Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER Journal of Environmental Economics and Management 51 (2006) 110–131

JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT

www.elsevier.com/locate/jeem

Optimal enforcement policy and firms' emissions and compliance with environmental taxes

Inés Macho-Stadler*, David Pérez-Castrillo

Department of Economics & CODE, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Received 22 April 2004
Available online 19 July 2005

Abstract

In a market where firms with different characteristics decide upon both the level of emissions and their reports, we study the optimal audit policy for an enforcement agency whose objective is to minimize the level of emissions. We show that it is optimal to devote the resources primarily to the easiest-to-monitor firms and to those firms that value pollution the less. Moreover, unless the budget for monitoring is very large, there are always firms that do not comply with the environmental objective and others that do comply; but all of them evade the environmental taxes.

Keywords: Environmental taxes; Optimal audit policy

1. Introduction

Environmental protection is a priority and a challenge in many countries. Economic activity generates negative external effects that producers do not internalize. Taxes and standards are the common policy instruments to regulate the environmental quality. The traditional approach to discuss the optimal environmental policy has been to assume that polluters comply with the environmental regulation. However, firms' compliance is not guaranteed.

*Corresponding author. Fax: +34 93 581 24 61
E-mail addresses: ines.macho@uab.es (I. Macho-Stadler), david.perez@uab.es (D. Pérez-Castrillo).

0095-0696/$ - see front matter
doi:10.1016/j.jeem.2005.06.001

The aim of this paper is to study the optimal audit policy in a situation where firms may evade environmental taxes. We analyze the effects of the possibility of evasion (combined with the optimal auditing policy) on the level of environmental emissions of a population of firms. We also study how the optimal policy varies with the characteristics of the firms.

We follow some recent environmental policy literature that has incorporated compliance issues.[1] We assume that the tax policy is not perfectly enforceable; in particular we consider that environmental taxes may be evaded by under-reporting emissions. This becomes possible when government monitoring is imperfect because firms cannot be monitored with high probability (it is costly), or because even when monitored, the true level of emissions of a firm is difficult to identify. Inspection policies combined with sanctions provide a key tool for the provision of incentives to reduce environmental deviations.

Cropper and Oates [5] define two types of environmental problems that may give rise to different environmental violations. First, pollution may arise from an accident: be caused by negligence or be a random act of nature. This type of problem is not considered in our paper. Second, a firm may intentionally violate the law by not complying with a regulatory standard, or by not paying the appropriate emission taxes. That is the type of violation that we analyze in this paper. In addition, we make an important distinction between the emission level and the taxes paid by a firm. Firms explicitly choose an emission—report combination and they may comply better with the environmental target than with the tax obligation or vice versa.

We first analyze the impact of the audit policy on an individual firm. The audit policy has a deterrence effect on both the firm's actual level of emission and its reported emission. When the firm does not fear any inspection, then it pollutes freely, and reports no pollution. When faced with a positive (but small) audit pressure, the firm decreases its emission level, and continues to report no pollution. It is only when the audit pressure is strong that the firm begins reporting more truthfully its actual level of emission. Hence, initially, *auditing has much a stronger deterrence effect on the emission decisions than on the reporting of them*. We claim this characteristic of the firm's behavior facing audits is good news. In particular, it has been extensively argued that pollution taxes should be considered for their environmental effects not for their revenue potential.

Second, we consider the optimal policy when the enforcement agency faces a population of firms with different characteristics. We develop the analysis under the assumption that the only objective of the enforcement agency is to minimize the level of emissions, as raising revenue is not an issue for the agency. We show that when firms differ in the effectiveness of the audit (some are more difficult to detect than others), then it is optimal to go first after the firms that are easy to audit. As the budget for audit increases, more firms will be monitored and the audit intensity on inspected firms increased as well. We also analyze the case where firms differ in their private gains from emissions. In this case, the optimal enforcement policy concentrates its efforts on those firms that value pollution less.

We show that, as it is expected, an increase in the budget (more monitoring) will induce pollution to fall. However, unless the enforcement agency's budget is very large, it will allocate its

[1] The compliance issue based on monitoring (or inspections) and fines is of general interest in many fields. For a general review of the compliance literature see Polinsky and Shavell [27]. For environmental problems see Cohen [3] and Heyes [15].

auditing effort among the firms in such a way that all of them will report not to have polluted. That is, an increase in the budget will not induce a rise in compliance with environmental taxes. We want to highlight the importance of this distinction: *There are always firms that do not comply with the environmental objective, and others that do comply*; but *all of them evade the environmental taxes*.

Our result seems to be at odds with a well-established result by Harford [11]. This author concludes that "the actual waste level of the firm does not directly depend upon the size of the fine or the probability of discovery of the violation." That is, increases in the budget would lead to more compliance with the taxes, but not to lower emission levels. This result was obtained from the analysis of the interior solution of the compliance decision of a single firm. Harford [11] also studied the corner solutions and argued that the interior solution is the sensible one. Our analysis points out that when the enforcement agency decides upon the distribution of the auditing intensity in a population of firms, it often allocates its limited resources in a manner that induces firms not to behave in the way described by Harford [11].[2]

There is an increasing literature on environmental regulations and more recently on the enforcement issue.[3] Harford [12], Helland [13], Innes [17], Kaplow and Shavell [18], Livernois and McKenna [20] and Malik [24], among others, have also considered self-reporting as an important element in enforcement policies. They show that self-reporting combined with an audit strategy increases compliance.

Swierzbinski [31] and Bontems and Bourgeon [1] study an informational aspect complementary to the one we address in this paper. They consider a model of environmental taxes where the regulator that designs the environmental policy may observe the emission levels (through a costly audit), but does not know the firms' abatement costs. They show that the threat of monitoring alters the usual result that states that firms over-estimate their abatement costs.

Finally, some empirical papers (see for example Dasgupta et al. [6], and Foulon et al. [8]) document the effect of monitoring and enforcement actions on the level of pollution emissions (for a review, see Cohen [4]. They provide evidence on the fact that both inspection and threat of an inspection are useful in reducing pollution emission.

The paper is organized as follows. In Section 2, we present the model and the firm's decision on both the emission level and the payment of taxes. Section 3 analyzes the optimal policy when there is a population of heterogenous firms that differ in their opportunities to evade; while in Section 4, we suppose that firms differ in the gains of pollution. In Section 5, we discuss on the generality of the results, and Section 6 concludes. Finally, an Appendix includes all the proofs.

2. The model and the firm's decision

In this section, we present the basic model and consider the decision of a single competitive firm. For the purpose of our model, we concentrate on the decision of the firm concerning its true

[2] The model used by Harford [11] has some differences with the model we present. However, the argument we give in this paragraph is robust to changes in our model which would make it similar to the one used by Harford [11].

[3] Cropper and Oates [5], Cohen [3], Heyes [15] and Sandmo [29] provide extensive reviews of the literature.

and reported level of emission. We use a generalization of the framework used, for example, by Sandmo [30].

The firm chooses the level of emission e, where $e \in [0, E]$. Hence, E is the emission level of the firm when pollution is free. The firm's benefits from emission e are represented by the function $\lambda g(e)$, with $\lambda > 0$, and $g(.)$ increasing and concave: $g'(e) > 0$ and $g''(e) < 0$ for all $e \in (0, E)$. Also, we assume for simplicity that $g'(0) = +\infty$ and $g'(E) = 0$, so that a small level of emission has a big marginal impact on the firm's profits, while the marginal profits at very high emission levels are very small. Parameter λ introduces a simple way to parametrize the gains of the firm (usually due to cost reduction) when polluting. A firm with higher λ is a firm whose private benefits from polluting are higher.

In order to control pollution, emissions are taxed at rate $t > 0$. We suppose that t is exogenously given; it is set by the government. It may be equal to the marginal social damages of emissions evaluated at the social optimum, taking into account the problem of enforcement.

Under environmental taxes, the *profits* of a firm with parameter λ that produces a level of emissions e and pays the taxes corresponding to e (i.e., there is perfect monitoring of emissions) are

$$\Pi(e; \lambda) = \lambda g(e) - te.$$

We denote by $e^*_\lambda = e^*(\lambda)$ the optimal level of emissions under perfect monitoring for a firm with parameter λ. The level e^*_λ is characterized by

$$\lambda g'(e^*_\lambda) = t.$$

The optimal level of emissions under perfect enforceability e^*_λ is increasing in λ and decreasing in t.

If the level of emissions is not perfectly monitorable cost free, then the auditing strategy of the enforcement agency and the reporting strategy of the firm (in addition to its emission strategy) are strategic decisions. We denote by α the probability that the enforcement agency will audit the emissions of the firm.[4] However, α is not necessarily the probability that an evader is caught, since an audit does not always allow the firm's true level of emissions to be uncovered. The probability that the true emission level of the firm is *identified* through an audit is $\rho \in [0, 1]$. Parameter ρ may be understood as the difficulty in detecting a violation or finding strong evidence that allows the sanctioning of the firm. Some pollutants persist in the environment longer than others; some can be more exactly assigned to the activity of a particular firm than others. The parameter ρ reflects these differences. A firm with a lower ρ is a firm that has more room for evading, since its emissions are harder to identify when audited.[5]

[4]For now, we suppose that the probability of being audited is independent of the report made by the firm. We think this is a sensible hypothesis. Moreover, in Section 5, we will show that restricting attention to this class of policy is without loss of generality in many scenarios.

[5]In our approach the type ρ of a firm is exogenous. Some authors have considered this characteristic to be the choice of the firm. Malik [23] considers a model where firms may engage ex ante in activities which serve to reduce the probability of being caught and fined. Heyes [14] studies a model where "uninspectability" is endogenously decided by firms.

The firm may choose a report z that does not coincide with the true emission level e. A firm never reports a higher emission level than the real one (since it involves paying higher taxes), so $z \leqslant e$. When it reports a level of emissions inferior to the real one—if it is audited and its true emission level is identified—, then in addition to paying the evaded taxes, a penalty is imposed to the firm. This penalty takes the form of the function $\theta(e-z)$, increasing and convex in the level of evasion: $\theta(0)=0, \theta'(x)>0$ and $\theta''(x)>0$ for $x>0$.[6] Sandmo [30] also assumes that the penalty on underdeclared emissions is progressive and argues that this is the most natural hypothesis. The main difference between our framework and Sandmo's is that we suppose that a firm is characterized by (λ, ρ), parametrizing differences in productivity and evasion possibilities.[7]

Therefore, the expected profits of a firm with parameters (λ, ρ) facing an audit probability α, when it chooses an emission level e and it reports z can be written as[8]

$$E\Pi(\lambda, \rho, \alpha; e, z) = \lambda g(e) - tz - \rho\alpha t[e-z] - \rho\alpha\theta(e-z) \quad \text{for } z \leqslant e. \tag{1}$$

The firm chooses the optimal levels e^0 and z^0 in order to maximize the expected profits (1). If the solution is interior, the first-order conditions are

$$\frac{\partial E\Pi}{\partial e} = \lambda g'(e) - \rho\alpha t - \rho\alpha\theta'(e-z) = 0, \tag{2}$$

$$\frac{\partial E\Pi}{\partial z} = -t + \rho\alpha t + \rho\alpha\theta'(e-z) = 0. \tag{3}$$

The next proposition establishes the optimal behavior of the firm:

Proposition 1. *For a given tax rate t, audit probability α, and penalty function $\theta(.)$, the optimal emission and report decisions (e^0, z^0) for the firm with parameters (λ, ρ) are*

(a) *If $\rho\alpha = 0$, then $e^0 = E$ and $z^0 = 0$.*
(b) *If $\rho\alpha \in (0, \frac{t}{\theta'(e^*_\lambda)+t})$, then $e^0 \in (e^*_\lambda, E)$ as defined by (4) and $z^0 = 0$, with*

$$\lambda g'(e^0) - \rho\alpha t - \rho\alpha\theta'(e^0) = 0. \tag{4}$$

(c) *If $\rho\alpha \in [\frac{t}{\theta'(e^*_\lambda)+t}, \frac{t}{\theta'(0)+t})$, then $e^0 = e^*_\lambda$ and $z^0 \in [0, e^*_\lambda)$ as defined by (5):*

$$[1-\rho\alpha]t = \rho\alpha\theta'(e^*_\lambda - z^0). \tag{5}$$

(d) *If $\rho\alpha \geqslant \frac{t}{\theta'(0)+t}$, then $e^0 = e^*_\lambda$ and $z^0 = e^*_\lambda$.*

[6]The penalty may be monetary or not. For example, in Canada, a list of firms that either do not comply with the existing regulation or whose environmental performance is of concern, is published annually. Both the community and the market act on it (see e.g., Lanoi et al. [19], for evidence on this aspect). Community pressure and other forms of informal sanction have been explored, for example, by Brooks and Sethi [2]. Penalties may also include the costs that responsible firms must pay, for cleaning-up the effects of the violations of the environmental regulation.

[7]The main purpose of Sandmo [30] is to explore the conditions under which the efficiency property of taxes continues to hold under imperfect compliance.

[8]To help the reading of equations, throughout the paper we only use parenthesis (.) for functions, as in $\theta(e-z)$ while we use brackets [.] for multiplications, as in $t[e-z]$, which means t times $e-z$.

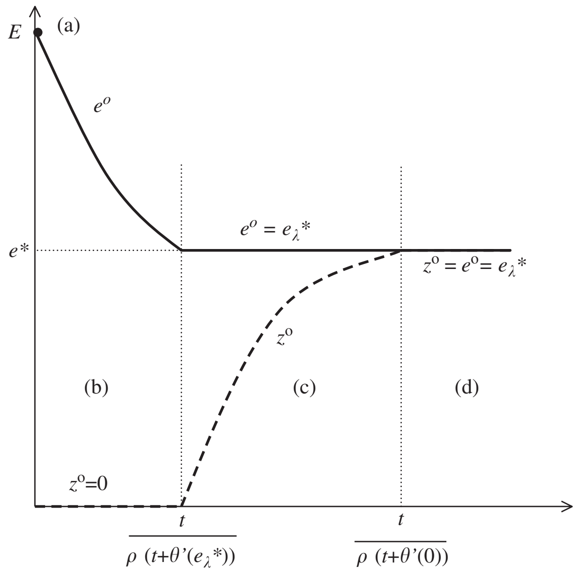

Fig. 1. Firm's best decision in terms of the emission level and the report.

The solution in terms of emissions and reports as a function of the audit probability α (for a firm that is caught with probability ρ) is illustrated in Fig. 1.[9] Since $\alpha \in [0,1]$, it may be the case that Region (d) in Fig. 1 does not exist. This happens when $t/\rho[\theta'(0)+t] \geqslant 1$, for example because $\theta'(0)=0$, or because ρ and/or $\theta'(0)$ are low enough. It can also be the case that both Regions (c) and (d) do not exist, what happens when $t/\rho[\theta'(e_\lambda^*)+t] \geqslant 1$, i.e., ρ is very low. Note, also, that the limits of the regions separating the interior and the corner solutions depend on both λ and $g(.)$ via e_λ^*. Finally, if the penalty function would be linear, then $\theta'(0)=\theta'(e_\lambda^*)$, and Region ($c$) would vanish.

If the firm is not subject to any audit ($\alpha = 0$), or it is impossible for the agency to prove that it has polluted ($\rho = 0$), then the firm does not fear an inspection. Hence, it pollutes freely while claiming to be a clean firm, that is, $e^0 = E$ and $z^0 = 0$. As the pressure on the firm increases (i.e., as we go from Region (a) to (b), with $\rho\alpha$ increasing), the firm decreases its level of emissions, while still reporting that it is clean. This is an important insight from the analysis of the model: *when auditing is not too frequent*, *its deterrence effect on emissions is much stronger than its effect on the report*. This result is independent of the objective function of the environmental agency, since it is derived from the analysis of the behavior of the firm. However, it is particularly good news for an agency that is (as we will assume from the next section on) mainly concerned about emissions, rather than with catching under-reporting firms.

When the audit pressure is strong, the firm chooses the "minimum" level of emission e_λ^* (the level that the firm would choose under perfect monitoring) and also makes a more honest report.

[9]Note that a similar figure can be drawn as a function of ρ for any level of α. It suffices to take into account that, for example, the cut-off $\alpha = t/\rho[\theta'(0)+t]$ will become $\rho = t/\alpha[\theta'(0)+t]$.

This corresponds to Region (*c*), where there is an interior solution for both emissions and report. This is the case that leads Harford [11] to reach the conclusion that emissions are not affected by the probability of auditing.[10] This is a very well-known result in the literature cited, for example, in Cohen [4]. Region (*c*) is also the region analyzed in Sandmo [30], where the optimal emission level is obtained, even if the taxes collected are not the ones corresponding to that emission level. Finally, if the perceived audit pressure $\rho\alpha$ is even stronger (Region (*d*)), the firm's decision is the same as under perfect monitoring, that is, $e^0 = z^0 = e^*_\lambda$.

Auditing firms aims at two apparently offsetting effects: (ex ante) deterrence and (ex post) detection. Some stylized facts[11] suggest that increased monitoring (a higher α) leads to higher detection coupled with higher deterrence. That is, the detection effect outweighs any deterrent effect. Our result is compatible with these stylized facts. Indeed, in the (possibly most relevant) Region (*b*), increasing α makes the firm more compliant, since it decreases its emission level. Moreover, the probability of the firm being caught increases, since it is still underreporting.

Finally, we state (without its easy proof) a corollary on the comparative statics of the optimal firm's emission and report with respect to the different parameters.

Corollary 1. (i) *The firm's optimal emissions* e^0 *are increasing in* λ *and decreasing in* t *when* $\rho\alpha > 0$. *Moreover,* e^0 *is non-increasing in* α *and* ρ.

(ii) *The firm's optimal report* z^0 *is non-decreasing in* α, ρ, *and* λ *and non-increasing in* t.

3. Firms differ in their possibilities to evade

In this section, we consider the optimal monitoring policy for the enforcement agency when it is in charge of auditing a population of firms that are heterogeneous with respect to their opportunities to evade. That is, we assume here that all firms obtain the same benefits from polluting, and we normalize $\lambda = 1$, but the probability of uncovering evasion varies across them, that is firms differ in their parameter ρ. The population of firms, parametrized by ρ, is distributed over the interval [0, 1], according to the density function $f(\rho)$, with $f(\rho) > 0$, for all $\rho \in [0, 1]$, whose cumulative function is $F(\rho)$. The enforcement agency has complete information about the type of each firm and can design an audit policy that discriminates among them.

We assume that *the only objective of the enforcement agency is to minimize total emissions*. That is, following e.g. Garvie and Keeler [10], we assume that the enforcement agency does not intend to raise money. Its objective is to achieve the highest level of compliance given its enforcement budget. (A lump-sum tax on firms makes possible to raise money without inducing any distortion.)[12] We denote by B the budget that the agency can devote to auditing and we normalize

[10] Harford [11] also analyzes Region (*b*) where reported wastes are equal to zero. He disregards this case as: "It would be irrational to set penalties so low that no pollution tax at all was collected."

[11] See for example Epple and Visscher [7] and Cohen [4].

[12] The first principle that according to the OECD [25] must be considered as overarching is: "Environmental quality should be the ultimate goal of enforcement agencies rather than the process of compliance control and enforcement per se. This implies changing the traditional focus on such performance indicators as the number of inspections, the number of detected violations versus violators brought to compliance, or the level of imposed and collected penalties." Another reason to exclude penalties from the objective function is that penalties are often not monetary, as discussed in the previous section.

the cost of one audit to one, so that B is the number of audits that the agency can carry out. Hence, the enforcement agency decides on the probability of auditing each type of firm, that is, it chooses $(\alpha(\rho))_{\rho\in[0,1]}$ in order to solve the following program:

$$\text{Min} \int_0^1 e(\rho)f(\rho)\,d\rho$$

$$\text{s.t.} \int_0^1 \alpha(\rho)f(\rho)\,d\rho \leqslant B$$

$$e(\rho) \in \text{argmax } E\Pi(\rho, \alpha(\rho); e, z),$$

where $E\Pi(\rho, \alpha; e, z)$ are the expected profits defined in (1) for $\lambda = 1$.

As we have seen by Proposition 1, the minimum emission level that the agency can achieve from any firm is e^* (we denote $e^* \equiv e_1^*$ since $\lambda = 1$ through out this section). Let us define

$$\hat{\alpha}(\rho) \equiv \frac{t}{\rho[\theta'(e^*) + t]}$$

as the minimum audit probability that induces a level of emissions e^* from a firm of type ρ (when $\alpha \leqslant \hat{\alpha}(\rho)$, the firm reports $z = 0$, the report is positive for $\alpha > \hat{\alpha}(\rho)$). Note that this probability level is "feasible" only when $\hat{\alpha}(\rho) \leqslant 1$, i.e., $\rho \geqslant \widehat{\rho}$, where:

$$\widehat{\rho} \equiv \frac{t}{\theta'(e^*) + t}.$$

A firm whose parameter ρ is lower than $\widehat{\rho}$, pollutes more than e^*, even if the audit probability is $\alpha = 1$, since the probability of being discovered when audited is low.

The minimum total pollution level that the agency can achieve (even with an unlimited budget) is

$$e_\rho^{\text{MIN}} \equiv \int_0^{\widehat{\rho}} e^{**}(\rho)f(\rho)\,d\rho + [1 - F(\widehat{\rho})]e^*,$$

where $e^{**}(\rho)$ is implicitly defined by

$$g'(e^{**}(\rho)) - \rho t - \rho\theta'(e^{**}(\rho)) = 0 \quad \text{for } \rho \in (0, \widehat{\rho}],$$

and $e^{**}(0) = E$. The first term in the expression for e_ρ^{MIN} measures the emissions of the firms that over pollute even when audited with probability one, choosing $e^{**}(\rho) > e^*$. The second term adds up the pollution of the firms that may be induced to choose the level of emission e^*.

What is the minimum budget that the enforcement agency needs in order to achieve e_ρ^{MIN}? The firms whose ρ belongs to $[0, \widehat{\rho}]$, should be audited with probability $\alpha(\rho) = 1$. On the other hand, those firms whose ρ belongs to $(\widehat{\rho}, 1]$, only need to be audited with probability $\hat{\alpha}(\rho)$. Therefore, the budget necessary to achieve e_ρ^{MIN} is

$$\overline{B}_\rho \equiv F(\widehat{\rho}) + \int_{\widehat{\rho}}^1 \hat{\alpha}(\rho)f(\rho)\,d\rho.$$

Given that the objective of the enforcement agency is to minimize emissions, the next proposition formally states an immediate consequence of the previous analysis.

Proposition 2. *When $B \geqslant \overline{B}_\rho$, the agency sets an audit policy that satisfies*: $\alpha(\rho) = 1$, *for* $\rho \in [0, \widehat{\rho}]$, *and* $\alpha(\rho) \in [\hat{\alpha}(\rho), 1]$, *for* $\rho \in (\widehat{\rho}, 1]$.

When the budget allocated to the enforcement agency is large enough, it will set a policy which achieves the minimum total pollution level possible, e_ρ^{MIN}. Increases in $\alpha(\rho)$, with respect to $\hat{\alpha}(\rho)$, for $\rho > \widehat{\rho}$, do not affect the firms' level of emission, they only increase the firms' report.

In the remainder of this section, we consider situations where $B < \overline{B}_\rho$, that is, where the agency does not have resources to achieve e_ρ^{MIN}. Next lemma highlights an important characteristic of the optimal audit policy in these situations.

Lemma 1. *When $B < \overline{B}_\rho$, the agency never sets an auditing probability higher than $\hat{\alpha}(\rho)$ for a type-ρ firm.*

Indeed, if it was the case that $\alpha(\rho) > \hat{\alpha}(\rho)$ for some ρ, decreasing this probability and increasing the audit pressure over those firms ρ' for which $\alpha(\rho') < \mathrm{Min}\{\hat{\alpha}(\rho'), 1\}$, would lead to a reduction in the total level of emissions. Note that this result depends on the assumption that the agency only cares about emission levels and it is not bothered by the fact that the report contains no relevant information in equilibrium.

The previous lemma leads to the following Proposition.

Proposition 3. *If $B < \overline{B}_\rho$, then the firm's report will be* $z(\rho) = 0$, *for all* $\rho \in [0, 1]$, *when the enforcement agency implements the optimal auditing policy*.

Proposition 3 states a result that is quite surprising at first sight: unless the agency's budget is very large (larger that $\overline{B}_\rho$), all the firms in the economy will be reporting that they do not pollute.[13] Understanding the result requires going back to Proposition 1. That proposition stated that increasing monitoring makes a firm first (Region (*b*)) decrease its emissions until a minimum level e_λ^*, while keeping the report $z^0 = 0$. When the monitoring is strong enough so that the firm decides e_λ^* (Region (*c*)), then increasing pressure only affects its reporting level, making it closer to the true emission. When the auditing agency only cares about emissions, the effect on the report is unimportant. Hence, it is not until all the firms are led to their minimum level of emissions (and this requires a budget of at least $\overline{B}_\rho$), that the agency induces them to report more truthfully. That is, under our assumptions, a probability of inspection that leads the firms to an interior solution in their compliance decision can never be part of optimizing behavior on the agency's side. This is the main reason behind the differences between our conclusions and those by Harford [11].

Before analyzing how the agency allocates the budget among the different types of firms, we comment on the allocation of resources to firms that have equal opportunities to evade. It is intuitive that the agency "should" apply the same policy to two identical firms. This is certainly the case if the firm's optimal emission is a (decreasing and) convex function of the probability of auditing. Indeed, under convexity, auditing one firm with a higher probability than other identical firms does not minimize the emission: monitoring both firms with average probability would decrease total pollution.[14]

[13] In other words, firms only report those emissions that do not really need monitoring to be identified.

[14] The property of convexity of the firm's emission with respect to the probability of auditing (which is not necessary to establish Proposition 3) seems reasonable and it helps to easily characterize the optimal policy addressed to a population of firms, since it allows the use of the marginal analysis. However, the main qualitative features of the policy

The next proposition characterizes the optimal auditing policy for budgets lower than $\overline{B}_\rho$. In particular, it shows that the auditing strategy will be biased toward targeting the easier-to-audit firms—the ones whose emissions are easier to identify. Corollary 2 complements the proposition by stating the firms' behavior facing the optimal auditing policy.

Proposition 4. *When $B<\overline{B}_\rho$ and the firm's emission is a decreasing and convex function of the probability of auditing, there exist $\rho_a(B)$ and $\rho_b(B)$, with $0<\rho_a(B)<\rho_b(B)\leqslant 1$, such that the optimal audit policy $\alpha(\rho)$ satisfies the following*:

(I) *If* $\rho\leqslant\rho_a(B)$, *then* $\alpha(\rho)=0$,
(II) *if* $\rho\in(\rho_a(B),\rho_b(B))$, *then* $\alpha(\rho)\in(0,\mathrm{Min}\{\hat{\alpha}(\rho),1\})$, *with* $\rho\alpha(\rho)$ *increasing in* ρ, *and*
(III) *if* $\rho\geqslant\rho_b(B)$, *then* $\alpha(\rho)=\mathrm{Min}\{\hat{\alpha}(\rho),1\}$.

Corollary 2. *When $B<\overline{B}_\rho$ and the firm's emission is a decreasing and convex function of the probability of auditing, the firms' optimal emission level $e^0(\rho)$ facing the optimal policy is the following*:

(I) *If* $\rho\leqslant\rho_a(B)$, *then* $e^0(\rho)=E$,
(II) *if* $\rho\in(\rho_a(B),\rho_b(B))$, *then* $e^0(\rho)$ (*defined by* (2) *for* $z=0$ *and* $\alpha=\alpha(\rho)$) *is decreasing in* ρ, *and*
(III) *if* $\rho\geqslant\rho_b(B)$, *then* $e^0(\rho)=e^*$.

Fig. 2 illustrates Proposition 4 and Corollary 2. When the enforcement agency does not have the budget necessary to achieve the minimum pollution possible e_ρ^{MIN}, then it has incentives to discriminate among firms. The agency first targets those firms whose non-compliance is easier to verify, that is, firms with higher ρ. For the firms with the highest ρs, that is, in Region (III) (which only exists when the budget B is high enough), the agency exerts the maximum auditing pressure, leading those firms to their lowest level of emissions e^*. In this region, the audit pressure $\alpha(\rho)$ decreases with ρ, because the easier it is to identify pollution, the lower the audit probability necessary to induce e^*.

The agency also audits with some probability those firms with intermediate values of the parameter ρ, Region (II), the total perceived pressure $\rho\alpha(\rho)$ increasing in ρ. Hence, easier-to-catch firms produce lower levels of emissions. Finally, the agency decides not to audit those firms whose pollution is very difficult to detect. All these firms will pollute as much as they wish, that is $e^0(\rho)=E$.

(*footnote continued*)
should remain if there are regions of parameters where the firm's emission is a concave function of the probability of auditing (although the policy will not be a continuous function of the firms' characteristics).

Formally, emissions are a convex function of the auditing probability if the function $h(x)$ defined below is increasing:

$$h(x)\equiv\frac{t+\theta'(e(x))}{g''(e(x))-x\theta''(e(x))}.$$

Note that $g'''(.)>\mathrm{Max}\{0,\theta'''(.)\}$ is a sufficient (although far from necessary) condition for $h(x)$ to be increasing. For example, an easy case where it holds is $g(e)=e^r$ and $\theta(e)=e^s$, with $r\in(0,1)$ and $s\in[1,2]$. Also, note that $h(x)<0$ since $\theta'(.)>0$, $\theta''(.)>0$ and $g''(.)\leqslant 0$.

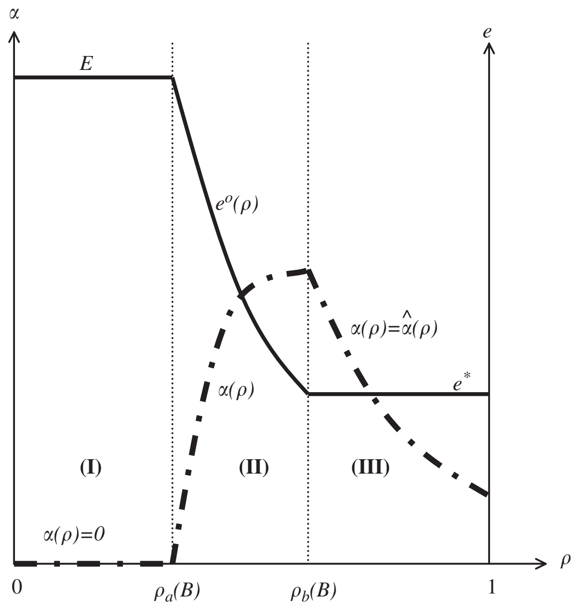

Fig. 2. Optimal audit policy (α) and induced level of emissions (e) as a function of ρ.

It is worthwhile pointing out that a result similar to Proposition 4 holds if the objective function of the enforcement agency is to minimize the budget necessary to achieve a given level of total emissions. For any level of total emissions $\bar{e}$ that the agency would wish to implement, there exist two cut-off values, $\rho_a(\bar{e})$ and $\rho_b(\bar{e})$, that define three Regions (I), (II) and (III) where the optimal audit policy follows the same pattern as in Proposition 4.

Note that according to Proposition 4, at the optimum, the most polluting firms will never be inspected. Of course this result relies on our assumption that only the global amount of emissions matters. If pollution has strong local effects then the Agency will have a more complex objective function than the one we assumed. The Agency will also have to act according to the requirements of politicians or public opinion.[15] Note also that the equalization of marginal profits (or marginal abatement costs), which is a property of pigovian taxation under full compliance, does not hold in our environment. In fact, this equalization only holds for those firms whose parameter ρ is larger or equal than $\rho_b(B)$, while for the other firms the marginal profits (or marginal abatement costs) are smaller given the optimal monitoring policy.

Finally we relate Proposition 4 to results obtained in Macho-Stadler and Pérez-Castrillo [21], where we explored the optimal audit policy for an income tax agency when confronted by a population such that the effectiveness of the audit policy differs from one taxpayer to the next. In

[15]See, for example, Ruiz [28] for a recent discussion of the impact of the emissions produced by the " Prestige", that sunk off Galicia (Spain) on 19 November 2002. The accidental spill from this oil tanker was spectacular and prompted the reaction of the population and the politicians. However, Ruiz [28] claims that the relative toxicity of the oil released by the "Prestige" was in fact equivalent to 10% the amount of an ordinary period. This every-day pollution receives much less attention and enforcement resources since it is much less spectacular.

the income tax evasion model also, the easy-to-control agents perceive a higher pressure than difficult-to-control agents. However, there are important differences between the two analyses. Our previous paper dealt with an adverse selection problem with respect to the (exogenous) income of the taxpayers, while the present paper analyzes a moral hazard problem with respect to the emissions. A problem that, in addition, generates an (endogenous) adverse selection problem concerning a second decision: the report.[16] Moreover, while the income tax agency was assumed to maximize tax receipts, in the present paper the environmental agency is assumed to care only about total emissions.

We close this section by showing how the audit strategy changes with the budget.

Proposition 5. *The cut-off levels* $\rho_a(B)$ *and* $\rho_b(B)$, *identified in Proposition* 4, *satisfy the following property*: $\rho_a(B)$ *is decreasing and* $\rho_b(B)$ *is non-increasing in B. Moreover, the optimal audit pressure* $\alpha(\rho)$ *is increasing in B, for all* $\rho \in (\rho_a(B), \rho_b(B))$.

Proposition 5 shows that when the budget for audit increases, more firms (from the population of firms easy to monitor) will comply with the environmental standards, and some other firms that are harder to monitor will be subject to audit. Moreover, except for the firms whose audit pressure is either zero, or high enough, the audit intensity will also increase with the budget (and hence the emission level will decrease).

4. Firms differ in their gains of pollution

In this section, we characterize the optimal monitoring policy when firms differ in their gains from emissions. Given the similarities between this analysis and the one developed in the previous section, we concentrate here on the main result and intuitions.

We consider that the enforcement agency faces a population of firms parametrized by λ (the parameter that measures the gains of the firms), distributed over the interval $[\underline{\lambda}, \overline{\lambda}]$, $0 < \underline{\lambda} < \overline{\lambda}$ according to the density function $\varphi(\lambda)$, with $\varphi(\lambda) > 0$ for all $\lambda \in [\underline{\lambda}, \overline{\lambda}]$. We consider for simplicity that the monitoring technology is perfect, i.e., $\rho = 1$, the qualitative results are not altered if one analyzes the situations with $\rho < 1$.

In Section 2, we denoted by e^*_λ the emission level decided by a firm with parameter λ which is subject to perfect monitoring. This is the minimum emission level that the enforcement agency can achieve through its monitoring strategy. Moreover, when $\rho = 1$, the firm will indeed pollute e^*_λ if and only if the probability of auditing is higher or equal to $t/[\theta'(e^*_\lambda) + t]$, which is always smaller than 1. Hence, the minimum industry pollution level, with no constraint on the budget, is defined by

$$e_\lambda^{\mathrm{MIN}} \equiv \int_{\underline{\lambda}}^{\overline{\lambda}} e^*_\lambda \varphi(\lambda)\, d\lambda.$$

[16] Because of the different nature of the asymmetric information highlighted before, the methods of proof of the two papers are drastically different.

The budget necessary to achieve $e_{\lambda}^{\mathrm{MIN}}$ is

$$\overline{B}_{\lambda} \equiv \int_{\underline{\lambda}}^{\overline{\lambda}} \frac{t}{[\theta'(e_{\lambda}^{*}) + t]} \varphi(\lambda)\, d\lambda.$$

The next proposition characterizes the optimal monitoring policy:

Proposition 6. (i) *When $B \geqslant \overline{B}_{\lambda}$, the agency chooses an audit policy that implies $\alpha(\lambda) \geqslant \frac{t}{[\theta'(e_{\lambda}^{*})+t]}$, for all $\lambda \in [\underline{\lambda}, \overline{\lambda}]$. Firms' emission levels are $e^{0}(\lambda) = e_{\lambda}^{*}$.*

(ii) *When $B < \overline{B}_{\lambda}$, there exists $\lambda(B)$, with $\underline{\lambda} < \lambda(B) \leqslant \overline{\lambda}$, such that*

(ii(I)) *For firms with $\lambda \geqslant \lambda(B)$, then $\alpha(\lambda) = 0$. A firm's emission level is $e^{0}(\lambda) = E$.*

(ii(II)) *For firms with $\lambda < \lambda(B)$, then $\alpha(\lambda) > 0$. A firm's emission level $e^{0}(\lambda)$ is increasing in λ.*

Proposition 6 shows that when the firms differ in the gains from emissions, the agency biases its strategy against those firms that value pollution less. Having less incentives to pollute, the firms with less gains from polluting will be more deterred by the auditing, hence the monitoring will have a stronger effect on those firms. On the other hand, the agency prefers not to devote resources to firms that place strong value on emissions (i.e., firms with very high λ). For those firms, polluting is so valuable that the marginal deterrence effect of the audit is small.

We would like to emphasize here that the optimality of the proposed audit policy relies very much on the fact that the agency can *announce and commit* on the audit strategy before the firms take any decision. Hence, the main objective of the policy is to deter firms from polluting. If the auditing is decided after the firms have polluted and reported, the equilibrium emerging in this game is different. In a quite different environment, Persico [26] analyzes the equilibrium allocation of policing between two groups of citizens with different legal (and similar criminal) earning opportunities. He assumes that the police cannot commit ex ante on its strategy of monitoring, and concludes that the police will devote more resources to the group with more modest earning opportunities, that is, to the group that (in relative terms) finds the criminal activity more profitable. The difference in the commitment capabilities in our model and Persico's one explains why the outcomes are different. In fact, Persico [26] also argues that, under some conditions, (ex ante) constraining the police to behave in a more fair way with respect to both groups increases the (ex ante) effectiveness of the police work.

For completeness, let us mention that a similar analysis can be developed for a population of firms differing along the two dimensions, λ and ρ. We denote

$$\hat{\alpha}(\rho, \lambda) \equiv \frac{t}{\rho[\theta'(e_{\lambda}^{*}) + t]}$$

the minimum audit probability that induces e_{λ}^{*} from a firm with parameters (λ, ρ). This probability is only feasible when $\hat{\alpha}(\rho, \lambda) \leqslant 1$, i.e., $\rho \geqslant \widehat{\rho}_{\lambda} \equiv t/[\theta'(e_{\lambda}^{*}) + t]$. If we suppose, for simplicity, that the distribution of firms along these two dimensions is independent, the minimum total pollution level that the agency can achieve is

$$e_{\rho,\lambda}^{\mathrm{MIN}} \equiv \int_{\underline{\lambda}}^{\overline{\lambda}} \left[\int_{0}^{\widehat{\rho}_{\lambda}} e_{\lambda}^{**}(\rho) f(\rho)\, d\rho + [1 - F(\widehat{\rho}(\lambda))] e_{\lambda}^{*} \right] \varphi(\lambda)\, d\lambda,$$

where $e_\lambda^{**}(\rho)$ is implicitly defined by

$$\lambda g'(e_\lambda^{**}(\rho)) - \rho t - \rho\theta'(e_\lambda^{**}(\rho)) = 0.$$

The budget necessary to achieve e_ρ^{MIN} is

$$\overline{B}_{\rho,\lambda} \equiv \int_{\underline{\lambda}}^{\overline{\lambda}} \left[F(\widehat{\rho}_\lambda) + \int_{\widehat{\rho}_\lambda}^{1} \hat{\alpha}(\rho,\lambda) f(\rho)\, d\rho \right] \varphi(\lambda)\, d\lambda.$$

When the budget is not too high, $B < \overline{B}_{\rho,\lambda}$, then the agency first targets those firms that value pollution less and whose non-compliance is easier to verify, that is, firms with low λ and high ρ. The firms with very low ρ and/or very high λ will not be subject to audit in the optimal audit policy.

5. A general audit policy

We have considered a model where the probability α that a firm is audited is independent of the report. This is also the framework adopted in most of the literature, including the related papers by Harford [11,12], Kaplow and Shavell [18], Innes [17], and Sandmo [30]. We made this reasonable hypothesis because it simplifies the analysis. In general, however, the audit probability can depend not only on the firm's characteristics (ρ and λ), but also on the firm's report z. We prove, and briefly discuss here, a result that shows that restricting attention to policies which are independent of the report z does not reduce the effectiveness of control in many interesting cases.

As in the previous section, we denote by $\hat{\alpha}(\rho,\lambda)$ the minimum audit probability that induces e_λ^* from a firm with parameters (λ,ρ).

Proposition 7. *Consider an optimal general auditing function $\alpha(z)$, and let z^0 be the optimal firm's report given $\alpha(z)$. Suppose that $\alpha(z^0) \leqslant \mathrm{Min}\{\hat{\alpha}(\rho,\lambda),1\}$. Then, the audit policy where the agency audits any report with probability $\alpha(z^0)$ is equivalent to the policy $\alpha(z)$. Therefore, the best policy that the agency can implement with a budget $\alpha \leqslant \mathrm{Min}\{\hat{\alpha}(\rho,\lambda),1\}$ leads the firm to report $z=0$ and it is equivalent to the policy where the agency audits any report with the same probability α.*

Proposition 7 shows that, if the agency does not want to achieve emission levels below e_λ^*, then it can restrict attention to policies where the audit probability does not depend on the report. The agency cannot achieve a better result through more general audit policies. Note that, if the tax rate t is optimally designed, e_λ^* is the optimal emission level from a social point of view under perfect monitoring. On the other hand, if the tax rate is not optimal, and the agency can credibly use sophisticated auditing schemes with audit probabilities depending on the firm's report, then it may have incentives to propose different audit functions than the ones we use.

We remark that although the optimal audit policy $\alpha(z)$ does depend on the firm's characteristics (hence, it is not robust to the introduction of asymmetric information on the gains from pollution λ and on the difficulty of detecting a violation ρ), the result stated in Proposition 7 does not depend on such information. Indeed, in the proof of Proposition 7, we show that the optimal audit policy $\alpha(z)$ leads the firm to report $z^0 = 0$. For any firm with characteristics (ρ,λ) satisfying $\alpha(0) \leqslant \mathrm{Min}\{\hat{\alpha}(\rho,\lambda),1\}$, the policy that sets an audit probability $\alpha(0)$ for any report is equivalent to the policy $\alpha(z)$.

Finally, Proposition 7 (as well as the previous results in this paper) states an important property of the optimal audit policy from the agency's point of view, without taking into account any potential internal conflict inside the agency. However, it is conceivable that detecting lying polluters will be good for the inspectors' career perspectives. In this sense, it is worth noticing that the audit policy proposed in Proposition 7 is precisely the one that maximizes the level of collected penalties for the given budget (although it does not maximize the total amount of money raised by the agency).[17] Hence, this proposition suggests that if there was a problem of incentives between the government and the agency, rewarding the agency in terms of the amount of penalties collected is more efficient than rewarding it as a function of the taxes collected.

6. Conclusion

In this paper, we have aimed at better understanding the role of environmental monitoring in determining firms' emission decisions and firms' tax compliance behavior. Our results predict that, when facing a population of heterogenous firms, the enforcement agency will focus on the ‘‘easier'' enforcement targets: easier-to-detect firms and those firms that value pollution less. Hence, the results help to explain why some firms and/or some industries are more monitored than others when the environmental agency targets the total level of emissions. This conclusion is in accordance with stylized facts (see Cohen [3]). This ‘‘discriminatory'' audit strategy allows more efficient results to be obtained in terms of total emissions.

We have also shown that the optimal auditing policy may very well lead to a reasonable level of emissions, coupled with a very high level of environmental tax evasion. In fact, we should expect that firms will be reporting very low emissions. Since the ultimate goal of enforcement agencies is environmental quality, and not tax revenue, one should not be surprised to see that their best policy leads firms to comply with the environmental objective but to evade environmental taxes.

Our results suggest that the insights obtained when focusing in the analysis of the optimal audit of a single firm may be drastically different from the conclusions derived from the analysis of the population of firms. In contrast to Harford [11] and Sandmo [30], we show that (except if the budget is very large) it is never optimal to audit a firm to the extend where it reports a positive level of emissions. It is better to allocate the resources among the firms so that they evade taxes but further comply with the environmental objective.

In our model, we abstract from many interesting elements of the environmental enforcement problem that are complementary to our analysis. Let us briefly comment on some of them.

We concentrate on the enforcement aspect of the environmental problem, and we do not address the question of how environmental taxes and the enforcement agency's budget are decided. These tools may be the choice of the central authority, who may consider social welfare or political interest in the decision-making process. In our model, the enforcement agency maximizes compliance with the environmental target. The general environmental policy will be decided at an earlier stage.

We assume that sanctions are costless to the enforcement agency. In fact, one may argue that prosecuting and enforcing the payment of fines may be costly for the regulator. This aspect may

[17]The proof of this result is included in the proof of Proposition 7.

reduce the agency interest in enforcing the environmental target, but will not change the nature of our results. We also assume that all the participants are risk-neutral. Risk aversion, or wealth constraints, may be important in some cases. In particular, bankruptcy and insolvency are problems that should be taken into account. However, we have argued that penalties are often non monetary (incarceration, reputation, firm's image, etc.).

In our model, the probability of inspection is endogenous (and contingent to the firm's pollution report). However, we do not consider the possibility that the probability of being inspected increases with the level of emissions of a firm. It may be the case that the firm's emission level may attract the attention of the environment agency via some kind of signal so that the probability of being audited increases with the level of evasion. Prior information in environmental enforcement has been considered by several authors. Harford [11,12] assumes that the exogenous probability of auditing is an increasing function of the wastes emissions. Heyes [16] presents a model where the firm is subject to a " light" inspection that may trigger a real audit. Franckx [9] proposes the use of ambient inspections before deciding on the auditing of a particular firm (see also Macho-Stadler and Pérez-Castrillo [22], for an analysis of the use of prior information in tax evasion models).

Finally, we have adopted the principal-agent approach. Hence, we have assumed perfect commitment (that often is justified based on the reputation concern of the enforcement agency). This is the most common approach. In fact, this is the most optimistic one, since it is the best scenario for enforcement issues. Given that, according to our results, firms report no emission but are inspected with small probability, press and opposition politicians may use such a situation in order to attack an Agency that does not react in a tougher way. Hence, commitment may be difficult to sustain. Some authors have recently considered the enforcement problem (monitoring and emission strategies) as the sequential equilibrium outcome of a game, where the enforcement agency has no-commitment capacity (see for instance Franckx [9]).

Acknowledgments

We thank Pierre Courtois, Laurent Franckx, Nicolas Porteiro, and two referees for very helpful comments. This research was partially conducted while we were visiting CES at Munich. We gratefully acknowledge the financial support from the Ministerio de Ciencia y Tecnología (BEC2003-01132) and the Generalitat de Catalunya (Barcelona Economics program of CREA and 2001SGR-00162).

Appendix

Proof of Proposition 1. First, we check the second-order conditions:

$$\frac{\partial^2 E\Pi}{\partial e^2} = \lambda g''(e) - \rho\alpha\theta''(e-z) < 0,$$

$$\frac{\partial^2 E\Pi}{\partial z^2} = -\rho\alpha\theta''(e-z) < 0$$

and

$$\frac{\partial^2 E\Pi}{\partial e^2}\frac{\partial^2 E\Pi}{\partial z^2} - \left[\frac{\partial^2 E\Pi}{\partial e \partial z}\right]^2 = -\rho\alpha\lambda g''(e)\theta''(e-z) \geqslant 0.$$

The emission level e^0 maximizing (1) is always strictly positive. Also, it is strictly lower than the maximum level E if and only if $\rho\alpha > 0$. If $\rho\alpha = 0$, that is, we are in Region (*a*), then it is easy to check that the firm chooses $e^0 = E$ and $z^0 = 0$. For the rest of the proof, we suppose $\rho\alpha > 0$, hence $e \in (0, E)$.

The report z is interior if and only if $\rho\alpha\theta'(0) < [1-\rho\alpha]t < \rho\alpha\theta'(e^0)$. When $\rho\alpha\theta'(0) \geqslant [1-\rho\alpha]t$, that is, we are in Region (*d*), the corner solution is $z^0 = e^0$ (the firm reports honestly) and then the firm's maximum is $e^0 = e^*_\lambda$. When $[1-\rho\alpha]t \geqslant \rho\alpha\theta'(e^0)$, the firm reports $z^0 = 0$. It chooses e^0 satisfying (2) for $z^0 = 0$, i.e. e^0 satisfies (4). Such a pair, e^0 satisfying (4) and $z^0 = 0$, is indeed a candidate solution if and only if $[1-\rho\alpha]t \geqslant \rho\alpha\theta'(e^0)$ for the proposed e^0. Given (4), the previous inequality is equivalent to $t \geqslant \lambda g'(e^0)$, i.e., $e^0 \geqslant e^*_\lambda$. This corresponds to the candidate (that will be optimum) in Region (*b*) (the case $e^0 = e^*_\lambda$ also appears when we analyze interior solutions). When both the emission level and the report are interior (Region (*c*)), adding Eqs. (2) and (3) we obtain $\lambda g'(e) = t$, i.e., $e^0 = e^*_\lambda$. The optimal report in this region z^0 is defined by (3) for $e = e^*_\lambda$, that is, it is given by Eq. (5). □

Proof of Proposition 4. In order to prove Proposition 4 we first state that $\rho_1\alpha_1 \leqslant \rho_2\alpha_2$ for $\rho_1 < \rho_2$. We state and prove two lemmas.

Lemma A.1. *The enforcement agency audits two identical firms with the same probability.*

Proof Lemma A.1. Denote by α_1 and α_2 the probabilities of auditing identical firms 1 and 2 with equal parameters ρ. First, when $\mathrm{Min}\{\alpha_1, \alpha_2\} \geqslant \mathrm{Min}\{\hat{\alpha}(\rho), 1\}$, the enforcement agency achieves the best-possible outcome, since either $e_1 = e_2 = e^*$ or $e_1 = e_2 = e^{**}(\rho)$. No reallocation of resources among those firms is possible. But since $B < \overline{B}_\rho$, no probability can be higher than $\mathrm{Min}\{\hat{\alpha}(\rho), 1\}$, hence $\alpha_1 = \alpha_2 = \mathrm{Min}\{\hat{\alpha}(\rho), 1\}$. Second, in Region (*b*) of Proposition 1, where $\alpha < \mathrm{Min}\{\hat{\alpha}(\rho), 1\}$ and $e^0 > e^*$, e^0 is a convex function of α. Auditing one firm with a probability $\alpha_1$lower than α_2 does not minimize the emission: a monitoring probability equal to $(\alpha_1 + \alpha_2)/2$ applied to both firms would result in lower total emissions than $e_1 + e_2$. □

Lemma A.2. *The emission levels e_1 and e_2 of two firms with parameters ρ_1 and ρ_2 satisfy $e_1 \geqslant e_2$ if and only if $\rho_1\alpha_1 \leqslant \rho_2\alpha_2$. Also, $e_1 > e_2$ if and only if $\rho_1\alpha_1 < \rho_2\alpha_2$ and $\alpha_1 < \hat{\alpha}(\rho_1)$.*

Proof of Lemma A.2. Suppose $\rho_1\alpha_1 \leqslant \rho_2\alpha_2$. First, if $\alpha_1 \geqslant \hat{\alpha}(\rho_1)$, i.e., $\rho_1\alpha_1 \geqslant \widehat{\rho}$, then also $\rho_2\alpha_2 \geqslant \widehat{\rho}$, i.e., $\alpha_2 \geqslant \hat{\alpha}(\rho_2)$. Therefore, $e_1 = e_2 = e^*$. Second, if $\rho_1\alpha_1 < \widehat{\rho} \leqslant \rho_2\alpha_2$, then $e_2 = e^* < e_1$. Third, let us assume that $\rho_2\alpha_2 < \widehat{\rho}$. Take Eq. (2) for $z^0 = 0$ (since in this region $z_1 = z_2 = 0$): $\lambda g'(e) = \rho\alpha[t + \theta'(e)]$. This equation defines a negative relationship between e and $\rho\alpha$ since $g(.)$ is concave and $\theta(.)$ is convex. Finally, it is easy to check that the conditions are not only necessary, but also sufficient.

Let us consider now two firms, 1 and 2, such that $\rho_1 < \rho_2$. Denoting $\alpha_i = \alpha(\rho_i)$ and $e_i = e^0(\rho_i)$, for $i = 1, 2$, we prove that $\rho_1\alpha_1 \leqslant \rho_2\alpha_2$. In the following argument, we assume $\alpha_1 \in (0, \mathrm{Min}\{\hat{\alpha}(\rho_1), 1\}]$, $\alpha_2 < \mathrm{Min}\{\hat{\alpha}(\rho_2), 1\}$. Suppose for a contradiction that $\rho_1\alpha_1 > \rho_2\alpha_2$ and consider a decrease in α_1 by $\delta > 0$ (δ small enough) that induces a saving of $f(\rho_1)\delta$ in auditing costs, and an increase in α_2, financed through this saving. This implies an increase in α_2 equal to $\delta f(\rho_1)/f(\rho_2)$.

The change in the total level of emissions after this reallocation of the budget is (notice that the marginal effects take place in Region (*b*) of Proposition 1):

$$-f(\rho_1)\frac{\partial e_1}{\partial \alpha_1}\delta + f(\rho_2)\frac{\partial e_2}{\partial \alpha_2}\frac{f(\rho_1)}{f(\rho_2)}\delta = f(\rho_1)\delta[-\rho_1 h(\rho_1\alpha_1) + \rho_2 h(\rho_2\alpha_2)], \tag{6}$$

where $h(x) \equiv \frac{t+\theta'(e(x))}{g''(e(x))-x\theta''(e(x))}$. Since $h(x)$ is increasing (see footnote 13), $h(\rho_1\alpha_1) > h(\rho_2\alpha_2)$, both expressions being negative. Therefore, $[-\rho_1 h(\rho_1\alpha_1) + \rho_2 h(\rho_2\alpha_2)] < 0$ and total emissions decrease after the reallocation of the budget previously proposed. Therefore, setting α_1 and α_2 such that $\rho_1\alpha_1 > \rho_2\alpha_2$ cannot be optimal.

Now, we can distinguish three regions:

(I) By the previous argument $\alpha_1 = 0$ when $\alpha_2 = 0$. Hence, there exists $\rho_a(B) \geqslant 0$ such that $\alpha(\rho) = 0$ for all $\rho \leqslant \rho_a(B)$.
(II) It is immediate after the argument developed before.
(III) The previous argument shows that $\alpha_2 = \text{Min}\{\hat{\alpha}(\rho_2), 1\}$ whenever $\alpha_1 = \text{Min}\{\hat{\alpha}(\rho_1), 1\}$. Hence, there exists $\rho_b(B) \in (\rho_a(B), 1]$ such that $\alpha(\rho) = \text{Min}\{\hat{\alpha}(\rho), 1\}$ for all $\rho \geqslant \rho_b(B)$.

To show that $\rho_a(B) > 0$ suppose, for a contradiction, that $\alpha(\rho) > 0$ for all $\rho > 0$ and take a fixed $\overline{\rho} \in (0, \rho_b(B))$. Given the argument developed before Eq. (6), it must be the case that $\rho h(\rho\alpha(\rho)) = \overline{\rho} h(\overline{\rho}\alpha(\overline{\rho}))$, for all $\rho \in (0, \rho_b(B))$. Since $\overline{\rho} h(\overline{\rho}\alpha(\overline{\rho}))$ is constant, it must happen that $h(\rho\alpha(\rho))$ tends to $+\infty$ when ρ tends to zero. However, this is not possible, given the expression for $h(.)$, since $\theta'(e)$ is bounded when $e \in [e^*, E]$ and $g''(e) < 0$ for all $e \in [e^*, E]$.

We also show that when B is large enough (but still smaller than $\overline{B}_\rho$), there exists a value $\rho_b(B) < 1$ that does separate regions (II) and (III). To prove this statement by contradiction, suppose $\rho_b(B) = 1$, that is, $\alpha(\rho) < \text{Min}\{\hat{\alpha}(\rho), 1\}$ for all $\rho \in (\rho_a, 1]$. Then, in the interval $(\rho_a, 1]$, the parameters ρ are in region (II), hence $\rho h(\rho\alpha(\rho)) = h(\alpha(1))$. Since $\alpha(1) \leqslant \hat{\alpha}(1)$, then $\rho h(\rho\alpha(\rho)) \leqslant h(\hat{\alpha}(1))$. Denote $\widetilde{\alpha}(\rho)$ the value satisfying $\rho h(\rho\alpha(\rho)) = h(\widetilde{\alpha}(1))$. Given that $h(.)$ is an increasing function and $\rho < 1$, it is the case that $\widetilde{\alpha}(\rho) < \hat{\alpha}(\rho)$ for all $\rho \in (\widehat{\rho}, 1)$. The cost of a policy where the firms with parameter $\rho \in (\widehat{\rho}, 1)$ are monitored with probability $\widetilde{\alpha}(\rho)$ is strictly lower than $\overline{B}_\rho$ (since at this budget the probability of auditing such firms is $\hat{\alpha}(\rho)$); let us denote the maximum cost by $\overline{b}_\rho < \overline{B}_\rho$. Then, the cost of any monitoring policy involving $\rho_b(B) = 1$ is lower than $\overline{b}_\rho < \overline{B}_\rho$. Therefore, $\rho_b(B) < 1$ for any budget $B \in (\overline{b}_\rho, \overline{B}_\rho)$. □

Proof of Proposition 5. A higher B must imply the increase in the audit probability of at least one type-ρ firm and we shall show that no firm may now be under a lower audit pressure. We first notice, by Proposition 4, that the optimal audit policy for a particular B is easily characterized once we know $\alpha(\rho^0)$ for any ρ^0 in Region (II), where $0 < \alpha(\rho^0) < \text{Min}\{\hat{\alpha}(\rho^0), 1\}$. Indeed, for any ρ in Region (II), it must be the case that the change that leads (in the proof of Proposition 4) to Eq. (6) is not profitable, i.e., $\rho h(\rho\alpha(\rho)) = \rho^0 h(\rho^0\alpha(\rho^0))$. This equality implicitly defines $\alpha(\rho)$ for any ρ in Region (II), while $\alpha(\rho) = 0$ in Region (I) and $\alpha(\rho) = \text{Min}\{\hat{\alpha}(\rho), 1\}$ in Region (III). Hence, let $\beta(\rho)$ be implicitly defined by

$$h(\rho\beta(\rho)) = \frac{\rho^0 h(\rho^0\alpha(\rho^0))}{\rho} \quad \text{if } h(0) < \frac{\rho^0 h(\rho^0\alpha(\rho^0))}{\rho}$$

and

$$\beta(\rho) = 0 \quad \text{if } h(0) \geqslant \frac{\rho^0 h(\rho^0 \alpha(\rho^0))}{\rho}.$$

(The value $\beta(\rho)$ is well defined because $h(.)$ is an increasing function.) Take $\alpha(\rho) = \mathrm{Min}\{\beta(\rho), \hat{\alpha}(\rho), 1\}$. Then, it is easy to check in a manner similar to the previous argument that the optimal policy is $\alpha(\rho)$. The function $\beta(\rho)$ is weakly increasing in $\alpha(\rho^0)$ (it is strictly increasing if $\beta(\rho) > 0$). Hence, $\alpha(\rho)$ is weakly increasing in $\alpha(\rho^0)$. In other words, when a particular $\rho = \rho^0$ is audited more regularly, no other ρ can be audited with less probability. Consequently, the level of emissions $e(\rho)$ is also a weakly increasing function in $e(\alpha(\rho^0))$. A higher B must imply the increase in the audit probability of at least one type-ρ firm, and by the previous argument no firm may be now under a lower audit pressure. Hence, a higher B leads to a lower ρ_a and ρ_b. Moreover, the audit intensity increases for all firms that are not at a corner solution. □

Proof of Proposition 6. Part (i) is trivial since $B \geqslant \overline{B}_\lambda$ allows the setting up of a policy with $\alpha(\lambda) \geqslant t/[\theta'(e^*_\lambda) + t]$ for all $\lambda \in [\underline{\lambda}, \overline{\lambda}]$, which leads to the best-possible outcome for the agency.

For part (ii), we first claim that, for the same reasons as in Proposition 3, when $B < \overline{B}_\lambda$ the auditing policy is such that $\alpha(\lambda) \leqslant t/[\theta'(e^*_\lambda) + t]$ and it induces all firms to report zero. That is, the policy lies in regions (*a*) or (*b*) of Proposition 1. Second, consider two firms, with $\lambda_1 > \lambda_2$, $\alpha(\lambda_1) > 0$, and $\alpha(\lambda_2) < t/[\theta'(e^*_{\lambda_2}) + t]$. We are going to prove that $\frac{\alpha(\lambda_1)}{\lambda_1} < \frac{\alpha(\lambda_2)}{\lambda_2}$. For this, we analyze the consequences of a decrease in $\alpha(\lambda_1)$ by $\delta > 0$ (δ small enough) that induces a saving of $\varphi(\lambda_1)\delta$ in auditing costs, and an increase in $\alpha(\lambda_2)$ financed with this amount. This implies an increase in $\alpha(\lambda_2)$ equal to $\delta\varphi(\lambda_1)/\varphi(\lambda_2)$. The change in the total level of emissions after this reallocation of the budget is (note that the relevant marginal effects happen in Region (*b*) of Proposition 1):

$$-\varphi(\lambda_1)\frac{\partial e^0(\lambda_1)}{\partial \alpha(\lambda_1)}\delta + \varphi(\lambda_2)\frac{\partial e^0(\lambda_2)}{\partial \alpha(\lambda_2)}\frac{\varphi(\lambda_1)}{\varphi(\lambda_2)}\delta = \varphi(\lambda_1)\delta\left[-\frac{1}{\lambda_1}h\left(\frac{\alpha(\lambda_1)}{\lambda_1}\right) + \frac{1}{\lambda_2}h\left(\frac{\alpha(\lambda_2)}{\lambda_2}\right)\right].$$

To show that at the optimal auditing policy, $\frac{\alpha(\lambda_1)}{\lambda_1} < \frac{\alpha(\lambda_2)}{\lambda_2}$, suppose it is not the case, i.e., $\frac{\alpha(\lambda_1)}{\lambda_1} \geqslant \frac{\alpha(\lambda_2)}{\lambda_2}$. Since $h(x)$ is increasing then $h(\frac{\alpha(\lambda_1)}{\lambda_1}) \geqslant h(\frac{\alpha(\lambda_2)}{\lambda_2})$, both numbers being negative. Therefore, $-\frac{1}{\lambda_1}h(\frac{\alpha_1}{\lambda_1}) + \frac{1}{\lambda_2}h(\frac{\alpha_2}{\lambda_2}) < 0$, which implies that total emissions decrease after the reallocation of the budget.

(a) Take two firms, with $\lambda_1 > \lambda_2$. If $\alpha(\lambda_1) > 0$, then either $\alpha(\lambda_2) = t/[\theta'(e^*_{\lambda_2}) + t]$ or, by the previous argument, $\frac{\alpha(\lambda_1)}{\lambda_1} < \frac{\alpha(\lambda_2)}{\lambda_2}$.Therefore, in both cases, $\alpha(\lambda_2) > 0$ whenever $\alpha(\lambda_1) > 0$ and $\lambda_2 < \lambda_1$. Hence, if there is λ^0 such that $\alpha(\lambda^0) = 0$ then $\alpha(\lambda) = 0$ for all $\lambda > \lambda^0$. That is, there exists a $\lambda(B) \leqslant \overline{\lambda}$ such that $\alpha(\lambda) = 0$ for all $\lambda > \lambda(B)$ and $\alpha(\lambda) > 0$ for all $\lambda < \lambda(B)$. Moreover, $\lambda(B) > \underline{\lambda}$ since $\lambda(B) = \underline{\lambda}$ would imply that no firm is audited, which cannot be optimal when $B > 0$.

(b) Consider those firms with $\lambda < \lambda(B)$. When the optimal auditing policy lies at the corner at a certain region of the parameter space, $\alpha(\lambda) = t/[\theta'(e^*_\lambda) + t]$, then $e^0(\lambda) = e^*_\lambda$ in this region, which is an increasing function of λ. When the solution is interior, we know that $\frac{\alpha(\lambda_1)}{\lambda_1} < \frac{\alpha(\lambda_2)}{\lambda_2}$ when $\lambda_1 > \lambda_2$. We claim that this implies that $e^0(\lambda_1) > e^0(\lambda_2)$. Indeed, condition (4) for $\rho = 1$

(that is the relevant condition in region (*b*)) can be rewritten as

$$g'(e^0) - \frac{\alpha}{\lambda}t - \frac{\alpha}{\lambda}\theta'(e^0) = 0.$$

Hence, the emission e^0 is an increasing function of α/λ. □

Proof of Proposition 7. First, we note that the agency can propose an audit policy equivalent to $\alpha(z)$, where it audits with probability 1 any report different from z^0, and with probability $\alpha(z^0)$ the report z^0. Facing this policy, the firm will still decide to report z^0: its expected profits by reporting z^0 do not change, while the profits in case it chooses any other report are at most the same as before. Hence the two policies involve the same final emission level and the same cost (i.e., same probability of auditing). Therefore, for the proof we can restrict attention to the set of audit functions parametrized by (α^0, z^0), where α^0 is the probability with which the firm is audited when it reports z^0, any other report is audited with certainty. Moreover, the policy must be such that the firm does choose z^0.

We now show that the optimal policy always involves $z^0 = 0$. Indeed, given the policy (α^0, z^0), the optimal emission level $e(z^0)$ by the firm is determined by condition (2), for $z = z^0$ and $\alpha = \alpha^0$:

$$\lambda g'(e(z^0)) - \rho\alpha^0 t - \rho\alpha^0\theta'(e(z^0) - z^0) = 0. \tag{7}$$

It can be checked that $e(z^0)$ is an increasing function. Therefore, the best policy that the agency can possibly implement in order to minimize the level of emissions with a budget (probability) α^0 involves $z^0 = 0$ and $e^0 = e(z^0 = 0)$ implicitly defined by

$$\lambda g'(e^0) - \rho\alpha^0 t - \rho\alpha^0\theta'(e^0) = 0. \tag{8}$$

(Let us here remember that we are assuming that the firms do not face limited liability constraints; the amount of penalties possibly paid can only increase without bounds if the polluters have unlimited assets, or the penalties are non-monetary.)

We check that, given the policy $(\alpha^0, z^0 = 0)$, the firm indeed chooses $z^0 = 0$. We denote $\Pi(\alpha^0; \rho)$ the profits of a type $-\rho$ firm subject to the previous policy when it chooses $z^0 = 0$, that is,

$$\Pi(\alpha^0; \rho) \equiv \lambda g(e^0) - \rho\alpha^0 t e^0 - \rho\alpha^0\theta(e^0).$$

The function $\Pi(\alpha^0; \rho)$ is decreasing in α^0. The firm will choose $z^0 = 0$ if its profits are higher than those of its other options. Given that the profits when the firm does not report $z^0 = 0$ do not depend of α^0, we denote them by $\widetilde{\Pi}(\rho)$.

When $\rho \geqslant \widehat{\rho}$, the best a firm can do if it chooses $z > 0$ is reporting truthfully and polluting $e = e^*_\lambda$; hence, $\widetilde{\Pi}(\rho) = \lambda g(e^*_\lambda) - te^*_\lambda$. Given that $\Pi(\alpha^0; \rho)$ is decreasing in α^0, $\Pi(\alpha^0; \rho) \geqslant \widetilde{\Pi}(\rho)$ for all $\alpha^0 \leqslant \mathrm{Min}\{\hat{\alpha}(\rho, \lambda), 1\} = \hat{\alpha}(\rho, \lambda)$ if and only if $\Pi(\hat{\alpha}(\rho, \lambda); \rho) \geqslant \widetilde{\Pi}(\rho)$, i.e., $-\rho\hat{\alpha}(\rho, \lambda)te^*_\lambda - \rho\hat{\alpha}(\rho, \lambda)\theta(e^*_\lambda) \geqslant -te^*_\lambda$. Since $\rho\hat{\alpha}(\rho, \lambda) = t/[\theta'(e^*_\lambda) + t]$, the last inequality holds if and only if, $t/[\theta'(e^*_\lambda) + t] \leqslant te^*_\lambda/[te^*_\lambda + \theta(e^*_\lambda)]$, that is, $\theta(e^*_\lambda) \leqslant \theta'(e^*_\lambda)e^*_\lambda$, which always holds because $\theta(0) = 0$, $\theta'(.) > 0$ and $\theta''(.) > 0$.

When $\rho < \widehat{\rho}$, the firm's optimal report is $z = 0$ even when this report is audited with probability 1, as the rest of the reports. Hence, choosing $z^0 = 0$ when $\alpha(0) = \alpha^0$ cannot give lower profits to the firm than any alternative decision.

Therefore, we have proven that the optimal policy necessarily involves $z^0 = 0$ and e^0 defined by (8), where α^0 is the audit probability of $z^0 = 0$. The same firm's emission and compliance behavior, at the same monitoring cost, is achieved with a monitoring policy that audits every report with probability α^0.

Finally, we note that Eq. (7) implies that the function $e(z^0)$ satisfies:

$$\frac{de}{dz^0} = \frac{\rho\alpha^0\theta''(e(z^0) - z^0)}{\rho\alpha^0\theta''(e(z^0) - z^0) - \lambda g''(e(z^0))} < 1.$$

Therefore, $e(z^0) - z^0$ decreases with z^0. Hence, the expected amount of penalties paid by the firm $\rho\alpha^0\theta(e(z^0) - z^0)$ finds its maximum when $z^0 = 0$, that is, at the proposed audit policy. □

References

[1] Ph. Bontems, J.-M. Bourgeon, Optimal environmental taxation and enforcement policy, Europ. Econ. Rev. 49 (2) (2005) 409–435.

[2] N. Brooks, R. Sethi, The distribution of pollution: community characteristics and exposure to air toxics, J. Environ. Econ. Manage. 32 (1997) 233–250.

[3] M.A. Cohen, Monitoring and enforcement of environmental policy, in: H. Folmer, T. Tiltenberg (Eds.), The International Year-book of Environmental Resource Economics 1999/2000, Edward Elgar Publishing, 1999.

[4] M.A. Cohen, Empirical research on the deterrent effect of environmental monitoring and enforcement, Environ. Law Inst. News Anal. 30 (2000) 10245–10252.

[5] M.L. Cropper, W.E. Oates, Environmental economics: a survey, J. Econ. Lit. 30 (1992) 675–740.

[6] S. Dasgupta, B. Laplante, N. Maningi, H. Wang, Inspections pollution prices, and environmental performance: evidence from china, Ecol. Econ. 36 (2001) 487–498.

[7] D. Epple, M. Vischer, Environmental pollution: modeling occurrence detection and deterrence, J. Law Econ. 27 (1984) 29–60.

[8] J. Foulon, P. Lanoie, B. Laplante, Incentives for pollution control: regulation or information?, J. Environ. Econ. Manage. 44 (2002) 169–187.

[9] L. Franckx, The use of ambient inspections in environmental monitoring and enforcement when the inspection agency cannot commit itself to announced inspection probabilities, J. Environ. Econ. Manage. 43 (1) (2002) 71–92.

[10] D. Garvie, A. Keeler, Incomplete enforcement with endogenous regulatory choice, J. Public Econ. 55 (1994) 141–162.

[11] J.D. Harford, Firm behavior under imperfectly enforceable pollution standards and taxes, J. Environ. Econ. Manage. 5 (1978) 26–43.

[12] J.D. Harford, Self-reporting of Pollution and the firm's behavior under imperfectly enforceable regulation, J. Environ. Econ. Manage. 14 (3) (1987) 293–303.

[13] E. Helland, The enforcement of pollution control laws: inspection, violations and self-reporting, Rev. Econ. Statist. 80 (1) (1998) 141–153.

[14] A. Heyes, Environmental enforcement when Inspectability is endogenous: a model with overshooting properties, Environ. Resource Econ. 4 (5) (1994) 479–494.

[15] A. Heyes, Implementing environmental regulation: enforcement and compliance, J. Regul. Econ. 17 (2) (2000) 107–129.

[16] A. Heyes, A theory of filtered enforcement, J. Environ. Econ. Manage. 43 (1) (2002) 34–46.

[17] R. Innes, Remediation and self-reporting in optimal law enforcement, J. Public Econ. 72 (1999) 379–393.

[18] L. Kaplow, S. Shavell, Optimal law enforcement with self-reporting behavior, J. Polit. Economy 102 (3) (1994) 583–606.

[19] P. Lanoi, B. Laplante, M. Roy, Can capital markets create incentives for pollution control?, Ecol. Econ. 26 (1998) 31–41.

[20] J. Livernois, C.J. McKenna, Trust or consequences enforcing pollution standards with self-reporting, J. Public Econ. 71 (3) (1999) 415–440.

[21] I. Macho-Stadler, D. Pérez-Castrillo, Optimal auditing policy with heterogeneous incomes sources, Int. Econ. Rev. 38 (4) (1997) 951–986.
[22] I. Macho-Stadler, D. Pérez-Castrillo, Auditing with signals, Economica 69 (2002) 1–20.
[23] A. Malik, Avoidance, screening and regulatory enforcement, Rand J. Econ. 21 (3) (1990) 341–353.
[24] A. Malik, Self-reporting and the design of policies for regulating stochastic pollution, J. Environ. Econ. Manage. 24 (1993) 241–257.
[25] OECD, A decade of challenges and developments in reforming environmental enforcement agencies, 2001.
[26] N. Persico, Racial profiling, fairness, and effectiveness of policing, Amer. Econ. Rev. 92 (5) (2002) 1472–1497.
[27] A.M. Polinsky, S. Shavell, The economic theory of public enforcement of the law, J. Econ. Lit. 38 (1) (2000) 45–76.
[28] J.M. Ruiz, Oil spills versus shifting baselines, Mar. Ecol.-Prog. Ser. 282 (2004) 307–309.
[29] A. Sandmo, The Public Economics of Environment, Oxford University Press, Oxford, 2000.
[30] A. Sandmo, Efficient environmental policy with imperfect compliance, Environ. Resource Econ. 23 (2002) 85–103.
[31] J.E. Swierzbinski, Guilty until proven innocent: regulation with costly and limited enforcement, J. Environ. Econ. Manage. 27 (1994) 127–146.