolò (2008) has shown that the early adoption of a clean technology by a firm could be an informative signal to the Regulator that it should be made mandatory for the whole industry. Both of these papers use standard signaling games, in which the sender transfers only information about himself. The analysis here goes quite a few steps further by noting that participation entails an informational externality as the statistical properties of the pool of non-participating firms change. Given an expected uniform tax rate, only the relatively clean firms participate so that the dispersion of types within the pool of non-participating firms decreases. That externality is quite strong as it induces an unraveling process that may lead all firms to willingly certify their type, eliminating the informational problem altogether.

The popularity of voluntary agreements has faltered over the years (Morgenstern et als., 2007). Voluntary agreements may result in little gain if we take into account the transaction costs (Croci, 2005). I provide an explanation for this phenomenon to the extent that the marginal social value of participation to the agreement is endogenous. Participation is a private decision that the Regulator does not always deem desirable. I show that, except in a quite particular case, there is always a subset of participating firms whose contribution to social welfare is lesser than it would be without a voluntary agreement. If these firms are sufficiently numerous, instituting a voluntary agreement may actually lead to a decrease in welfare.

I present the model in the next section. At first, I consider the case of costless certification that yields full participation. I show how the voluntary agreement

model presented here is related to the literature on regulation under incomplete information according to the timing at which the options to enforce a regulation and to certify a firm's technology become available. Since the argument about bad voluntary agreements rests on transaction costs, I then consider the case of costly certification in section 2 with a two-periods version of the model. In section 3, I explain how the institution of a voluntary agreement could decrease social welfare. I prove that it can happen with a numerical example in a mathematical appendix.

In this paper, the informational externality is "continuous" in the sense that the participation of a small group of firms has a continuous effect on the Regulator's beliefs. This formalization is done for tractability, but I suspect that it does not represent well all plausible cases. I discuss this issue in the conclusion.

1   First and Second Best Regulation

I model a competitive single-good market with a perfectly elastic demand at price p and a population of firms with strictly positive, increasing and continuously differentiable supply functions for all positive prices. Let $s(p)$ denote a firm's supply. Its profit function π (net of fixed costs) is strictly increasing and strictly convex and, by Hotelling's lemma, it has s for derivative.

A firm is defined by its profit function and its pollution intensity (damage per unit produced) which I denote d. The intensities are distributed over an interval $\mathcal{D} = [d_0, d_1]$, with positive measure over any subinterval. I assume that the price p of the good is greater than d_1 so that all firms have a positive, socially efficient

level of production.[1]

Let $\hat{d}$ denote the expected pollution intensity conditional on being *no less* than d. Hence, the expected intensity can be written $\hat{d}_0$. I assume that this function is increasing and continuous.[2] There are positive measures of firms below and above the mean so that

$$d_0 < \hat{d}_0 < d_1.$$

Consider a firm (π, d) whose supply s equals the industry's *average* supply among firms with a given pollution intensity d. I assume that the intensities d are distributed independently of the firms' other attributes (the profit functions) so that $\hat{d}_0 s(p)$ denotes the average damage in the industry. When the Regulator imposes a unit tax t, the average supply decreases to $s(p - t)$. Average profit decreases to $\pi(p - t)$ but part of this reduction is the tax revenue $T = ts(p - t)$ that has no bearing on social surplus. On average, the social value of production of firms with intensity d amounts to[3]

$$\pi(p - t) - (d - t)s(p - t). \tag{1}$$

If we decompose profits as revenues minus costs, the tax revenues cancel out so that the production level determines the social value of production. But since supply strictly decreases with the tax rate, we can equivalently use either production

---

[1]To simplify the analysis, I discard consumer surplus. Assume that all production is exported abroad.

[2]In the conclusion, I show that these are actually strong economic assumptions.

[3]In the rest of the paper, π and s shall always refer to the average profit and supply functions keeping constant the intensity level d.

or the tax rate as an instrument to maximize (1). Doing so with the tax rate, we get the first-order condition[4]

$$(d - t^*)s'(p - t^*) = 0.$$

The optimal pigovian tax $t^* = d$ yields the first-best social value of production $\pi(p - d)$. Let

$$S^* = E(\pi(p - d)) > 0$$

denote the first-best average social value of production under complete information.

From the perspective of the Regulator, the environment is a credence good: the real damage to the environment that results from the firms' activities is *a priori* unknown. For instance, the level of harmful emissions produced by the use of a chemical may depend on the way a firm processes it, an information likely to be known only by the firm itself. To impose a pigovian discriminatory tax on a firm, the Regulator must know its pollution intensity. Since it is unrelated to the observable variables, either supply or profits, there is no point in having a discriminatory tax rate on either of these. Hence, without information about d, the best the Regulator can do is to set a uniform tax t that maximizes the expected social value of production

$$S(t) = \pi(p - t) - (\hat{d}_0 - t)s(p - t).$$

---

[4]The second order condition $s'(p - t^*) > 0$ is always satisfied.

The laissez-faire option is the special case where $t = 0$ and welfare amounts to $S(0)$. The second-best tax t^{**} solves the first-order condition

$$(\hat{d}_0 - t^{**})s'(p - t^{**}) = 0,$$

so that the best uniform tax level equals the *average* intensity. With such a tax, each firm pays the expected damages of its production

$$T^{**} = t^{**}s(p - t^{**}) = \mathrm{E}(ds(p - t^{**}))$$

and the expected social value of production reaches

$$S(t^{**}) = \pi(p - \hat{d}_0).$$

Since $t = t^{**}$ is socially preferred to $t = 0$, welfare increases: $S(t^{**}) > S(0)$. Jensen's inequality implies that

$$S^* = \mathrm{E}(\pi(p - d)) > \pi(p - \hat{d}_0) = S(t^{**}).$$

To summarize: $S^* > S(t^{**}) > S(0)$. At the firm level, though,

$$d_0 < \hat{d}_0 < d_1,$$

or equivalently

$$t^*(d_0) < t^{**} < t^*(d_1).$$

The second-best tax is higher (lower) than the first-best tax for clean (dirty) firms. As a consequence, clean (dirty) firms produce less (more) and make less (more) profit with t^{**} than they should

$$s(p - d_0) > s(p - t^{**}) > s(p - d_1),$$
$$\pi(p - d_0) > \pi(p - t^{**}) > \pi(p - d_1).$$

The second-best tax improves social welfare with comparison to the laissez-faire option but it is less efficient than the first-best scheme because it unduly constrains the production of relatively clean firms and it allows dirty firms to produce more than they should.

1.1 The First-Best for Free

The Regulator lacks information about the pollution intensity but that information is available from the firms themselves, although not all firms have an incentive to provide it: those for which $d > t^{**}$ would pay a higher rate with an optimal tax. But there is a solution to that problem: privatize the public bad so that an overtaxed firm has an incentive to inform the Regulator of its lesser than average pollution intensity.

Suppose the firms believe that the Regulator will set the uniform tax rate to

$$t_1 = t^{**} = \hat{d}_0.$$

All the clean firms whose polluting intensities are lesser than $\hat{d}_0$ then have an incentive to inform the Regulator that such a tax would unduly destroy social welfare in their case.[5] They would ask for discriminatory tax rebates $t_1 - d$ so that their effective tax rates would equal the first-best rates $t^*(d) = d$.

These private incentives generate an informational externality. If the firms within the left portion of the distribution come forward, certify their type and ask for a discriminating rate, the Regulator will learn that the conditional distribution of intensity within the remaining population is bounded below by $\hat{d}_0$ and it will set a new uniform rate

$$t_2 = \hat{t}_1 = \hat{\hat{d}}_0$$

for those firms. Since the conditional expectation of intensity within this subset is higher than $\hat{d}_0$, that uniform rate will be higher, $t_2 > t_1$, and all firms for which $t_1 \leq d < t_2$, that stayed silent in the first round will now come forward, show compelling evidence about their type and also ask for a discriminating rate. Again, the conditional expected intensity in the remaining population will shift to

---

[5]That the "cleanest" firms have an incentive to signal their type to the Regulator is in stark contrast with the results of Heyes (2005) who considers standards of abatement instead of pigovian taxes. There, a uniform average standard of abatement turns out to be uneconomical for the firms that have a higher marginal cost of abatement, presumably the "dirtiest" ones. They are the ones who benefit from a discriminatory, welfare-enhancing policy and who have an incentive to inform the Regulator of their type.

the right and the Regulator will revise its rate to

$$t_3 = \hat{t}_2 = \hat{\hat{d}}_0.$$

Again, $t_3 > t_2 > t_1$. This cascade effect stops when the tax rate reaches $t = d_1$, since $d_1 = \hat{d}_1$. At this point, all firms have willingly revealed their types and the Regulator is able to implement the first-best pigovian tax scheme at no cost by imposing a uniform tax rate $t = d_1$ and offering each firm a discriminatory tax rebate $t - d$.

This is an instance of Viscusi's (1978) unraveling process in markets with incomplete information. This process is quite powerful. Consider the simpler case of a monopoly that sells a zero cost good to an heterogenous continuum of buyers who may buy either one or zero unit. The buyers are differentiated by their willingness to pay d in $\mathcal{D}$. With no information about these valuations, the best the monopoly can do is to set a single price within $\mathcal{D}$ since all buyers will buy at the lowest posted price anyway. But with certification, the monopoly can screen all buyers by selling at the maximum price d_1 unless a buyer voluntarily certifies their valuation: such a buyer would then get a personalized rebate that covers the difference between d_1 and their willingness to pay. With no certification cost, all buyers would apply for a rebate and the monopoly would gather all the surplus in the market. Now, if the idea of "certifying" one's willingness to pay is dubious, that of certifying a firm's technology is routinely put into practice.

Back to the voluntary agreement model, we see that this result rests on two

assumptions:

1. The *Coercion* assumption: the Regulator can constrain the firms' production choices directly or with taxes.

2. The *Certification* assumption: firms can certify their type.

Without the Coercion assumption, the Regulator is pretty much emasculated; without the Certification assumption, it would have to rely on soft (non-verifiable) information of dubious quality since all firms would state that they have the lowest possible intensity in order to get the highest possible rebate.

Both assumptions make sense in the long-run since the Regulator channels vast powers and information diffusion is an increasing-return technology (the more people know something, the easier for someone else to learn it). In that sense, the regulation problem is acute in the short-run when either one of these assumptions does not hold.[6] Within a given timeframe, it is sufficient that the Coercion assumption will eventually hold to ensure that the regulatory process has some bite (Segerson and Miceli 1998). The only question is whether it holds before or after the Certification assumption.

The traditional literature on regulation under incomplete information assumes that the Coercion assumption holds before the Certification assumption (the possibility of certification is assumed away). To rationalize the recourse to voluntary agreements, I shall assume the opposite: In the next section, I develop a

---

[6]Regulation is not "free" in the long-run to the extent that new information asymmetries continuously emerge.

two-periods version of the model where the the Certification assumption holds sooner than the Coercion assumption, within the very specific setting of a voluntary agreement.

2 A Rationale For Voluntary Agreements

Consider a two-periods horizon with no discounting.[7] The Regulator can constrain the firms' production choices with taxes only in the second period. As before, a firm's type (its pollution intensity) is private information but signalling that type is now costly. Without a voluntary agreement, the firms produce and pollute at will in the first period and the uninformed Regulator imposes the second-best uniform tax in the second period. The social surplus amounts to the benchmark level $S(0) + S(t^{**})$.

A voluntary agreement is more than just a slogan as it needs human resources to work.[8] In this paper, I define a voluntary agreement as the most economical institutional arrangement through which the firms can communicate information to the Regulator in the short-run. I idealize it as a certification process (a perfect audit) of a firm's type that costs $c > 0$ (a deadweight loss) per unit produced.[9] Once certified, a firm's type becomes common knowledge. As in the previous

---

[7]I discard discounting to simplify the analysis but there is little point to introduce it anyway unless one assumes that the lengths of the two-periods (short and medium terms) are the same.

[8]In the Danish agreement scheme on industrial energy efficiency, where participating firms enjoyed a lower tax rate, the cost involved amounted up to 330,000 € per firm. Those of the Project XL in the United States were of the same magnitude (OECD, 2003).

[9]This makes the certification cost a variable cost although it should have a fixed cost interpretation. I do this for tractability; in particular, this is an easy way to relate the certification cost to the firm's size. This assumption does not affect the results in a significant way.

section, a certified firm gets a tax rebate in the second period.

2.1 The Regulator's Problem

The Regulator sets the tax rates once it has the power to do so: at the beginning of the second period. I assume that it then maximizes expected welfare conditional on the information gathered during the voluntary agreement: it offers the first-best discriminatory rate to each certified firm (equal to their type) and the second-best uniform rate to the others, computed for that pool of firms. Hence, it cannot commit himself in the first period to set a high uniform tax to induce more participation (if the firms did not respond and it learned little, it would have an incentive to lower that rate). I specify this to emphasize the idea that the information brought by a voluntary agreement is too valuable to commit ex ante to discard it ex post (see the discussion in the conclusion).

The institution of a voluntary agreement is not a strategic variable here. I simply compare the levels of social welfare achieved with and without it. On the other hand, the private decisions of firms to participate or not is an essential part of the argument.

2.2 Participation

Suppose the firms expect the second period uniform rate to be t. A firm participates in the voluntary agreement if the long-run benefit $\pi(p - d) - \pi(p - t)$ of getting a tax rebate $t - d$ covers the short-run opportunity cost $\pi(p) - \pi(p - c)$ of

certification. The gain in participating (the difference) is denoted $V(d, t)$.

$$V(d, t) = \pi(p - d) - \pi(p - t) - (\pi(p) - \pi(p - c)).$$

Notice that V is convex, decreasing in d and increasing in t. A firm with pollution intensity d that expects a second period uniform rate t participates if $V(d, t) \geq 0$.

To maintain a simple information structure, I assume thereafter that all profit functions are identical up to an affine transformation. In the previous section, we had $c = 0$ so that the participation condition resumed to $d \leq t$. Hence, participation depended on d but not on π and did not affect the statistical independence property between d and π. This is no longer the case so that, theoretically, participation could cause observable profits (or supply) and pollution intensity to be somewhat correlated. It wouldn't be certain that a uniform tax, independent of π, would necessarily yield the second-best welfare within the pool of non-participating firms. But if all profit functions are identical to some function π up to an affine transformation, then they will share the same participation function V up to such transformation as well and their decisions to participate will differ, as before, only if their types differ.

The certification cost must not be too high so that the cleanest firm wishes to participate. Hence, I assume that

$$V(d_0, t^{**}) > 0. \tag{2}$$

Since V increases with t, this inequality holds as well for any higher tax rate. Be-

sides, the dirtiest firm d_1 will never want to pay for certification since the uniform rate is bounded above by its type: it cannot expect to get a tax rebate by getting certified. This implies that $V(d_1, t) < 0$. To resume,

$$V(d_0, t) > 0 > V(d_1, t) \quad \text{for all } t \in [t^{**}, d_1]. \tag{3}$$

By continuity, there exists a firm $\delta(t)$ that is indifferent about participating or not; *i.e.* for which $V(\delta(t), t) = 0$. Using the implicit function theorem, we verify that a higher expected tax incite more firms to participate; that is, δ strictly increases with t. It follows that all firms such that $d \leq \delta(t)$ participate.[10] A firm dirtier than the marginal firm $\delta(t)$ does not participate. For the marginal firm

$$\pi(p - \delta(t)) - \pi(p - t) = \pi(p) - \pi(p - c) > 0,$$

so that $\delta(t) < t$. Increasing d slightly, we find a bunch of non-participating firms for which $\pi(p - d) > \pi(p - t)$; that is, firms that would benefit in the long-run by participating yet choose to abstain to economize on the certification cost.

2.3 Existence of an Equilibrium

In a rational expectations equilibrium, the expected tax rate t_e induces a participation $\delta(t_e)$ such that the second-best uniform rate within the pool of non-participating firms is $\hat{\delta}(t_e) = t_e$.

---

[10]Although $\delta(t)$ bounds the subset of participating firms, it says nothing about its size which depends on the distribution of types within the population.

Since V decreases with d, the l.h.s. of condition (3) ensures that $\delta(t^{**}) > d_0$ so that $\hat{\delta}(t^{**}) > \hat{d}_0 = t^{**}$. The function $\hat{\delta}$ is a composition of two increasing functions on $\mathcal{D}$; so it is increasing as well and, by Tarski's theorem, it has a fixed point t_e that denotes both the equilibrium expected and the realized value of the uniform rate. If $t_e = t^{**}$, we get the contradiction $t^{**} = \hat{\delta}(t^{**})$. It follows that $\hat{\delta}(t_e) = t_e > t^{**}$. For further reference, I denote $d_e = \delta(t_e)$ the marginal firm at the equilibrium point.

The existence of an equilibrium is depicted in Figure 1. The space is $\mathcal{D}^2$: the horizontal axis represents the pollution intensity d and the vertical axis, the tax rate t. Two strictly increasing continuous functions are drawn. The function $\hat{d}$ maps each intensity level onto a revised tax rate. In particular, $\hat{d}_0 = t^{**}$ and $\hat{d}_1 = d_1$. The function δ maps tax rate into a marginal firm that bounds above the set of participating firms. It is undefined around d_0 since no firm would agree to pay the certification cost to avoid such a low rate. It is always below t because only firms with $d < t$ will eventually want to participate. Yet $\delta(t^{**})$ exists and is strictly positive since we have assumed that the certification cost is low enough so that type d_0 has a strict interest in participating. This implies that δ crosses the t-axis below t^{**}. At the other end of $\mathcal{D}$, $\delta(t) < t$ implies that $\delta(d_1) < d_1$. These two facts imply the existence of an equilibrium point, such as point z, where the two functions cross. There, each function yields the inverse of the other: $\hat{\delta}(t_e) = t_e$.
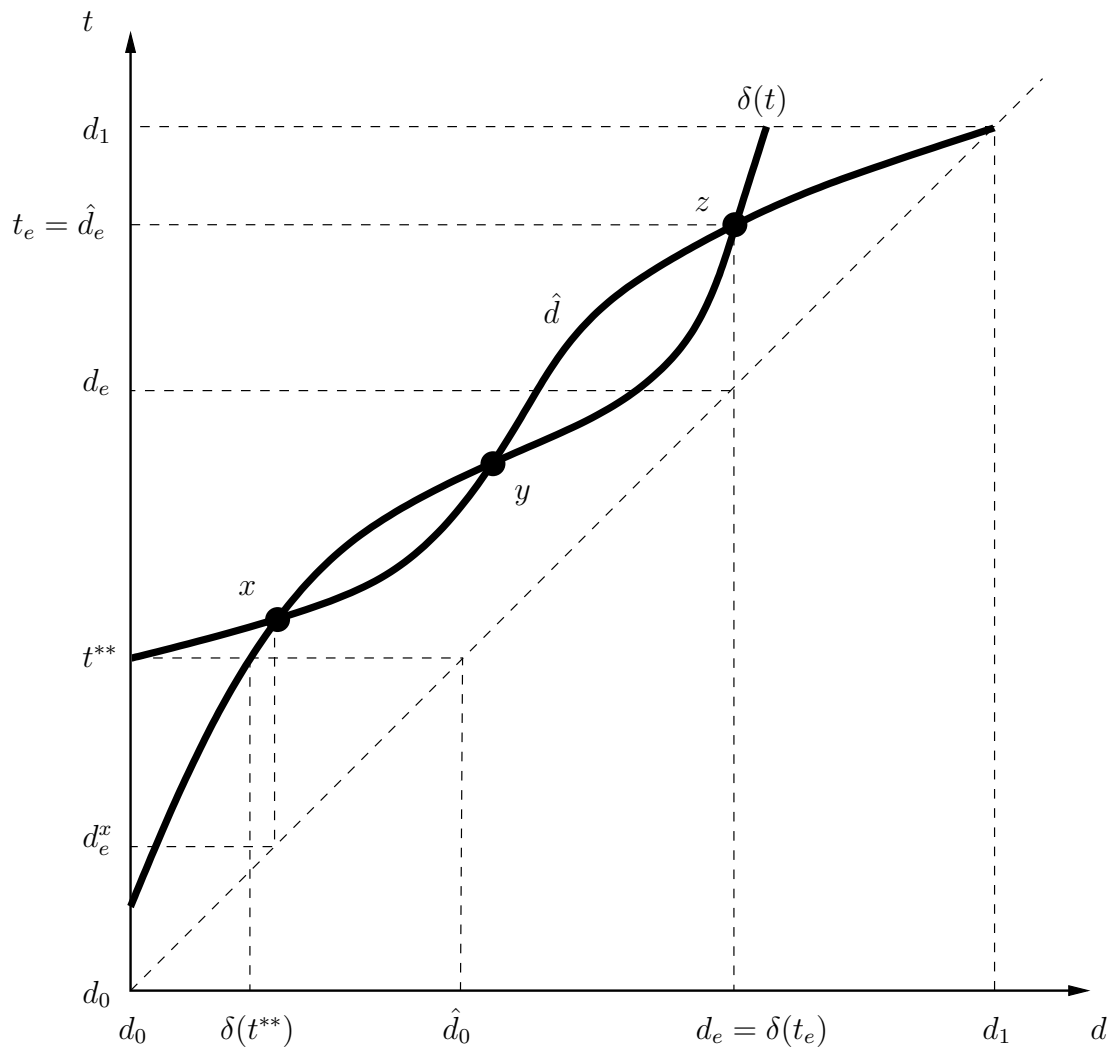
Figure 1: Equilibria.

2.4 Multiple Equilibria and Stability

Instances of multiple equilibria are possible. Recall that the higher the expected rate, the greater the participation. In equilibrium z, we have $d_e > t^{**}$. By this criteria, I shall refer to z as a high-participation equilibrium. By contrast, an equilibrium like x, where $d_e^x < t^{**}$, is a low-participation equilibrium (I use this characterization later).

The function δ does not depend on the distribution of types in any way. By contrast, the conditional expectation function $\hat{d}$ depends only on the distribution of types. So both curves are independent: given δ, we are free to draw any increasing function $\hat{d}$ that starts at point (d_0, t^{**}) and ends at point (d_1, d_1). As we consider different distributions, the equilibrium points are anywhere along δ within the points a and b. Stated differently, given any tax rate $t \in (t^{**}, d_1)$, there exists a distribution of types such that $t_e = t$ and $d_e = \delta(t)$. Likewise, given any participation level d between $\delta(t^{**})$ and $\delta(d_1)$, there exists a distribution of types such that $d_e = d$ and $t_e = \hat{d}$.

Multiple equilibria generically come in odd numbers as is the case in the figure. There are three equilibria in x, y and z where x and z are respectively the low and high equilibria I just described. Both are stable in the sense that if some firms expect a marginal discrepancy between the actual and equilibrium tax rates, their reaction will lead the Regulator to revise the rate by a lesser amount than this discrepancy. For instance, if the firms expect t to be $t_e - \Delta_1$, participation will decrease to $\delta(t_e - \Delta_1)$. But this should lead the Regulator to lower t by $\Delta_2 = t_e - \hat{\delta}(t_e - \Delta_1)$ which is less than Δ_1. Iterating on this reasoning, we see

that $\Delta_n \to 0$. The equilibrium y is unstable: any discrepancy between the firms' expectations about the tax rate and its equilibrium value will lead to an ever increasing or decreasing revision process until the expectations converge again to a new stable equilibrium.

3   Welfare Analysis

As the certification cost decreases, all firms eventually participate and welfare rises from $S(0) + S(t^{**})$ to $2S^*$. So a voluntary agreement enhances welfare if the certification cost is sufficiently low. The logically remaining questions are: given a positive certification cost, when does the institution of a voluntary agreement increases welfare? and can it reduce welfare?

   With certification, the total social contribution to welfare of non-participating firms always increases because, as the uniform tax rises, it gets closer on average to the marginal damage from their production.[11] It follows that if the institution of a voluntary agreement leads to a decrease of welfare, it is because the private benefit of certification for some firms is greater than its private cost, and lower than its social cost as I will show below. In that sense, a bad voluntary agreement involves too much participation. In addition, these firms must be sufficiently numerous so that their drag on social welfare overcomes the gains realized by the other firms. With a numerical example (in a mathematical appendix), I show that there exists such a thing as a bad voluntary agreement.

---

[11]That raise worsens the distortion for the cleanest ones among them, but their total contribution as a group necessarily rises.

In the rest of this section, I provide a partial characterization of good and bad voluntary agreements. It is partial because the outcome of a voluntary agreement depends on the distribution of types upon which I have laid out very few assumptions. Nevertheless, a lot can be inferred from the behavior of the cleanest firm. To get this characterization, I need two additional regularity assumptions about the supply function: (i) that it is concave and (ii) that the dirtier the firm, the higher its efficient level of total pollution $ds(p-d)$.[12]

Notwithstanding its effect on the information structure, participation entails two external effects that matter for the Regulator. A participating firm reduces its production by $s(p) - s(p-c)$ in the short-run and that brings an additional social benefit as pollution is proportionally reduced. In the long-run, the firm is taxed efficiently and the Regulator avoids the external social loss $(d - t^{**})s(p - t^{**})$ induced by the second-best uniform tax rate. Let

$$W(d) = d(s(p) - s(p-c)) + (d - t^{**})s(p - t^{**}),$$

an increasing affine function, resume these external effects.

A voluntary agreement increases a participating firm's contribution to social welfare if

$$V(d, t_e) + W(d) \geq \pi(p - t^{**}) - \pi(p - t_e). \tag{4}$$

The positive term on the r.h.s. is an adjustment to take into account that the Reg-

---

[12]These non-parametric restrictions on the technology may look obscure: Together, they imply that the dirtiest firm should not have to reduce its production by more than a half to be efficient.

ulator considers the benefit of participation in comparison to the base scenario, with no voluntary agreement, while each individual firm considers the benefit of participating given that such an opportunity exists and that the uniform tax rate is expected to rise up to t_e. [13]

Notice that the function $V + W$ is a convex function of d. Notwithstanding the distribution of types, a voluntary agreement increases welfare when (4) holds at the minimum of $V + W$ over the subset $[d_0, d_e]$ of participating firms. Otherwise, there exists a positive measure of types around the point that minimizes $V + W$ for which (4) does not hold. Again, if these types weigh sufficiently in the distribution, welfare will decrease.

The minimum of $V + W$ could be at either of the boundary points d_0 and d_e, or at an interior point d^* where its partial derivative is null. The partial derivative of $V + W$ with respect to d

$$s(p) + s(p - t^{**}) - s(p - c) - s(p - d) \tag{5}$$

amounts to the total *reduction* of production of a participating firm over the two-periods horizon.

Suppose $V + W$ reaches its minimum at d_0. This implies that (5) is positive

---

[13]Equation (4) is a rearrangement of the condition

$$\pi(p - c) - ds(p - c) + \pi(p - d) \geq \pi(p) - ds(p) + \pi(p - t^{**}) - (d - t^{**})s(p - t^{**})$$

that contrasts a firm's social contribution to welfare if it participates to the voluntary agreement (on the l.h.s.) to its contribution when no such agreement exists (on the r.h.s.). That inequality is independent of the equilibrium uniform tax rate. I subtract the term $\pi(p - t_e)$ on both sides to get (4).

at d_0 (the cleanest firm decreases its production). If, in addition, its social contribution to welfare increases; that is if (4) holds for type d_0, then it holds as well for all the other participating types to its right. Then a voluntary agreement unambiguously increases welfare. On the other hand, if (4) does not holds for type d_0, then it does not hold either for a subset of participating types to its right. Then a voluntary agreement could decrease welfare if this subset weights sufficiently in the distribution. In the two remaining cases, that of an interior minimum at d^* and that of a minimum at d_e, I show that (4) does not hold either for a subset of firms.

If the minimum is interior, there is a participating firm d^* that simply reallocates its production in time

$$s(p) + s(p - t^{**}) = s(p - c) + s(p - d^*). \tag{6}$$

Since it now produces less in the first period, it must be producing more in the second period. Hence $s(p - d^*) > s(p - t^{**})$ which implies that certification provides this firm a tax rebate $t^{**} - d^* > 0$. Again, so as to keep total production constant, $s(p - d^*) < s(p)$ implies that $s(p - c) > s(p - t^{**})$, so that $t^{**} > c$. Finally, the concave assumption (i) above implies that the tax rebate is lesser than the certification cost: $t^{**} - d^* < c$ (see the second appendix).

Now consider this property of convex functions: If π is strictly convex, $a > b$ and $0 < x < a - b$, then[14]

$$\pi(a) + \pi(b) > \pi(a - x) + \pi(b + x).$$

---

[14]The inequality is reversed if $x > a - b$.

Set $a = p$, $b = p - t^{**}$ and $x = c$, so that $a - b = t^{**} > c = x$, to get

$$\pi(p) + \pi(p - t^{**}) > \pi(p - c) + \pi(p - t^{**} + c),$$

and since $-t^{**} + c > -d^*$, that

$$\pi(p) + \pi(p - t^{**}) > \pi(p - c) + \pi(p - d^*),$$

or equivalently that

$$V(d^*, t_e) < \pi(p - t^{**}) - \pi(p - t_e). \tag{7}$$

Using (6), we have

$$W(d^*) = d^* s(p - d^*) - t^{**} s(p - t^{**}).$$

Because a dirtier firm should pollute more (assumption ii above) and $t^{**} > d^*$, then

$$W(d^*) < 0. \tag{8}$$

It follows from (7) and (8) that (4) does not hold at d^* nor over a subset of types around that value.

Lastly, if the minimum is at d_e, the derivative (5) is negative at d_e (the marginal

firm increases its production) so that

$$s(p) + s(p - t^{**}) < s(p - c) + s(p - d_e). \qquad (9)$$

This inequality implies that $t^{**} > d_e$ (the marginal firm gets a tax rebate). Multiply by d_e and subtract T^{**} on both sides to get

$$W(d_e) < d_e s(p - d_e) - t^{**} s(p - t^{**}).$$

Again, because a dirtier firm should pollute more (assumption ii above), the r.h.s. is negative so that $W(d_e)$ is negative as well. By definition, $V(d_e) = 0$ so that (4) does not hold at d_e. Thus, there is a subset of types to the left of d_e over which (4) does not hold either.

Notice that (9) cannot hold in a high-participation equilibrium. In such equilibrium, the uniform tax rate rises so much that the marginal firm is compelled to participate although it pays a higher tax and produces less in the end. With a zero certification cost, as in section 1, this is exactly what happens to the dirtiest firm.

4   Conclusion

From the firm's point a view, certification is an option to reallocate its production in time: a good reallocation involves a small reduction in current production (because of the certification cost) and a large expansion of future production; a dirty firm cannot expect a good reallocation and will not participate. From the Regula-

tor's point of view, such a reallocation makes sense for the cleanest firms but not for the others. In particular, there is no gain in certifying a type $d = t^{**}$ firm in the middle since it is already taxed at the first-best level. A voluntary agreement works as a deal between the cleanest firms and the Regulator.

However, the Regulator does not control entry so that, although the cleanest firms always participate, other parasitistic firms are drawn in as well. These firms waste the cost of certification to get a tax rebate without increasing social welfare since their total production does not change much. Only when the cleanest firms do contribute more by decreasing their total production can we be sure that certification will improve the contributions of all other participating firms; and that, because only then will we be sure that the latter will decrease their total production as well.

In this model, certification can only take place in the first period within the very specific setting of a voluntary agreement. Certification *is* the voluntary agreement: it stands for a sophisticated information-transmission mechanism that goes on within a voluntary agreement. Meetings, exchange of documents, on site controls, etc, by civil servants, and participating firms' employees result in the Regulator being informed, at a cost, about the firms' technology. Obviously, information about firms can come from other sources and at other moments in time, but the point is that proceeding early on a voluntary basis makes sense if the transaction costs are low. Allowing the firms to get certified at various points in time would obviously enrich the model (at the cost of getting more complex) and could provide insights on the issue of participation when there are multiple equilibria.

The Regulator may want to influence participation if the firms did not coordinate themselves on the best equilibrium in the first place.

Assuming that a voluntary agreement would enhance welfare, one way to do so would be for the Regulator to commit in advance to impose the (best) equilibrium uniform tax rate. Yet, the information structure does not always allows the Regulator to compute ex ante (as I do) the equilibrium second-best rate. Remember that I have assumed that the conditional expected intensity among the pool of non-participating firms was increasing and continuous as the marginal intensity (the type of the marginal firm) rises. These are strong assumptions as the following simple example will demonstrate.

Compare the following two cases. In the first case, the pollution intensity is distributed uniformly over $[0,1]$ so that $t^{**} = \frac{1}{2}$. If the cleanest firms within $[0, \Delta]$, where Δ is a small number, choose to participate, the ex post rate would rise marginally by $\frac{\Delta}{2}$ and that would motivate other firms to participate. In the second case, it is common knowledge that the distribution of firms is either $F(d) = d(2 - d)$ or $G(d) = d^2$ with equal probabilities. From an ex ante point of view, firms are distributed according to $\frac{1}{2}F(d) + \frac{1}{2}G(d) = d$ which is the uniform distribution. Since the ex ante distribution is the same as in the first case, t^{**} still equals $\frac{1}{2}$, but if the firms within $[0, \Delta]$ participate, the Regulator will immediately learn which of F or G is the true distribution. If it's F, the ex post tax will plunge to $\frac{1}{3} + \frac{2}{3}\Delta$; and if it is G, it will jump to $\frac{2}{3} + \frac{2}{3}\frac{\Delta^2}{\Delta+1}$. In the second case, the conditional expectation does not change continuously as participation increases and it is not even "increasing" or "decreasing" in a meaningful way. The informational

externality here has a discontinuous effect on the Regulator's beliefs. The point is that it would make little sense for the Regulator to commit ex ante to a second period tax rate. But the logic of certification remains unaltered: clean firms have a keen interest in coming forward and informing the Regulator about their type.

References

Ashby, Simon, Swee-Hoon Chuah, and Robert Hoffmann. 2004. "Industry Self-Regulation: A Game-Theoretic Typology of Strategic Voluntary Compliance." *International Journal of the Economics of Business*, 11(1): 91–106.

Croci, Edoardo. 2005. "The Economics of Environmental Voluntary Agreements." In *The Handbook of Environmental Voluntary Agreements*, ed. E. Croci, 3–30. Springer.

Dawson, Na Li, and Kathleen Segerson. 2008. "Voluntary Agreements with Industries: Participation Incentives with Industry-Wide Targets." *Land Economics*, 84: 97–114.

Denicolò, Vincenzo. 2008. "A signaling model of environmental overcompliance." *Journal of Economic Behavior & Organization*, 68(1): 293–303.

Heyes, Anthony G. 2005. "A signaling motive for self-regulation in the shadow of coercion." *Journal of Economics and Business*, 57(3): 238–246.

Lévêque, François. 1999. "Externalities, Collective Goods and the Requirement of a State's Intervention in Pollution Abatement." In *Voluntary Approaches in Environmental Policy. Economics Energy Environment*, , ed. Carlo Carraro and Fran cois Lévêque, 17–26. Kluwer Academic Publishers.

Morgenstern, Richard D., and William A. Pizer. 2007. *Reality Check: The Nature and Performance of Voluntary Environmental Programs in the United States, Europe, and Japan.* RFF Press.

OECD. 2003. "Policies to Reduce Greenhouse Gas Emissions in Industry — Successful Approaches and Lessons Learned: Workshop Report." OECD OECD and IEA Information Paper.

Price, Lynn, Jiang Yun, Ernst Worrell, Du Wenwei, and Jonathan E. Sinton. 2003. "Developement of an Energy Conservation Voluntary Agreement Pilot Project in the Steel Sector in Shandong Province." Ernest Orlando Lawrence Berkeley national Laboratory LBNL-51608.

Schmelzer, Dirk. 1999. "Voluntary Agreements in Environmental Policy: Negotiating Emission Reductions." In *Voluntary Approaches in Environmental Policy. Economics Energy Environment*, , ed. C. Carraro and F. Lévêque, 55–74. Kluwer Academic Publishers.

Segerson, Kathleen, and Thomas J. Miceli. 1998. "Voluntary Environmental Agreements: Good or Bad News for Environmental Protection?" *Journal of Environmental Economics and Management*, 36: 109–130.

Segerson, Kathleen, and Thomas J. Miceli. 1999. "Voluntary approaches to environmental protection: the role of legislative threats." In *Voluntary Approaches in Environmental Policy. Economics Energy Environment*, ed. C. Carraro and F. Lévêque, 105–120. Kluwer Academic Publishers.

Viscusi, W. Kip. 1978. "A Note on "Lemons" Markets with Quality Certification." *The Bell Journal of Economics*, 19(1): 277–279.

A Mathematical Appendix

A.1 Numerical example

The following numerical example describes instances of low and high-participation equilibria where social welfare is lower than the level reached without a voluntary agreement. All the firms have the same profit function $\pi(p) = p^2/2$, so that supply is a linear function of price $s(p) = p$. Set $p = 4$ and distribute d uniformly over $[0, 4]$, so that $t^{**} = 2$. If a firm is taxed at rate t, it produces $4 - t$ and contributes

$$\pi(p - t) - (d - t)s(p - t) = \frac{4 - t}{2}(4 + t - 2d)$$

to social welfare. Under the first-best scheme ($t = d$), it produces $4 - d$ and contributes $(4 - d)^2/2$. With second-best taxation ($t = 2$), it produces 2 and contributes $2(3 - d)$. When unregulated ($t = 0$), it overproduces 4 and contributes $4(2 - d)$.

Aggregated social welfare sums the contributions of each quartile of the population. For the three regimes, we get[15]

$$S^* = \int_0^4 \frac{(4-s)^2}{2} ds = 6.1\overline{6} + 3.1\overline{6} + 1.1\overline{6} + 0.1\overline{6} = 10.\overline{6},$$

$$S(t^{**}) = \int_0^4 2(3-s) \, ds = 5 + 3 + 1 - 1 = 8,$$

$$S(0) = \int_0^4 4(2-s) \, ds = 6 + 2 - 2 - 6 = 0.$$

Second best taxation improves the contributions of the second and third quartiles and reduces the social loss brought by the dirtiest firms, but at the expense of reducing the social contribution of the cleanest firms of the first quartile.

The private benefit to participate is

$$2V(d,t) = (4-d)^2 - (4-t)^2 + (4-c)^2 - 16.$$

Condition (2) commands that $c < 2$. Set $V(\delta(t), t) = 0$ to obtain the measure of participation $\delta(t)$

$$4 - \delta(t) = \sqrt{(4-t)^2 - (4-c)^2 + 16} \tag{10}$$

With participation $\delta(t)$, the second period rate rises from 2 to

$$\hat{\delta}(t) = \int_{\delta(t)}^4 \frac{s}{4 - \delta(t)} ds = \frac{4 + \delta(t)}{2}.$$

___

[15]Since the density is constant, I discard it in all computations. All reported aggregated payoffs are thus multiplied by 4.

Use (10) and the equilibrium condition $\hat{\delta}(t_e) = t_e$ to compute the equilibrium tax rate

$$4 - t_e = \sqrt{\frac{16 - (4 - c)^2}{3}}. \tag{11}$$

The equilibrium reluctant firm is $d_e = \delta(t_e)$. Plug back the equilibrium values into (10), and notice that

$$(4 - c)^2 = 16 - \tfrac{3}{4}(4 - d_e)^2. \tag{12}$$

In the first period, a participating firm contributes

$$\pi(p - c) - ds(p - c) = \frac{4 - c}{2}(4 - c - 2d).$$

In the second period, it efficiently contributes $(4 - d)^2/2$. The total welfare from participating firms amounts to

$$\int_0^{d_e} \frac{1}{2} \left[(4 - c)^2 - (4 - c)2d + (4 - d)^2\right] d\delta,$$
$$= \frac{d_e}{2} \left[(4 - c)^2 + cd_e + \tfrac{1}{3} d_e^2 - 8d_e + 16\right],$$
$$= \frac{d_e}{2} \left[cd_e - \tfrac{5}{12} d_e^2 - 2d_e + 20\right]. \tag{13}$$

(Use (12) in the last step to get rid of $(4 - c)^2$.)

A non-participating firm generates $4(2 - d)$ in the first period and

$$\frac{4 - t_e}{2}(4 + t_e - 2d)$$

in the second period. Substitute $t_e = (4 + d_e)/2$ and integrate to get the total welfare from non-participating firms

$$\int_{d_e}^{4} \frac{1}{8} \left[32(2 - s) + (4 - d)(12 + d - 4s) \right] ds = 8 - \frac{d_e}{2} \left[\frac{1}{4} d_e^2 - 7d_e + 28 \right]. \quad (14)$$

Total welfare with a voluntary agreement is the sum of (13) and (14) :

$$8 + \frac{d_e^2}{2}(c - m(d_e)), \quad (15)$$

where $m(\delta) = \frac{2}{3}\delta - 5 + \frac{8}{\delta}$. Re-parameterize the verification cost with a decreasing transformation on $\mathcal{D}$:

$$c(\delta) = 4 - \sqrt{16 - \frac{3(4 - \delta)^2}{4}}, \quad (16)$$

so that $d_e = \delta$ (the condition $c(\delta) < 2$ implies that $\delta > 0$ while $c(4) = 0$). This family of models is now indexed by a participation parameter δ instead of a certification cost c. Put differently, to get a participation of δ, we need to set a certification cost equal to $c(\delta)$. Substitute (16) for c in (15)

$$8 + \frac{\delta^2}{2} \left(c(\delta) - m(\delta) \right). \quad (17)$$

With no voluntary agreement, the Regulator achieves

$$S(0) + S(t^{**}) = 8.$$

Comparing this expression with (17), we see that the institution of a voluntary agreement decreases social welfare when $c(\delta) < m(\delta)$. This happens in any low-participation equilibrium (when $\delta < 2 = t^{**}$). Since $c(2) = 4 - \sqrt{15} < m(2) = \frac{1}{3}$, it also happens for some high-participation equilibria.

A.2 A participating firm that does not change its total production gets a tax rebate bounded by the certification cost

Equation (6) implies $t^{**} > c$ and $t^{**} > d$. Here I show that it also implies $t^{**} - d^* < c$. Suppose that $c \le d^*$. Since we have assume in i) that the supply function is concave, $-s$ is convex and we can apply the convexity property presented in the following paragraph of the text with $a = p$, $b = p - t^{**}$ and $x = c$ (so that $x = c < t^{**} = a - b$) to get

$$s(p - c) + s(p - t^{**} + c) > s(p) + s(p - t^{**}).$$

Compare this expression with (6) to conclude that $t^{**} - d^* < c$. If $c > d^*$, set $x = d^*$ and run the same argument.