January 2008

The Optimal Pricing of Pollution When Enforcement is Costly

JOHN K. STRANLUND
Department of Resource Economics
University of Massachusetts-Amherst

CARLOS A. CHÁVEZ
Departamento de Economía
Universidad de Concepción

MAURICIO G. VILLENA
Escuela de Negocios
Universidad Adolfo Ibáñez

Correspondence to: John K. Stranlund
Department of Resource Economics
214 Stockbridge Hall
80 Campus Center Way
University of Massachusetts-Amherst
Amherst, MA 01003, USA.
Phone: (413) 545-6328
Fax: (413) 545-5853
E-mail: stranlund@resecon.umass.edu.

Acknowledgements: We gratefully acknowledge financial support from Conicyt-Chile, under Project Fondecyt No. 1060679, and Fondecyt International Cooperation under Project No 7060098. Partial support was also provided by the Cooperative State Research Extension, Education Service, U. S. Department of Agriculture, Massachusetts Agricultural Experiment Station, and the Department of Resource Economics under Project No. MAS00871. Joe Moffitt and John Spraggon provided valuable comments and suggestions. We also thank seminar participants at the University of Massachusetts-Amherst and the University of California-Berkeley for valuable comments.

The Optimal Pricing of Pollution When Enforcement is Costly

Abstract: We consider the pricing of a uniformly mixed pollutant with a model of optimal, possibly firm-specific, emissions taxes and their enforcement under incomplete information about firms' abatement costs, enforcement costs, and pollution damage. We argue that optimality requires an enforcement strategy that induces full compliance by every firm, except possibly when a regulator can base the probabilities of detecting individual violations on observable correlates of violator's actual emissions. Moreover, setting aside several unrealistic special cases, optimality requires discriminatory emissions taxes, except when a regulator is unable to use observable firm-level characteristics to distinguish its expectations of the firms' abatement costs or the costs of monitoring them for compliance from other firms. In many pollution control settings, especially those that have been subject to various forms of environmental regulation in the past, regulators are not likely to be so ill-informed about individual firms. In these settings, policies that set or generate a uniform pollution price are inefficient. These policies include conventional designs involving uniform taxes and competitive emission trading with freely-allocated or auctioned permits.

Keywords: Compliance, Enforcement, Emissions Taxes, Monitoring, Asymmetric Information, Uncertainty

JEL Codes: *L51*, *Q58.*

1. Introduction

In a first-best world of environmental policy, an optimal tax to control emissions of a uniformly mixed pollutant involves a uniform per unit tax set equal to marginal damage from emissions at the efficient level of aggregate emissions. Alternatively, a competitive emissions trading program with either freely-allocated or auctioned permits will generate a uniform price for pollution that is the same as the first-best tax. In a first-best world, however, regulations do not have to be enforced and regulators have complete information about all the benefits and costs of pollution control. These assumptions are always violated in real world applications. In this paper, therefore, we consider the optimal pricing of pollution when compliance must be enforced and regulators have only incomplete information about firms' abatement costs, the costs of regulatory enforcement, and the damages from pollution. Our model is cast as the joint determination of optimal, possibly firm-specific, emissions taxes and their enforcement.[1]

Our efforts produce several new results. The first set of results concerns the determination of firm-specific tax/enforcement policies that achieve an uncertain distribution of individual emissions with minimum expected enforcement costs. In this sense we seek tax/enforcement policies that are cost-effective. A key feature of our work is that we assume throughout that it is not costly to collect emissions taxes from compliant firms, but that it is costly to collect penalties from noncompliant firms. We first demonstrate that under a constant expected marginal penalty for tax evasion, a cost-effective tax/enforcement policy requires sufficient enforcement effort to induce full compliance by all firms. In the theoretical literature on compliance with emissions taxes most authors simply assume that full compliance is not or cannot be achieved (e.g., Harford 1978, 1987; Sandmo 2002, Montero 2002, Cremer and Gahvari 2002, and Macho-Stadler and Perez-Castrillo 2006). This is also true in most theoretical analyses of the compliance and enforcement problem in emissions trading schemes (e.g., Malik 1990, Keeler 1991, Stranlund and Dhanda 1999, Montero 2002).[2] Without downplaying the

[1] In a recent contribution, Bontems and Bourgeon (2005) consider optimal environmental taxes under incomplete information and costly enforcement. They take a standard revelation approach that relies on eliciting truthful reports by firms of their "types". (See Lewis 1996 for a review of this approach). In their work, a policy consists of a type-specific lump sum tax, an emissions standard, a monitoring probability, and a fine for violating the standard. This is very different from the way environmental economists and policymakers usually think of emissions taxes. Emissions taxes are usually per unit taxes, no restrictions are placed on firms' emissions, and noncompliance occurs if a firm attempts to evade its tax liability by under-reporting its emissions. We take this approach in this paper.

[2] Still others in this literature restrict themselves to only full compliance outcomes (Malik 1992, Stranlund and Chavez 2000, Chavez and Stranlund 2003). Only Stranlund (2007) considers the optimal design of emission trading policies with costly enforcement. He demonstrates that when regulators are fully informed about firms' costs of controlling their emissions, to reach an exogenous aggregate emissions target the optimal policy calls for inducing full compliance. Under incomplete information about firm's abatement costs, the full-compliance result continues to hold if a constant marginal penalty for violations is a constant multiple of the price of emissions permits.

relevance of examining the performance of incentive-based policies when enforcement is not sufficient to induce full compliance, our work suggests that these situations may involve sub-optimal policy designs.

However, our assumption of a constant expected marginal penalty for tax evasion is not common in the theoretical literature on compliance with incentive-based policies. Most authors assume expected penalties that are some combination of strictly convex penalty functions, and probabilities of detecting violations that may depend on firms' emissions reports, the regulator's expectation of their emissions, or on their actual emissions. Unfortunately, these assumptions are not justified by appealing to actual enforcement strategies, nor has anyone examined whether there are efficiency justifications for these assumptions.

Consequently, to examine the robustness of our full-compliance result we ask whether it is possible to reduce the expected enforcement costs of achieving an uncertain distribution of individual emissions with a non-constant expected marginal penalty that results in some level of tax evasion. We find that a regulator cannot use noncompliance along with any combination of the firms' emissions reports, its expectations of their emissions, or strictly convex penalty functions to form an efficient tax/enforcement policy—all such strategies result in higher expected enforcement costs than inducing full compliance with a constant expected marginal penalty. However, monitoring effort can be reduced if the probabilities of detecting firms' violations can be based on observable correlates of their actual emissions. Whether such a strategy can also reduce expected enforcement costs depends on the tradeoff between reduced monitoring costs and the expected sanctioning costs that arise from punishing violations.

Our third result comes from determining optimal firm-specific emissions taxes under the assumption that it is not possible to improve on an enforcement strategy of inducing full compliance. Setting aside several unrealistic special cases, we show that discriminatory taxes are optimal except when observable firm characteristics do not provide a regulator with any information about the variation in the firms' marginal abatement costs or the marginal costs of monitoring them for compliance. In this case, the regulator has such poor information about individual firms that it cannot distinguish them from one another in a useful way. While this lack of information is certainly characteristic of many pollution control settings, regulators will not be so ill-informed in others. Particularly in developed countries, firms have been subject to some form of pollution control for many years. Consequently, we suspect that there are many emissions-control situations in which prior experience has provided regulators with information about the costs of monitoring different firms, and may have allowed regulators to determine how observable firm characteristics like output, levels and kinds of inputs, abatement and production technologies, etc., are jointly distributed with their abatement or monitoring costs. With this type of information, even though it is incomplete, optimality requires discriminatory pollution prices.

An important consequence of this result is that when regulators have enough information to distinguish firms from one another, any emissions control policy that sets or generates a uniform price cannot be optimal. In particular, the policies that drive our conventional wisdom about the value of incentive-based policies, like those involving Pigouvian taxes and competitive emissions trading, are actually suboptimal policies. [3]

The rest of the paper is organized as follows. In the next section we develop a model of compliance behavior under emissions taxes and constant expected marginal penalties. In section 3, we lay out the components of the costs of enforcing emissions taxes, and demonstrate that an optimal tax policy with constant expected marginal penalties must induce full compliance by

[3] Beyond the control of uniformly mixed pollutants from point sources, which is the setting for this work as well as all of the literature we discuss, it is well known that discriminatory emissions taxes are optimal when pollutants are spatially differentiated (see Xepapadeas (1997), chapter 2 section 8 for references, including zonal taxes).

every firm, even when a regulator has incomplete information about their abatement costs and the costs of enforcement. In section 4, we examine the robustness of our full compliance result by asking whether a regulator can combine noncompliance and more sophisticated enforcement strategies to reduce expected enforcement costs. In section 5 we use our full-compliance enforcement strategy to determine optimal emissions taxes under this strategy, and demonstrate that the optimal design will often involve discriminatory taxes. We conclude in section 6 with an extended discussion of how our results contribute to the theoretical literature on costly enforcement of incentive-based policies, as well as suggestions for future research.

2. A Model of Compliance Behavior under Emissions Taxes

The regulatory model of this paper is the standard one in which a regulator first commits itself to a tax policy and its enforcement and communicates all the elements of the policy to the regulated firms. The firms then make their optimal choices of emissions and compliance. In this section we examine these choices.

Throughout consider a fixed set N of n heterogeneous risk-neutral firms. These firms may or may not belong to the same industry, but each emits the same uniformly mixed pollutant. A regulator has incomplete information about the firms' abatement costs, but a unique feature of our work is that we allow for the possibility that regulators can use observable firm characteristics to distinguish firms from one another in meaningful ways. For example, past environmental regulations may have provided enough information to regulators to allow them to derive estimates of the parameters of firms' abatement costs as functions of observable production and abatement technologies, or levels and kinds of inputs and outputs. Consequently, let the abatement cost function of firm i be $C(q_i, x_i, \varepsilon_i)$, which is strictly decreasing and strictly convex in its emissions q_i. The variable ε_i is known to the firm but is a random variable from the regulator's perspective, and x_i is a vector of characteristics of the firm that the regulator can observe. Note that the functional form of abatement costs is constant across firms. This is not necessary for our results, but it does reduce our notational burden. Though the form of *C* does not vary, individual abatement cost functions differ according to differences in their observable characteristics and realizations of the random parameters.

The firm's emissions are taxed at rate t_i. It is required to submit a report of its emissions, r_i, and it is noncompliant if it attempts to evade some part of its tax liability by reporting $r_i < q_i$. The regulator cannot determine the firm's compliance status without a costly audit. Let π_i denote the probability that the regulator is able to make this determination. Like most other authors, we assume that monitoring produces a measure of emissions that is accurate enough to judge a firm's compliance status without error. The detection probability is common knowledge and the regulator commits to it at the outset. If monitoring reveals that *i* has under-reported its emissions, it faces a unit penalty of ϕ_i on $q_i - r_i > 0$. Obviously, under this specification the expected penalty is linear. This assumption is not common in the literature on compliance with emissions taxes or tradable emission permits, so we address the value of enforcement strategies that produce alternative forms of the expected penalty function in section 4.[4] The unit penalty may vary among firms to allow for the possibility that this may be part of an efficient policy. However, we restrict it to be no more than a maximal value $\bar{\phi}$, which does not vary across firms.

[4] Linear penalties (i.e., constant marginal penalties) are also not common in the literature. They are, however, very common for actual emissions trading schemes (see Boemare and Quirion 2002 for several examples). There is less documented evidence for actual emissions taxes. However, Poland's sulfur dioxide and nitrogen oxide taxes impose a constant fine of 10 times the tax for noncompliance (Zylics and Spyrka 1994). Under Sweden's tax on nitrogen oxide, violators must pay their unreported tax liability plus interest (personal communication with Claes Englund, officer of the Swedish NOx program). While this may not appear to be much of a deterrent, for our purposes all that matters is that it is a linear fine.

We also assume that $\phi_i > t_i$ throughout. This is a natural assumption because the penalty can be interpreted as recovering evaded taxes plus a punitive element of $\phi_i - t_i$ per unit of under-reported emissions. Perhaps more importantly, this assumption ensures that full compliance is a possible outcome throughout the paper.

To simplify our analysis we restrict it to policies that motivate all firms to reduce their emissions below what they would release in the absence of any sort of regulatory control, but that do not cause any firm to choose zero emissions. Moreover, we assume that each firm has sufficient assets so that the tax or penalty it pays cannot force it into bankruptcy. Under these assumptions firm *i* chooses it emissions and emissions report to solve:

$$\begin{aligned} &\min_{(q_i, r_i)} C(q_i, x_i, \varepsilon_i) + t_i r_i + \pi_i \phi_i (q_i - r_i) \\ &s.t. \qquad q_i - r_i \geq 0,\ r_i \geq 0. \end{aligned} \quad [1]$$

Restricting the firm to $q_i - r_i \geq 0$ follows from the fact that a firm will never have an incentive to report that its emissions are higher than they really are. Let $\mathcal{L}$ denote the Lagrange equation for [1] and let λ_i denote the multiplier attached to the constraint $q_i - r_i \geq 0$. The following first-order conditions are then both necessary and sufficient to determine the firm's optimal choices of emissions and emissions report:

$$\mathcal{L}_q = C_q(q_i, x_i, \varepsilon_i) + \pi_i \phi_i - \lambda_i = 0; \quad [2]$$

$$\mathcal{L}_r = t_i - \pi_i \phi_i + \lambda_i \geq 0,\ r_i \geq 0,\ r_i(t_i - \pi_i \phi_i + \lambda_i) = 0; \quad [3]$$

$$\mathcal{L}_\lambda = -(q_i - r_i) \leq 0,\ \lambda_i \geq 0,\ \lambda_i(q_i - r_i) = 0. \quad [4]$$

Making the common assumption that a firm will comply if it is indifferent between compliance and noncompliance, [3] reveals that a firm's optimal emissions report is:

$$r_i = \begin{cases} q_i \text{ if } & t_i \leq \pi_i \phi \\ 0 \text{ if } & t_i > \pi_i \phi. \end{cases} \quad [5]$$

Thus, the firm provides a truthful report of its emissions when the tax does not exceed the expected marginal penalty. When the tax does exceed the expected marginal penalty, it is cheaper for the firm to report zero emissions and face the expected penalty than to pay the tax.

At this stage, some may object to our formulation of the regulator's enforcement strategy on the grounds that it is implausible that a regulator would not react with an automatic audit if it received a report of zero emissions. While this is certainly true, we show in the next section that it will never be optimal to set a firm's tax and marginal expected penalty so that it reports zero emissions. Thus, even though a report of zero emissions is possible under our specification, it is only possible under a poorly designed policy.

When $t_i \le \pi_i \phi_i$ so that the firm reports its emissions truthfully, [3] becomes $t_i = \pi_i \phi_i - \lambda_i$. Combining this with [2] yields the familiar result that the firm chooses its emissions to equate its marginal abatement cost to the tax; that is, $C_q(q_i, x_i, \varepsilon_i) + t_i = 0$. However, when $t_i > \pi_i \phi_i$ and the firm under-reports its emissions, [4] indicates that $\lambda_i = 0$. In this case [2] becomes $C_q(q_i, x_i, \varepsilon_i) + \pi_i \phi_i = 0$; that is, a noncompliant firms chooses its emissions to equate its marginal abatement cost to the expected marginal penalty. Thus, a firm's optimal choice of emissions is:

$$q_i = \begin{cases} q(t_i, x_i, \varepsilon_i) & \mid\ C_q(q_i, x_i, \varepsilon_i) + t_i = 0, \quad \text{if} \quad t_i \le \pi_i \phi_i \\ q(\pi_i \phi_i, x_i, \varepsilon_i) & \mid\ C_q(q_i, x_i, \varepsilon_i) + \pi_i \phi_i = 0, \quad \text{if} \quad t_i > \pi_i \phi_i. \end{cases} \quad [6]$$

3. The costs of enforcing emissions taxes and the optimality of full compliance

The regulatory objective of this paper is to choose a system of firm-specific emissions taxes and enforcement strategies to minimize the expected social costs of the regulation. These costs include the regulator's expectations of aggregate abatement costs, pollution damage, and enforcement costs. In this section we focus on expected enforcement costs, which include the regulator's expectations of aggregate monitoring costs and the costs of collecting penalties from noncompliant firms. Under the assumption that collecting tax revenue from compliant firms is cheaper than collecting penalties from noncompliant firms, we state and prove a "cost-effective enforcement" result. Specifically, we show how a tax/enforcement policy with a constant expected marginal penalty should be designed to minimize the expected enforcement costs of inducing an arbitrary, fixed, but imperfectly known set of individual emissions.

We allow the costs of monitoring to vary across firms to reflect the possibility that the regulator will find it harder to determine the compliance status of some firms than others.[5] Monitoring costs may vary across firms for several reasons. The location of a firm may affect the costs of inspecting their emissions. Plants with more discharge points may be harder to monitor than others. The variation in abatement and production technologies, particularly if firms belong to different industries, may also produce variation in monitoring costs. Like our approach to modeling firms' abatement costs, a regulator is uncertain about the costs of monitoring individual firms, but it might possess information about how monitoring costs are correlated with observable firm characteristics. Consequently, let the cost of monitoring firm *i* be $m(\pi_i, x_i, \mu_i)$, which is increasing and convex in the detection probability, π_i. The regulator is uncertain about monitoring costs because it cannot observe the parameter μ_i, but it may have some information about how monitoring costs vary with the firms' observables, x_i. For simplicity we assume that aggregate monitoring costs are simply the sum of the individual monitoring costs functions; thus, the regulator's conditional expectation of aggregate monitoring costs is $E\left(\sum_N m(\pi_i, x_i, \mu_i)\right)$, where *E* denotes the expectation operator throughout the paper.

[5] This assumption is closely related to the assumption that individuals vary in their probabilities of apprehension, which was first analyzed by Bebchuk and Kaplow (1993). Macho-Stadler and Perez-Castrillo (2006) assume heterogeneous probabilities of apprehension in their study of enforcing emissions taxes.

We assume throughout that tax revenue and revenue from penalizing noncompliant firms are simple transfers with no real effects. Despite this, society is not indifferent about collecting them; in particular, penalizing noncompliant firms may involve significant costs. These include the government's costs of generating sufficient evidence to get a court to agree with their finding of a violation and the imposition of a penalty. Accused firms may mount costly challenges to any finding of noncompliance and the imposition of a penalty, and the government may respond with its own costly efforts to fight off these challenges.[6] On the other hand, a compliant firm reports the full extent of its emissions and, in doing so, essentially admits liability for these emissions. With this admission the government does not need to generate the evidence that would be necessary to impose a penalty for noncompliance. Moreover, a firm that admits its liability is not likely to challenge the imposition of the tax.

Therefore, we feel that is natural to assume that imposing and collecting penalties from noncompliant firms is more costly that collecting taxes from compliant firms, and incorporate this assumption into our model. To do so in a simple way, suppose that collecting emissions taxes is costless, but that $s_i > 0$ is the cost of collecting the penalty from firm *i* if it is caught evading its tax liability. Like the costs of monitoring individual firms, the regulator need not have complete information about the costs of collecting penalties from individual firms. Although s_i may be a function of the size of the firm's penalty and possibly its observable characteristics, our results do not depend on specifying these relationships. We do, however, assume that the aggregate expected cost of collecting penalties is linear in the costs of collecting

[6] Although it is perfectly reasonable to assume that penalizing firms is costly, none of the work in the literature on enforcing emissions taxes that we are aware of deals explicitly with these costs. In the literature on enforcing emissions trading policies, only Stranlund (2007) assumes that imposing and collecting penalties is costly. The assumption of costly sanctions is only a bit more common in the literature on enforcing emissions standards. For examples, see Malik (1993) and Arguedas (2007). Costly sanctions are also not very common in the much larger literature on optimal law enforcement; however, see Polinsky and Shavell (1992) for an analysis of how costly sanctions affect the determination of optimal law enforcement.

penalties from individual firms. If N^{nc} denotes the subset of firms that are noncompliant, the aggregate expected cost of collecting penalties is $\sum_{N^{nc}} \pi_i s_i$.[7]

Despite our weak assumptions about the expected costs of enforcing emissions taxes, we are able to prove the following proposition concerning the optimal enforcement of these policies.

Proposition 1: Consider a tax/enforcement policy, (t_i, π_i, ϕ_i), i = 1, … , n, with $t_i < \phi_i \le \overline{\phi}$ for each i. Suppose that firms react to this policy with emissions q_i, i = 1, … , n. This distribution of emissions is achieved with minimum expected aggregate enforcement costs if and only if $t_i = \pi_i \overline{\phi}$ for each i = 1, … , n. With taxes and monitoring set in this way, each firm is compliant.

Proof: The proof proceeds by first showing that any policy involving $t_i \ne \pi_i \phi_i$ for some i can be modified to reduce enforcement costs without changing the distribution of emissions. First suppose that $t_i > \pi_i \phi_i$ for some i. Then, [5] indicates that $r_i = 0$, and [6] indicates that $q_i = q(\pi_i \phi_i, x_i, \varepsilon_i)$; that is the firm is fully noncompliant. Alternatively, hold π_i constant so that aggregate monitoring costs do not change, but reduce t_i so that $t_i = \pi_i \phi_i$. The firm will then choose $r_i = q_i$ so that it is now compliant, but it does not change it emissions because $q(\pi_i \phi_i, x_i, \varepsilon_i) = q(t_i, x_i, \varepsilon_i)$. Moreover, changing t_i in this way does not affect the decisions of any of the other firms. However, reducing t_i to $\pi_i \phi_i$ eliminates the expected costs of penalizing the firm; hence, aggregate expected enforcement costs are reduced.

[7] Any enforcement strategy is likely to involve fixed monitoring costs and sanctioning costs, which we do not model. Adding these fixed costs does not change any of the results of our work as long as they are not so high that it is optimal to forego regulation altogether.

Now suppose that $t_i < \pi_i \phi_i$ for some *i*. In this case, [5] and [6] reveal that the firm is compliant so that $r_i = q(t_i, x_i, \varepsilon_i)$. However, if π_i is reduced so that $t_i = \pi_i \phi_i$, the firm does not change its choice of emissions and it remains compliant. This change in π_i does not affect the decisions of the other firms, but aggregate expected enforcement costs are reduced because expected monitoring costs decrease.

Therefore, minimizing expected aggregate enforcement costs requires $t_i = \pi_i \phi_i$ for each *i* = 1, … , *n*. Clearly, given the tax rates, monitoring of all firms can be minimized by setting the unit penalties as high as is allowed while maintaining the equality between the tax rates and the expected marginal penalties. Therefore, minimizing expected aggregate enforcement costs requires $t_i = \pi_i \overline{\phi}$ for each *i* = 1, … , *n*. QED.

It is important to note that the proposition holds despite the regulator's uncertainty about the firms' abatement costs. This uncertainty implies that the regulator is uncertain about the distribution of individual emissions that will result from a particular policy; however, whatever q_i, *i* = 1, … , *n*, results, the expected enforcement costs of holding the firms to this distribution of emissions are minimized by choosing $t_i = \pi_i \overline{\phi}$, *i* = 1, … , *n*. Moreover, the proposition holds despite the regulator's uncertainty about monitoring and sanctioning costs. All the regulator has to know is that expected aggregate monitoring costs are increasing in individual monitoring levels, and expected aggregate sanctioning costs are increasing in the costs of penalizing individual firms.[8]

[8] Clearly, inducing full compliance requires the regulator to commit to a monitoring strategy, as we have assumed. If it was unable to commit it would have an incentive to reduce its monitoring effort ex post if it knew that all firms were compliant. While a few other authors have modelled enforcement of environmental regulations under the assumption that a regulator cannot commit to monitoring strategies (e.g. Grieson and Singh 1990 and Franckx

The cost-effectiveness of inducing full compliance depends on three assumptions that differ from the rest of the literature on enforcing incentive-based environmental policies. Our assumption that it is costly to collect penalties from noncompliant firms is crucial, because the fundamental value of inducing full compliance is to avoid these costs. In the absence of sanctioning costs, Proposition 1 does not hold because society would be completely indifferent between allowing noncompliance and inducing full compliance. Moreover, we have given the regulator the freedom to choose firm-specific tax rates. All others assume a uniform tax rate that is often fixed. Finally, no one else to our knowledge specifies enforcement strategies that produce constant expected marginal penalties. We now examine the robustness of our full-compliance policy recommendation under non-constant expected marginal penalties.

4. The Robustness of the Optimality of Full Compliance

The focus on positive violation choices in the literature on emissions tax enforcement is accomplished in part with the assumption that expected marginal penalties are functions of the firms' choices of emissions, emissions reports, and evaded taxes. For example, Harford (1978 and 1987) and Sandmo (2002) assume strictly convex expected penalty functions that involve monitoring probabilities that depend on firms' emissions and their emissions reports, and penalty functions that are strictly convex in the firms' violations. That is, they assume expected penalties of the form $\pi_i(q_i, r_i) f_i(q_i - r_i)$, where f_i is a penalty function.[9] In this section we ask the following question: Can a regulator design a tax/enforcement policy with a non-constant

---

2001), we maintain the more common assumption of regulatory commitment for two reasons. First there is clear value to the ability to commit that derives from the cost-effectiveness of inducing full compliance. Second, we observe real cases in which regulators do commit to enforcement strategies that achieve full (or nearly full) compliance. For example, the EPA's SO_2 and NO_X Trading programs were designed to achieve full compliance and have largely succeeded in this regard.

[9] Malik (1990) and vanEgteren and Weber (1996) assume the same form in the context of emissions trading, except that a firm's emissions report is replaced with a firm's permit holdings.

expected marginal penalty that leads to lower expected enforcement costs than inducing full compliance with a constant expected marginal penalty? Recall that aggregate expected monitoring and sanctioning costs are linear in the costs of monitoring and sanctioning individual firms. Thus, we can answer this question from the perspective of alternative tax/enforcement policies for a single firm and apply the result in the aggregate.

Let us introduce a small amount of new notation for this section. Denote a tax/enforcement policy for firm *i* as p_i, consisting of a tax and expected penalty function, under which the firm optimally chooses its emissions $q\left(p_i, x_i, \varepsilon_i\right)$ and emissions report $r\left(p_i, x_i, \varepsilon_i\right)$. Although the regulator cannot determine the true values of $q\left(p_i, x_i, \varepsilon_i\right)$ and $r\left(p_i, x_i, \varepsilon_i\right)$ ex ante, it does know the firm's decision criterion, and hence can form conditional expectations of the firm's emissions and emissions report, $E\left(q\left(p_i, x_i, \varepsilon_i\right)\right)$ and $E\left(r\left(p_i, x_i, \varepsilon_i\right)\right)$. To conserve notation let $\overline{q}\left(p_i\right) = E\left(q\left(p_i, x_i, \varepsilon_i\right)\right)$ and $\overline{r}\left(p_i\right) = E\left(r\left(p_i, x_i, \varepsilon_i\right)\right)$.

We consider three policy types that induce the same expected emissions from the firm. The first is the policy of Proposition 1, which we denote as $p_i^c = \left[t_i^c, \pi_i^c \phi_i (q_i - r_i)\right]$. (We do not set the unit penalty at its maximum level in this section because, as will be obvious shortly, our analysis depends on being able to vary ϕ_i). The superscript *c* indicates that the tax and the detection probability are chosen to induce full compliance, given the unit penalty ϕ_i. Recall that motivating the firm to be compliant with minimal monitoring requires $t_i^c = \pi_i^c \phi_i$. The regulator's expectation of the firm's emissions under p_i^c is $\overline{q}\left(p_i^c\right)$.

The other policies, $p_i^q = \left[t_i^q, \pi_i(q_i, r_i) f_i(q_i - r_i)\right]$ and $p_i^{\overline{q}} = \left[t_i^{\overline{q}}, \pi_i\left(\overline{q}\left(p_i^{\overline{q}}\right), r_i\right) f_i(q_i - r_i)\right]$, feature non-constant expected marginal penalties. Under p_i^q, the detection probability is based in part on the firm's actual emissions. Of course, it is not possible to do this directly because a firm's emissions are unknown until it is actually audited. Sandmo (2002) recognizes this, but justifies conditioning a regulator's monitoring strategy on a firm's emissions by assuming that emissions produce observable correlates that a regulator can use to allocate its monitoring effort. Perhaps higher emissions are associated with more smoke leaving a stationary pollution source, or elevated ambient concentrations of a pollutant can be linked to higher emissions from a particular source. When emissions do not have such observable correlates an alternative is that the regulator forms an expectation of the firm's emissions given a tax/enforcement policy and then uses this expectation to refine its monitoring strategy. (Malik 1990 suggests this approach). This leads to policy $p_i^{\bar{q}}$, under which the detection probability is conditioned on the regulator's expectation of the firm's emissions $\bar{q}\left(p_i^{\bar{q}}\right)$. Note that holding the firm's expected emissions constant across the policies will likely require that they involve different tax rates.

To simplify our analysis we assume that the functional forms of π_i and f_i are the same under p_i^q and $p_i^{\bar{q}}$, and that π_i and f_i' are linear functions. The latter assumptions imply that the regulator's prior expectation of the detection probability it will have to maintain under policy $p_i \in \left(p_i^q, p_i^{\bar{q}}\right)$ is $E\left[\pi\left(q\left(p_i, x_i, \varepsilon_i\right), r\left(p_i, x_i, \varepsilon_i\right)\right)\right] = \pi\left(\bar{q}\left(p_i\right), \bar{r}\left(p_i\right)\right)$, and that its expectation of the firm's marginal penalty is $E\left[f'\left(q\left(p_i, x_i, \varepsilon_i\right) - r\left(p_i, x_i, \varepsilon_i\right)\right)\right] = f'\left(\bar{q}\left(p_i\right) - \bar{r}\left(p_i\right)\right)$.

The results of this section are based on comparisons of the detection probability under p_i^c to the regulator's expectations of the detection probabilities under p_i^q and $p_i^{\bar{q}}$ that induce the same expected emissions from the firm. Clearly, detection probabilities are easily adjusted by changing marginal penalties. So our results do not depend on arbitrary differences in marginal penalties, we assume that the equilibrium expected marginal penalties under p_i^q and $p_i^{\bar{q}}$ are equal to the constant marginal penalty under p_i^c; that is, $\phi_i = f'\left(\bar{q}\left(p_i\right) - \bar{r}\left(p_i\right)\right)$ for $p_i \in \left(p_i^q, p_i^{\bar{q}}\right)$. The proof of the following proposition is in the Appendix.

Proposition 2: Suppose that a regulator wishes to induce a fixed level of expected emissions from a firm. Relative to inducing the firm's compliance with a constant expected marginal penalty, a regulator can reduce its expected monitoring of the firm if and only if the probability of detection is a strictly increasing function of the firm's actual emissions and the regulator implements an enforcement strategy that it expects will result in the firm's noncompliance.

Since inducing full compliance with a constant expected marginal penalty does not incur expected sanctioning costs, implementing an enforcement strategy that results in firms' noncompliance can only reduce expected enforcement costs if it involves significantly lower detection probabilities. Since expected detection probabilities are lower only if they are increasing in the firms' actual emissions, the following Corollary follows immediately from Proposition 2.

Corollary: Any tax/enforcement policy that the regulator expects will result in firms' noncompliance and that features any combination of a strictly convex penalty function, the firms' emissions report, and the regulator's expectations of the firms' emissions will produce higher expected enforcement costs than a policy that motivates the firms to be compliant with constant expected marginal penalties.

Proposition 2 and its Corollary are new results that clarify the value of enforcement strategies that are commonly assumed by others. Despite the complexity of these strategies, it appears that the only way a regulator can improve on the simple strategy of inducing full compliance with constant expected marginal penalties is if the probabilities of detecting the violations of noncompliant firms are strictly increasing functions of the firms' actual emissions. Interestingly, a strictly convex penalty, firms' emissions reports, and the regulator's expectations of the firms' emissions cannot be used to reduce enforcement costs.[10]

Note carefully that Proposition 2 does not tell us that making detection probabilities increasing functions of the firms' actual emissions and allowing them to be noncompliant will lead to lower expected enforcement costs. This will happen only when the expected value of reduced monitoring effort outweighs the additional expected costs of penalizing noncompliance; however, there is nothing in our model that guarantees this.[11] Moreover, it may not be possible to base a monitoring strategy on a firm's actual emissions. We have already noted that a regulator cannot do this directly because a firm's emissions are hidden until it is actually audited. It may be possible to do so indirectly if emissions have observable correlates that a regulator can use, but not all types of emissions have such useful correlates. Thus, the opportunities to use noncompliance to reduce expected enforcement costs may be quite limited. Certainly, these opportunities are much more limited than what is implied by the existing literature on enforcing emissions taxes.

---

[10] The Corollary is related to the basic result of Stranlund (2007) in his study of minimizing the aggregate abatement and enforcement costs of holding firms to a fixed aggregate cap on emissions with competitive emissions trading. Under the assumption that monitoring probabilities are invariant to firms' choices of emissions and permit holdings, Stranlund demonstrates that any enforcement strategy that features a strictly convex penalty and allows noncompliance involves higher expected enforcement costs than a strategy that induces full compliance with a constant marginal penalty. In view of the results of this section it would be interesting to reconsider Stranlund's results under monitoring probabilities that might depend on the firms' choices.

[11] We have not fully explored this tradeoff here, so it might be interesting in future work to determine the parameterizations of our model under which noncompliance and detection probabilities that are increasing in firms' emissions lead to lower expected enforcement costs.

Even if it is possible to use firms' actual emissions indirectly and their noncompliance to reduce expected enforcement costs, this involves uncertainty that inducing compliance with a constant expected marginal penalty does not. Under the former policy a regulator can never be certain of the firms' compliance choices and, consequently, it cannot be certain of the resources it will expend on monitoring and punishing noncompliance. On the other hand, all of these variables are completely known under a policy of inducing full compliance—the regulator is certain that all firms will be compliant, sanctioning costs are zero, and monitoring effort is fully determined by the marginal penalty and individual tax rates, all of which are under the control of the regulator. In the real world, regulators may prefer this high level of control to an expectation that enforcement costs may be reduced.

5. Optimal emission taxes under incomplete information and costly enforcement

Given our obvious pessimism about the value of using firms' noncompliance to reduce the expected costs of enforcing emissions taxes, in this section we incorporate our full compliance strategy of Proposition 1 to determine optimal emissions taxes under incomplete information and costly enforcement. Our primary focus now is on whether an optimal policy involves discriminatory taxes or whether a regulator should set a single tax that applies to all firms.

Clearly, if a regulator's information about individual firms is so poor that it is unable to distinguish them from one another in useful ways, it has no basis for choosing discriminatory taxes. While such poor information may be characteristic of some pollution control settings, it certainly is not a universal feature. As we stated in the introduction, we suspect that in many situations, particularly those in which firms have been subject to control policies in the past, regulators can probably observe individual firm characteristics that provide some information about their unknown abatement and monitoring cost parameters.

Suppose that pollution damage is an imperfectly known, increasing function of aggregate emissions, $D\left(\sum q_i, \delta\right)$, where δ is a random variable. The regulator knows the joint distribution of the firms' unknown abatement cost and monitoring cost parameters, their observable characteristics, and the unknown damage parameter. With this knowledge it can form an expectation of the social costs of pollution and its control, conditional on its observations of the firms' characteristics:

$$E\left\{\sum C(q_i, x_i, \varepsilon_i) + \sum m(\pi_i, x_i, \mu_i) + D\left(\sum q_i, \delta\right)\right\}. \quad [7]$$

Since the regulator will enforce the optimal policy so that all firms are compliant, from Proposition 1 it constrains the minimization of [7] by choosing $(t_i, \pi_i),\ i = 1, \ldots, n$, so that

$$t_i = \pi_i \overline{\phi},\ i = 1, \ldots, n. \quad [8]$$

From [6], under this policy the regulator knows that the firms will choose their emissions so that

$$C_q(q_i, x_i, \varepsilon_i) + t_i = 0,\ i = 1, \ldots, n, \quad [9]$$

which implicitly define their emissions as

$$q_i = q(t_i, x_i, \varepsilon_i)\ i = 1, \ldots, n. \quad [10]$$

Substituting [8] and [10] into [7] gives us the regulator's conditional expectation of the social cost function in terms of well-enforced, firm-specific tax rates:

$$E\left\{\sum C(q(t_i, x_i, \varepsilon_i), x_i, \varepsilon_i) + \sum m(t_i / \overline{\phi}, x_i, \mu_i) + D\left(\sum q(t_i, x_i, \varepsilon_i), \delta\right)\right\}. \quad [11]$$

Assuming that [11] is strictly convex in $(t_1, \ldots, t_n)$ and that optimality calls for a positive tax for each firm, the following first-order conditions uniquely identify the optimal tax rates for each firm *k*:

$$E\left(C_q(q(t_k,x_k,\varepsilon_k),x_k,\varepsilon_k)q_t(t_k,x_k,\varepsilon_k\right)+E\left(m_\pi(t_k/\bar{\phi},x_k,\mu_k)\right)\big/\bar{\phi}$$
$$+E\left(D'\left(\sum q(t_i,x_i,\varepsilon_i),\delta\right)q_t(t_k,x_k,\varepsilon_k)\right)=0,\ k=1,\ldots,n.$$

Substitute [9] into these and rearrange the results to obtain

$$t_k=\frac{E\left(D'\left(\sum q(t_i,x_i,\varepsilon_i),\delta\right)q_t(t_k,x_k,\varepsilon_k)\right)}{E\left(q_t(t_k,x_k,\varepsilon_k)\right)}+\frac{E\left(m_\pi(t_k/\bar{\phi},x_k,\mu_k)\right)}{\bar{\phi}E\left(q_t(t_k,x_k,\varepsilon_k)\right)},\ k=1,\ldots,n. \qquad [12]$$

Using the definition of the covariance between random variables, the first term on the right hand side of equations [12] is

$$E\left(D'\left(\sum q(t_i,x_i,\varepsilon_i),\delta\right)\right)+\frac{Cov\left(D'\left(\sum q(t_i,x_i,\varepsilon_i),\delta\right),q_t(t_k,x_k,\varepsilon_k)\right)}{E\left(q_t(t_k,x_k,\varepsilon_k)\right)},\ k=1,\ldots,n, \qquad [13]$$

where *Cov* denotes the covariance operator. Moreover, use [10] to obtain

$$q_t(t_k,x_k,\varepsilon_k)=-1\big/C_{qq}(q_k,x_k,\varepsilon_k),\ k=1,\ldots,n. \qquad [14]$$

These indicate that the firms' marginal responses to their taxes are equal to the reciprocal of the slopes of their marginal abatement cost functions. Substitute [13] and [14] into [12] to obtain

$$t_k=E\left(D'\left(\sum q(t_i,x_i,\varepsilon_i),\delta\right)\right)+\frac{Cov\left(D'\left(\sum q(t_i,x_i,\varepsilon_i),\delta\right),-1\big/C_{qq}(q_k,x_k,\varepsilon_k)\right)}{E\left(-1\big/C_{qq}(q_k,x_k,\varepsilon_k)\right)}$$

$$+\frac{E\left(m_\pi(t_k/\bar{\phi},x_k,\mu_k)\right)}{\bar{\phi}E\left(-1\big/C_{qq}(q_k,x_k,\varepsilon_k)\right)},\ k=1,\ldots,n. \qquad [15]$$

Note that the first term on the right hand side is the regulator's expectation of marginal damage. Since this term appears in all of the equations in [15], our final proposition follows immediately.

Proposition 3: An optimal policy of well-enforced emissions taxes under incomplete information about firms' abatement and monitoring costs involves discriminatory taxes if and only if

$$\frac{Cov\left(D'\left(\sum q(t_i,x_i,\varepsilon_i),\delta\right),-1/C_{qq}(q_j,x_j,\varepsilon_j)\right)}{E\left(-1/C_{qq}(q_j,x_j,\varepsilon_j)\right)}+\frac{E\left(m_\pi(t_j/\bar{\phi},x_j,\mu_j)\right)}{\bar{\phi}E\left(-1/C_{qq}(q_j,x_j,\varepsilon_j)\right)}$$

$$\neq\frac{Cov\left(D'\left(\sum q(t_i,x_i,\varepsilon_i),\delta\right),-1/C_{qq}(q_k,x_k,\varepsilon_k)\right)}{E\left(-1/C_{qq}(q_k,x_k,\varepsilon_k)\right)}+\frac{E\left(m_\pi(t_k/\bar{\phi},x_k,\mu_k)\right)}{\bar{\phi}E\left(-1/C_{qq}(q_k,x_k,\varepsilon_k)\right)}, \qquad [16]$$

for some *j* and *k*.

Proposition 3 indicates that there are three potential sources of variation in optimal individual emissions taxes: variation in the regulator's conditional expectations of the marginal costs of monitoring the firms, $E\left(m_\pi(t_k/\bar{\phi},x_k,\mu_k)\right)$, $k=1,\ldots,n$; variation in its conditional expectations of the reciprocal of the slopes of the firms' marginal abatement cost functions, $E\left(-1/C_{qq}(q_k,x_k,\varepsilon_k)\right)$, $k=1,\ldots,n$, and variation in the covariances between marginal damage and the reciprocal of the slopes of the firms' marginal abatement cost functions, Before we discuss these sources of variation let us highlight what appears to be fundamental justification for choosing a uniform tax to control a uniformly mixed pollutant. This is when the firms' observable characteristics do not provide the regulator with any information about the variation of the slopes of the firms' marginal abatement costs, or the marginal costs of monitoring them for compliance. In this case the terms on both sides of equation [16] do not vary across firms, and the regulator chooses a uniform tax because it cannot distinguish the firms from one another. Thus, the fundamental justification for setting a uniform tax to control a uniformly mixed pollutant is that a regulator has very poor information about individual firms.[12]

[12] There are also several unrealistic special cases under which a uniform tax is efficient. For example, suppose that the covariance terms in [15] are equal to zero. Then, optimality calls for a uniform tax if marginal monitoring costs are equal to zero, the firms' marginal monitoring costs and the slopes of their marginal abatement costs are the same, or if their observable characteristics are the same.

When a regulator has somewhat better, but still incomplete, information about individual firms, their optimal tax rates will vary. Admittedly, the number of distinct tax rates may be small if a regulator has only coarse information about individual firms. For example, suppose that a control situation involves the firms from a number of industries and that the regulator knows something about how monitoring or abatement costs differ across the industries, but is unable to distinguish firms within industries. In this case the number of distinct tax rates may simply be equal to the number of industries involved. Or, imagine a setting involving the emissions of the firms in a single industry that use only a small number of distinct abatement technologies to control their emissions. If this piece of information is the only characteristic that a regulator can use to distinguish the firms' abatement or monitoring costs, the number of tax rates may be equal to the number of available control technologies. Depending on the degree of heterogeneity in the population of regulated firms, more detailed information about each of them may lead to a greater number of distinct tax rates.

Now let us look into the sources of variation of individual tax rates when a regulator can distinguish at least some firms from at least some others. The first terms on both sides of [16] are interesting because they do not depend at all on the costs of enforcement. That is, even if one assumes zero enforcement costs, incomplete information about firms' abatement costs can produce variation in optimal tax rates when the regulator has some information that allows it to distinguish the slopes of the firms' marginal abatement cost functions from one another. Note that the first terms on both sides of [16] are zero if the covariance terms are zero. This would occur if marginal damage is a known constant, or if the slopes of the firms' marginal abatement cost functions are known constants. This latter assumption is important because it is common to model uncertainty about abatement costs as a random shift of only the intercept of marginal abatement costs. Under this assumption and the equally common and unrealistic assumption that enforcement is free, optimality requires a uniform tax set equal to the regulator's expectation of marginal damage under the tax.[13]

However, we are mainly interested in how enforcement costs induce discriminatory emission taxes, which is captured by the variation in $E\left(m_{\pi}(t_k/\bar{\phi}, x_k, \mu_k)\right)\Big/\bar{\phi}E\left(-1\big/C_{qq}(q_k, x_k, \varepsilon_k)\right)$, k = 1, …, n. Since this term is negative, the optimal tax on a firm will tend to be lower as this term is lower. This is intuitive because it reflects the regulator's expectation of the increase in monitoring costs associated with inducing lower emissions from a firm with a well-enforced tax. Inducing a marginal decrease in the emissions of a firm with a more steeply sloped marginal abatement cost curve (i.e., higher $C_{qq}(q_k, x_k, \varepsilon_k)$) requires a relatively greater increase in its tax and, consequently, a relatively greater increase in monitoring to maintain the firm's compliance. Therefore, to conserve monitoring costs, optimal taxes will tend to be lower for firms that the regulator expects have steeper marginal abatement cost functions. For the same reason, tax rates will tend to be lower for firms that the regulator expects are more difficult to monitor, and hence, have higher marginal monitoring cost functions.

Suppose that the slopes of the firms' marginal abatement cost functions vary and the regulator has some information about this variation. Then the optimal tax rates will vary across the firms even if their monitoring cost functions do not vary. Thus, discriminatory taxes do not require variation in enforcement costs. This is particularly noteworthy because it implies that simply recognizing that enforcement is costly will often be sufficient justification for imposing discriminatory taxes.

[13] To our knowledge, the result that uncertainty about abatement costs by itself can produce discriminatory pollution prices is new. Given the focus of this paper on the role of enforcement costs, we do not explore this issue in depth, but believe that it may be an interesting area for future research.

6. Concluding Discussion

We have examined the optimal pricing of a uniformly mixed pollutant when enforcement is costly and regulators have incomplete information about firms' abatement costs and the costs of enforcement. We have argued that an optimal policy calls on regulators to devote sufficient enforcement resources to induce full compliance by all firms, except, possibly, when the probabilities of detecting violations are increasing functions of observable correlates of their actual emissions. Moreover, enforcement costs will typically induce discriminatory pollution prices, except when regulators have very poor information about individual firms. To conclude this paper we discuss some of the ways that our results contribute to the literature on the design of incentive-based policies, and suggest ways to continue this line of research.

As we noted in the introduction, the related literature has not dealt squarely with the possibility that inducing full compliance may be a component of an optimal tax policy. Harford (1978, 1987), Cremer and Gahvari (2002), and Sandmo (2002) all restrict their analyses by assuming that firms' violations are always interior choices, but make little attempt to justify this modeling restriction. Montero (2002), in his study of price vs. quantity regulation with costly enforcement, explicitly assumes that monitoring costs are large enough and penalties are restricted enough so that full compliance is not socially optimal (page 439). Our full compliance result depends on our assumptions that the unit penalty each firm faces exceeds its tax rate; that collecting penalties is costly but collecting taxes is not, and that regulators are not able to use information about firms' actual emissions to conserve enforcement costs. Given these assumptions, optimality requires full compliance regardless of monitoring costs. To be sure, monitoring costs will affect the optimal tax rates, but they do not affect the decision to induce full compliance.

Our results about the optimality of inducing full compliance stand in stark contradiction to results published recently in this journal. Macho-Stadler and Perez-Castrillo (2006) assume that a budget-constrained regulator seeks to minimize aggregate emissions by choosing an audit policy to enforce a fixed uniform emissions tax. In their model the probability of detection is independent of a firm's emissions and emission report, but violations are punished with a strictly convex penalty. Under these conditions, they show that it will never be optimal for a regulator to induce full compliance unless its budget is very large. In fact, their results suggest that an optimal monitoring strategy will cause firms to almost always report zero emissions.

In contrast, it is easy to demonstrate that our full-compliance result continues to hold if we use our model to address Macho-Stadler's and Perez-Castrillo's policy problem. This is so because we can cast the problem of minimizing aggregate emissions with a fixed enforcement budget as a two-stage process; that is, first determine the tax/enforcement strategy that minimizes the expected enforcement costs of holding each firm to an arbitrary level of emissions, and then choose individual tax rates to minimize aggregate emissions, given the regulator's enforcement budget. Under Macho-Stadler's and Perez-Castrillo's assumption that the probabilities of detecting noncompliance are independent of the firms' actual, our Proposition 2 and its Corollary imply that the solution to the first stage of this problem is to induce full compliance with constant expected marginal penalties. Thus, a tax/enforcement policy that produces the greatest amount of environmental protection that is enforceable with a fixed budget requires that all firms be compliant.[14]

[14] Moreover, it is easy to show that a budget-constrained regulator will choose differential tax rates if it has information on the variation of the firms' marginal abatement costs and marginal monitoring costs.

There are two reasons for the difference between this policy recommendation and that of Macho-Stadler and Perez-Castrillo. First, we assume that sanctioning noncompliant firms is costly, while they assume these costs away. Thus, given a level of emissions from a firm, in their model society is completely indifferent about whether the firm is compliant or not. We argue that this is simply not plausible. The second reason is that Macho-Stadler and Perez-Castrillo take a uniform emissions tax as exogenous, while we allow a regulator to choose firm-specific taxes and insist that these be chosen with an accompanying enforcement strategy to minimize expected enforcement costs. Thus, Macho-Stadler and Perez-Castrillo limit their analysis in a way that does not allow a regulator to truly maximize the productivity of its enforcement budget in limiting aggregate emissions.

Of course, there are other features of the regulatory environment that we, along with all others in the literature, have assumed away that might give regulators the opportunity to conserve enforcement costs by allowing noncompliance. An anonymous reviewer suggested to us that we could interpret our detection probability as the probability of detecting *and* punishing noncompliant choices. If the likelihoods of convicting noncompliant firms are higher when their violations are higher, then the probabilities of detection and punishment would be increasing functions of the firms' actual emissions. In this case, our Proposition 2 suggests that a regulator may be able to exploit an increasing likelihood of punishing violators to reduce its monitoring effort.

Moreover, we have taken the standard approach of assuming that a regulator commits to a policy and communicates this to the firms before they make their decisions. However, regulators might be unable or unwilling to commit to policies that produce full compliance. In fact, a regulator may be motivated to keep elements of a policy hidden, in particular its monitoring strategy. If firms are averse to the ambiguity about the likelihood that they will be inspected, a regulator may be able to use their noncompliance to conserve enforcement costs. We have also assumed that monitoring produces a perfectly accurate judgment of a firm's compliance status. Inducing full compliance when monitoring produces errors may not be desirable, particularly given that some compliant firms may be punished for violations they did not commit. All of these features deserve rigorous investigation in future work.

Our second contribution, that enforcement costs will often induce discriminatory emissions taxes, is not widely known, but we cannot claim that it is entirely new. Cremer and Gahvari (2002) also examine the optimal design of an emissions tax that is costly to enforce. While they determine optimal taxes and their enforcement jointly as we do, they limit their analysis to policies that generate positive violations by all firms, and they assume complete information about firms' abatement and enforcement costs. They obtain a uniform tax for particular industries because of their assumption of identical firms within industries; however, they also recognize that the tax rate may vary across industries in part because of differences in marginal enforcement costs and abatement costs. One could easily use their results to argue that discriminatory taxes are likely to be optimal in an industry composed of heterogeneous firms. Besides showing that optimality will often require full compliance, we extend Cremer and Gahvari's work by determining the extent to which discriminatory taxes remain optimal when regulators have incomplete information about abatement and enforcement costs.

Malik (1992) provided an early hint that policies that generate a uniform pollution price are likely to be inefficient when one accounts for enforcement costs. He models a competitive emissions trading program under complete information that is enforced to achieve full compliance, and demonstrates that emission trading leads to a distribution of emission control that does not minimize the sum of aggregate abatement and enforcement costs. An important distinction between our work and Malik's is that he is concerned with the optimal distribution of emissions while we are concerned with the optimal distribution of emissions prices. The two approaches are clearly complementary, but our pricing approach illuminates what we believe is the fundamental reason for the sub-optimality of emissions trading that Malik identifies: a competitive emissions trading policy leads to a uniform price, while enforcement costs typically call for discriminatory prices.

The most important implication of our discriminatory-pricing result is also rather obvious. When regulators have sufficient information to set discriminatory taxes, any policy that sets or generates a single price for pollution cannot be efficient. These policies include standard designs involving uniform emissions taxes and competitive emissions trading. But it is the single pollution price that drives much of our understanding of incentive-based control and that leads to the most important reason for designing and implementing these policies. That reason is the ability of these policies to induce a distribution of individual emissions control that minimizes the aggregate abatement costs of achieving some aggregate emissions target, even when the target cannot be guaranteed because of incomplete information. A single pollution price motivates firms to choose emissions so that their marginal abatement costs are equal in equilibrium, and this forms the set of necessary conditions for minimizing aggregate abatement costs. Clearly, when it is optimal to set discriminatory prices, firms' marginal abatement costs will differ, and aggregate abatement costs will not be minimized. Thus, the main justification for implementing policies that price pollution is not valid when discriminatory prices are optimal.

Moreover, discriminatory pricing greatly complicates the regulatory choice between emissions taxes and emissions trading under uncertainty that began with Weitzman's (1974) seminal work. The canonical analysis of price-based versus quantity-based emissions control features a uniform tax versus a competitive emissions trading program that produces a uniform permit price. Even those who have suggested combining taxes and tradable permits maintain this approach. For example, the models of Roberts and Spence (1976) and Kwerel (1977) produce a single equilibrium price. Montero (2002) is the only work that compares taxes and tradable permits when enforcement is costly, but he does not notice that enforcement costs may cause discriminatory pricing. Our results about the optimality of discriminatory pollution prices warrant a reexamination of the choice between emissions taxes and emissions trading, as well as efforts to combine these instruments.[15]

We recognize, however, that even when regulators have sufficient information to impose discriminatory taxes, other considerations may limit their ability to do so. Legal prohibitions against discriminatory taxation may prevent regulators from implementing an optimal policy. Even if discriminatory taxes are lawful, they may not be politically feasible. In all likelihood, some firms will perceive discriminatory taxes as unfair and lobby against their use. We recognize as well that differentiated taxes can produce moral hazard problems. In our model a firm faces a lower tax than another if the regulator's expectation of the marginal monitoring costs associated with inducing a lower level of emissions is higher. Thus, as in Heyes (1994), firms may invest in reducing their "monitorability" by, for example, acquiring production or abatement technologies that make emissions monitoring difficult. In Heyes's model, firms seek to weaken the productivity of regulator's enforcement activities. In our model, firms may be motivated to do so to lower their tax rate. This deserves further investigation as well.

[15] The first task in this effort would be to modify conventional emissions trading schemes to produce the correct firm-specific prices. These modifications may involve permit trading ratios among firms that differ from the one-for-one trading schemes that characterize conventional emissions trading. Another option might be to combine one-for-one permit trading with firm-specific taxes to account for differences in the optimal pollution prices.

Appendix: Proof of Proposition 2

A policy of the form $p_i^c = \left[t_i^c, \pi_i^c \phi_i (q_i - r_i)\right]$ motivates the firm to report its true level of emissions, which is ensured by $t_i^c = \pi_i^c \phi_i$. Under this policy the firm chooses its emissions so that $-C_q\left(q_i, x_i, \varepsilon_i\right) = t_i^c$, from which the regulator calculates $q\left(t_i^c, x_i, \varepsilon_i\right)$ and forms its expectation of the firm's emissions $E\left(q\left(t_i^c, x_i, \varepsilon_i\right)\right) = \bar{q}\left(p_i^c\right)$. Substitute this expectation into $-C_q\left(q_i, x_i, \varepsilon_i\right) = t_i^c$ to obtain the identity

$$-E\left(C_q\left(\bar{q}\left(p_i^c\right), x_i, \varepsilon_i\right)\right) \equiv t_i^c. \qquad \text{[A.1]}$$

Now let us construct the policy $p_i^q = \left[t_i^q, \pi_i(q_i, r_i) f(q_i - r_i)\right]$ that induces the same level of expected emissions as under p_i^c. The firm's expected costs under p_i^q are

$$C(q_i, x_i, \varepsilon_i) + t_i^q r_i + \pi_i(q_i, r_i) f_i(q_i - r_i). \qquad \text{[A.2]}$$

Assuming that the firm chooses positive emissions, an emissions report, and a non-negative violation, *i*'s optimal choices of these values are determined by:

$$C_q(q_i, x_i, \varepsilon_i) + \left(\partial \pi_i / \partial q_i\right) f_i(q_i - r_i) + \pi_i(q_i, r_i) f_i'(q_i - r_i) - \lambda_i = 0; \qquad \text{[A.3]}$$

$$t_i^q + \left(\partial \pi_i / \partial r_i\right) f_i(q_i - r_i) - \pi_i(q_i, r_i) f_i'(q_i - r_i) + \lambda_i = 0; \qquad \text{[A.4]}$$

$$-(q_i - r_i) \le 0,\ \lambda_i \ge 0,\ \lambda_i(q_i - r_i) = 0, \qquad \text{[A.5]}$$

where, λ_i is the Lagrange multiplier attached to the constraint $q_i - r_i \ge 0$, and $\partial \pi_i / \partial q_i$ and $\partial \pi_i / \partial r_i$ are the constant marginal effects of the firm's emissions and report on $\pi_i(q_i, r_i)$.

The regulator uses [A.3]—[A.5] to calculate $q\left(p_i^q, x_i, \varepsilon_i\right)$ and $r\left(p_i^q, x_i, \varepsilon_i\right)$, and its expectations of the firm's emissions and emissions report, $E\left(q\left(p_i^q, x_i, \varepsilon_i\right)\right) = \bar{q}\left(p_i^q\right)$ and $E\left(r\left(p_i^q, x_i, \varepsilon_i\right)\right) = \bar{r}\left(p_i^q\right)$. Substitute $\bar{q}\left(p_i^q\right)$ and $\bar{r}\left(p_i^q\right)$ into [A.3] and [A.4] to obtain the identities:

$$\begin{aligned} &E\left(C_q\left(\bar{q}\left(p_i^q\right), x_i, \varepsilon_i\right)\right) + \left(\partial \pi_i / \partial q_i\right) f_i\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right) \\ &\quad + \pi_i\left(\bar{q}\left(p_i^q\right), \bar{r}\left(p_i^q\right)\right) f_i'\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right) - \lambda_i \equiv 0; \end{aligned} \qquad \text{[A.6]}$$

$$t_i^q + \left(\partial \pi_i / \partial r_i\right) f_i\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right)$$
$$- \pi_i\left(\bar{q}\left(p_i^q\right), \bar{r}\left(p_i^q\right)\right) f_i'\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right) + \lambda_i \equiv 0; \qquad [A.7]$$

Now combine [A.6] and [A.7] to obtain

$$-E\left(C_q\left(\bar{q}\left(p_i^q\right), x_i, \varepsilon_i\right)\right) \equiv t_i^q + \left(\partial \pi_i / \partial q_i + \partial \pi_i / \partial r_i\right) f_i\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right). \qquad [A.8]$$

If p_i^q is to induce the same expected level of emissions as p_i^c, then p_i^q is constructed so that $\bar{q}\left(p_i^c\right) = \bar{q}\left(p_i^q\right)$. Obviously this implies $-E\left(C_q\left(\bar{q}\left(p_i^q\right), x_i, \varepsilon_i\right)\right) = -E\left(C_q\left(\bar{q}\left(p_i^c\right), x_i, \varepsilon_i\right)\right)$. Thus, equating [A.8] and [A.1] yields

$$t_i^q = t_i^c - \left(\partial \pi_i / \partial q_i + \partial \pi_i / \partial r_i\right) f_i\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right). \qquad [A.9]$$

We are now ready to compare the detection probability under p_i^c to the expected detection probability under p_i^q. To do so, substitute $t_i^c = \pi_i^c \phi_i$ into [A.9] and substitute the result into [A.7] to obtain

$$\pi_i^c \phi_i - \pi_i\left(\bar{q}\left(p_i^q\right), \bar{r}\left(p_i^q\right)\right) f_i'\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right) = \left(\partial \pi_i / \partial q_i\right) f_i\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right) - \lambda_i \qquad [A.10]$$

So that the difference between $\pi_i\left(\bar{q}\left(p_i^q\right), \bar{r}\left(p_i^q\right)\right)$ and π_i^c does not depend on the difference in marginal penalties, substitute $\phi_i = f_i'\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right)$ into [A.10] and rearrange terms to obtain

$$\pi_i^c - \pi_i\left(\bar{q}\left(p_i^q\right), \bar{r}\left(p_i^q\right)\right) = \frac{\left(\partial \pi_i / \partial q_i\right) f_i\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right) - \lambda_i}{\phi_i}. \qquad [A.11]$$

Since $\lambda_i \geq 0$ and $\lambda_i = 0$ when $\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right) > 0$, [A.11] indicates that $\pi_i\left(\bar{q}\left(p_i^q\right), \bar{r}\left(p_i^q\right)\right) < \pi_i^c$ if and only if $\left(\partial \pi_i / \partial q_i\right) f_i\left(\bar{q}\left(p_i^q\right) - \bar{r}\left(p_i^q\right)\right) > 0$. Clearly, this requires that the regulator expects the firm to be noncompliant and that the probability of detection under p_i^q is a strictly increasing function of the firm's emissions.

There remains the possibility that the detection probability may be lower under the policy $p_i^{\bar{q}} = \left[t_i^{\bar{q}}, \pi_i\left(\bar{q}\left(p_i^{\bar{q}}\right), r_i\right) f_i(q_i - r_i)\right]$ than under $p_i^c = \left[t_i^c, \pi_i^c \phi_i\right]$. Under $p_i^{\bar{q}}$ the regulator calculates $\bar{q}\left(p_i^{\bar{q}}\right)$ and commits to $\pi_i\left(\bar{q}\left(p_i^{\bar{q}}\right), r_i\right)$ before the firm makes its choices. Consequently, the firm treats $\overline{q}\left(p_i^{\overline{q}}\right)$ as a constant, implying that $\partial\pi_i/\partial q_i = 0$. Proceeding in exactly the same way as above, it is straightforward to show that the detection probability under p_i^c and the expected detection probability under $p_i^{\overline{q}}$ are related by $\pi_i^c - \pi_i\left(\overline{q}\left(p_i^{\overline{q}}\right),\overline{r}\left(p_i^{\overline{q}}\right)\right) = -\lambda_i/\phi_i$. Since $\lambda_i \geq 0$, $\pi_i^c \leq \pi_i\left(\overline{q}\left(p_i^{\overline{q}}\right),\overline{r}\left(p_i^{\overline{q}}\right)\right)$, which indicates that the probability of detection under $p_i^{\overline{q}}$ cannot be less than the probability of detection under p_i^c. This completes the proof of Proposition 2. QED.

References

Arguedas, Carmen. 2007. "To Comply or Not To Comply? Pollution Standard Setting under Costly Monitoring and Sanctioning".” Forthcoming in *Environmental and Resource Economics*.

Boemare, Catherine and Phillipe Quirion. 2002. “Implementing Greenhouse Gas Trading in Europe: Lessons from Economic Theory and International Experiences.” *Ecological Economics* 43: 213-230.

Bebchuk, Lucian A. and Louis Kaplow. 1993. “Optimal Sanctions and Differences in Individual’s Likelihood of Avoiding Detection.” *International Review of Law and Economics* 13, 217-224.

Bontems, Phillipe and Jean-Marc Bourgeon. 2005. “Optimal Environmental Taxation and Enforcement Policy.” *European Economic Review* 49(2), 409-435.

Chavez, Carlos A. and John K. Stranlund. 2003. “Enforcing Transferable Permit Systems in the Presence of Market Power.” *Environmental and Resource Economics* 25 (1): 65-78.

Cremer, Helmuth and Firouz Gahvari. 2002. **“**Imperfect Observability of Emissions and Second-Best Emission and Output Taxes.” *Journal of Public Economics* 85, 385–407.

Franckx, Laurent. 2002. “The Use of Ambient Inspections in Environmental Monitoring and Enforcement When the Inspection Agency Cannot Commit Itself to Announced Inspection Probabilities.” *Journal of Environmental Economics and Management* 43(1): 71–92.

Grieson, Ronald E., and Nirvikar Singh. 1990. “Regulating Externalities Through Testing.” *Journal of Public Economics* 41(3): 369–387.

Harford, Jon. 1987. “Self-Reporting of Pollution and the Firm’s Behavior Under Imperfectly Enforceable Regulations.” *Journal of Environmental Economics and Management* 14: 293-303.

Harford, Jon. 1978. “Firm Behavior Under Imperfectly Enforceable Pollution Standards and Taxes.” *Journal of Environmental Economics and Management* 5: 26-43.

Heyes, Anthony. 1994. “Environmental Enforcement when “Inspectability” is Endogenous. A Model with Overshooting Properties” *Environmental and Resource Economics* 4(5): 479-494.

Keeler, Andrew. 1991. “Noncompliant Firms in Transferable Discharge Permit Markets: Some Extensions.” *Journal of Environmental Economics and Management* 21: 180-189.

Kwerel, Evan. 1977. “To Tell the Truth: Imperfect Information and Optimal Pollution Control,” *The Review of Economic Studies* 44(3): 595-601.

Lewis, Tracy. 1996. “Protecting the Environment When Costs and Benefits are Privately Known,” *Rand Journal of Economics* 27 (4), 819-847.

Macho-Stadler, Ines and David Perez-Castrillo. 2006. “Optimal Enforcement Policy and Firm’s Emissions and Compliance with Environmental Taxes.” *Journal of Environmental Economics and Management* 51(1), 110-131.

Malik, Arun S. 1993. “Self Reporting and the Design of Policies for Regulating Stochastic Pollution.” *Journal of Environmental Economics and Management* 24: 241-257.

Malik, Arun S. 1992. “Enforcement Cost and the Choice of Policy Instruments for Controlling Pollution.” *Economic Inquiry* 30: 714-721.

Malik, Arun S. 1990. “Markets for Pollution Control when Firms are Noncompliant.” *Journal of Environmental Economics and Management* 18: 97-106.

Montero, Juan-Pablo. 2002. “Prices versus Quantities with Incomplete Enforcement.” *Journal of Public Economics* 85: 435-454.

Polinsky, A. Mitchell and Steven Shavell. 1992. "Enforcement Costs and the Optimal Magnitude and Probability of Fines." *Journal of Law and Economics* 35(1): 133-148.
Roberts, Marc J. and Michael Spence. 1976. "Effluent Charges and Licenses under Uncertainty," *Journal of Public Economics* 5(3-4): 193-208.
Sandmo, Agnar. 2002. "Efficient Environmental Policy with Imperfect Compliance." *Environmental and Resource Economics* 23(1): 85-103.
Stranlund, John K. and Kanwalroop K. Dhanda. 1999. "Endogenous Monitoring and Enforcement of a Transferable Emissions Permit System." *Journal of Environmental Economics and Management* 38(3): 267-282.
Stranlund, John K. and Carlos A. Chavez. 2000. "Effective Enforcement of a Transferable Emissions Permit System with a Self-Reporting Requirement." *Journal of Regulatory Economics* 18(2): 113-131.
Stranlund, John K. 2007. "The Regulatory Choice of Noncompliance in Emissions Trading Programs." *Environmental and Resource Economics* 38(1), 99-117.
vanEgteren, Henry and Mariane Weber. 1996. "Marketable Permits, Market Power, and Cheating." *Journal of Environmental Economics and Management* 30: 161-173.
Weitzman, Martin L. 1974. "Prices vs. Quantities," *Review of Economic Studies* 41(4): 477-91.
Xepapadeas, Anastasios. 1997. *Advanced Principles of Environmental Policy*. Edward Elgar Publishing, Northampton, Massachusetts.
Zylics, Tomasz and Joanna Spyrka. 1994. "Poland" in *Use of Economic Instruments in Environmental Policy in Central and Eastern Europe. Case studies of Bulgaria, the Czech Republic, Hungary, Poland, Romania, the Slovak Republic, and Slovenia*. Edited by Jürg Klarer. Available at: www. Rec. org/REC/Publication/EcInstrumens/Poland.html