University of Massachusetts Amherst
Department of Resource Economics
Working Paper No. 2004-4

Optimal Noncompliance in Emissions Trading Programs

John K. Stranlund[1]

Abstract:

This paper characterizes optimal noncompliance for an emissions trading policy that seeks to achieve a given aggregate emissions target at least cost. The total costs of achieving the target consist of aggregate abatement costs, monitoring costs, and the expected costs of collecting penalties from noncompliant firms. Optimal noncompliance depends in large measure on whether the marginal penalty for individual violations is increasing or constant. The primary result of this work is that any policy that achieves an aggregate emissions target with a linearly increasing marginal penalty, regardless of whether it involves perfect or imperfect compliance, is more costly than an alternative policy that induces perfect compliance with a constant marginal penalty.

Keywords: Noncompliance, Emissions Trading, Environmental Regulation, Enforcement

JEL Classification: K42, L51, Q50

[1] John K. Stranlund
Department of Resource Economics
214 Stockbridge Hall
80 Campus Center Way
University of Massachusetts
Amherst, MA 01003
E: stranlund@resecon.umass.edu P: 413-545-6328 F: 413-545-5853

March 2004

Optimal Noncompliance in Emissions Trading Programs

John K. Stranlund*
Department of Resource Economics
University of Massachusetts-Amherst

* Send correspondences to John K. Stranlund, Department of Resource Economics, 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.

Optimal Noncompliance in Emissions Trading Programs

Abstract: This paper characterizes optimal noncompliance for an emissions trading policy that seeks to achieve a given aggregate emissions target at least cost. The total costs of achieving the target consist of aggregate abatement costs, monitoring costs, and the expected costs of collecting penalties from noncompliant firms. Optimal noncompliance depends in large measure on whether the marginal penalty for individual violations is increasing or constant. The primary result of this work is that *any* policy that achieves an aggregate emissions target with a linearly increasing marginal penalty, regardless of whether it involves perfect or imperfect compliance, is more costly than an alternative policy that induces perfect compliance with a constant marginal penalty. *JEL Codes*: *L51, Q28.*

1. Introduction

By exploiting the power of a market to allocate pollution control responsibilities, well-designed emissions trading programs promise to achieve environmental quality goals more cheaply than traditional command-and-control regulations. It is obvious though that the full potential of emissions trading cannot materialize if these programs are not enforced well. In recognition of this fact, a sizable theoretical literature exists that examines the consequences of noncompliance and the design of enforcement strategies for emissions trading programs. There is, however, a significant omission in this literature—there is no published work that examines what the optimal level of noncompliance should be for emissions trading programs.

Authors of papers in this literature often address questions of compliance behavior or the design of enforcement strategies under the assumption that enforcement is not, or cannot be sufficient to induce perfect compliance [Malik (1990, 2002), Keeler (1991), van Egteren and Weber (1996), Stranlund and Dhanda (1999), Montero (2002)]. Others restrict their analysis to full-compliance outcomes [Malik (1992), Stranlund and Chavez (2000), Chavez and Stranlund (2003)]. In practice we find examples of emissions trading programs with significant noncompliance as well as examples with near-perfect compliance. Montero, Sanchez, and Katz (2002) argue that the development of an emissions trading program for total suspended particulates in Santiago, Chile has been hampered by weak enforcement and significant noncompliance. On the other hand, Stranlund, Chavez, and Field (2002) contend that the enforcement strategies of the Sulfur Dioxide Allowance Trading program and the RECLAIM program of southern California were clearly designed to achieve very high rates of compliance, and have been largely successful in doing so.

This paper addresses a fundamental policy question: To achieve a fixed aggregate emissions target cost-effectively, should emissions trading programs be designed and implemented to achieve perfect compliance, or does allowing a certain amount of noncompliance reduce the costs of reaching the emissions target? It is assumed that the policy instruments available to a regulator are the aggregate supply of emission permits and monitoring to check individual firms for noncompliance. Unlike much of the literature on optimal law enforcement (see Polinsky and Shavell (2000) for a review), this work is not concerned with choosing optimal penalty 'levels'. However, it turns out that optimal noncompliance depends in large measure on the structure of penalties for individual violations. The combination of optimal noncompliance and the structure of penalty schedules provides several new results:

- With a given linearly increasing marginal penalty schedule, a simple parametric condition involving the relative costs of monitoring and collecting penalties determines whether the optimal level of noncompliance is zero or positive. The fundamental trade-off in this setting is between light monitoring and allowing a certain amount of noncompliance to conserve on monitoring costs, and more intense monitoring to induce perfect compliance to conserve on the expected costs of collecting penalties.
- With a given constant marginal penalty the optimal amount of noncompliance is zero. This is true because allowing noncompliance cannot reduce the amount of monitoring necessary to achieve the aggregate emissions standard. Inducing perfect compliance to eliminate the expected costs of collecting penalties minimizes the costs of achieving the emissions standard.
- For any policy that achieves an aggregate emissions target with a linearly increasing marginal penalty, regardless of whether it involves perfect or imperfect compliance, there is an alternative policy involving perfect compliance and a constant marginal penalty that achieves the emissions target at lower cost.[1]

---

[1] There is no work in the literature that compares the efficiency properties of alternative penalty schedules. However, Keeler (1991) provides a positive comparison of emissions trading to

This last result is the primary contribution of this work because it suggests a strong policy recommendation—to achieve an aggregate emissions target at least cost, an emissions trading program should include a constant marginal penalty, the supply of permits should be set equal to the emissions target, and monitoring should be sufficient to induce perfect compliance.

The optimality of perfect compliance may run counter to one's intuition, and some may be surprised by this conclusion. Certainly, most papers in the literature on compliance and enforcement of emissions trading programs have focused on imperfect compliance, and one finds several assumptions in this literature that are made to preclude perfect compliance. One common assumption is that monitoring is so costly that perfect compliance is not socially optimal. Indeed, the results of this paper indicate that this may be true with an increasing marginal penalty. However, the dominance of perfect compliance and a constant marginal penalty revealed in this paper does not depend at all on the costs of monitoring. A related assumption that limits analyses to imperfect compliance is that penalties cannot be set high enough to achieve perfect compliance.[2] Although no upper bound is placed on penalties in this analysis, the cost-effectiveness of perfect compliance and a constant marginal penalty does not depend on the freedom to choose an arbitrarily high marginal penalty. Finally, other authors assume that enforcement resources are simply insufficient to induce full compliance.[3] While this is certainly true in many real instances of environmental policy enforcement, the main result of this paper suggests that in designing an emissions trading program, regulators should allocate sufficient enforcement resources to achieve perfect compliance.

---

emissions standards under exogenous enforcement strategies that involve increasing, constant, and decreasing marginal penalties. He finds that the structure of marginal penalty functions plays an important role in determining whether aggregate emissions and aggregate noncompliance are higher or lower under an emissions trading program. In contrast, this paper is concerned with deriving endogenous enforcement strategies and the determination of whether marginal penalties should be increasing or constant. Decreasing marginal penalties are not considered in this paper.

[2] For just one example among several, Montero (2002) assumes that monitoring is too expensive for perfect compliance to be optimal and that penalties are restricted to some maximum level. Obviously, if penalties were not restricted, then perfect compliance could be achieved with only minimal monitoring (Becker (1968)). Thus, if one assumes that monitoring is so costly that perfect compliance is never optimal, this assumption must be accompanied by a suitably limited penalty.

[3] For example, Stranlund and Dhanda (1999) examine the choice of enforcement strategy by a budget-constrained enforcer who does not have sufficient resources to induce perfect compliance.

The optimality of a constant marginal penalty may also be surprising, given that such a penalty is not common in the literature on compliance and enforcement of emissions trading programs. Interestingly, constant marginal penalties appear to be much more common for actual and proposed emissions trading programs than in the literature.[4]

Since the conclusions of this paper runs contrary to the most common assumptions made in the literature about compliance levels and the structure of penalties, they should cause analysts in this area to more carefully evaluate their assumptions about the sort of penalty they are using and whether their focus should be on perfect or imperfect compliance.

The rest of the paper proceeds as follows. Sections 2 and 3 focus on the consequences of employing a linearly increasing marginal penalty. Section 2 characterizes equilibrium violations, given an enforcement strategy and permit supply, while section 3 characterizes optimal noncompliance (i.e., optimal permit supply) and monitoring, given an increasing marginal penalty schedule. The analysis turns to a constant marginal penalty in Section 4, where the result that perfect compliance is optimal with such a penalty is proved. Section 5 establishes the cost-effectiveness of a constant marginal penalty and perfect compliance over any policy involving an increasing marginal penalty. A discussion of this result is contained in section 6, and section 7 concludes.

2. Equilibrium Noncompliance with an Increasing Marginal Penalty

The analysis begins with a characterization of equilibrium violations and emissions in an emissions trading program, given a linearly increasing marginal penalty function, a fixed monitoring probability, and a fixed supply of permits. Doing so sets the stage for the next section, which provides the analysis of optimal noncompliance given a linearly increasing

[4] See Boemare and Quirion (2002) for examples. There is quite a lot of variation in how actual constant marginal penalties are set. The SO_2 Allowance program employs a fixed (in real terms) financial penalty. The EPA's recent Clear Skies proposal calls for a unit penalty that is three times the clearing price in the most recent auction of permits (US EPA 2003a). Many policies employ an offset penalty whereby a firm's excess emissions in one period are deducted from its allocation of permits in the next period. The SO_2 and Clear Skies programs include a one-to-one offset to complement the financial penalties of these programs. The US EPA's Ozone Transport Commission NOx Budget Program employs a 3-to-1 offset as its primary penalty for noncompliance.

marginal penalty. Most of the results of this section can be found elsewhere in the literature, so some of them are presented without proofs, while others are proved in the appendix.

Throughout consider a fixed set of n heterogeneous, risk-neutral firms. The abatement cost function for firm i is $c_i(e_i)$, which is strictly decreasing and convex in the firm's emissions e_i. The firm is allocated l_i^0 emissions permits initially and chooses to hold l_i permits. Each permit confers the legal right to release one unit of emissions. Assume competitive behavior in the permit market so that all permit trades take place at a constant price p. Permits cannot be banked; hence, the analysis throughout is static

A regulator has perfect knowledge of each firm's permit holding, but cannot observe emissions without a costly audit. Let π denote the probability that the regulator audits each firm. An audit perfectly reveals a firms emissions, and hence, its compliance behavior. The monitoring probability is common knowledge, and the regulator commits itself to it at the outset. The assumption that the monitoring probability is the same for all firms will be justified shortly. If a firm is noncompliant, its emissions exceed the number of permit it holds and its violation is $v_i = e_i - l_i > 0$. If a firm is compliant, $e_i - l_i \leq 0$ and $v_i = 0$. Violations are penalized according to a quadratic penalty function, $\phi v_i + \gamma v_i^2/2$, where $\phi > 0$ and $\gamma > 0$. When the analysis turns to a constant marginal penalty schedule in a later section, γ will be set equal to zero.

Assume that the intercept of the marginal penalty schedule, ϕ, is greater than the equilibrium price of permits. This assumption allows full compliance to be a possible outcome throughout the paper.[5] No upper bound on marginal penalties is imposed, but none of the results of this paper rely on setting arbitrarily high penalties.

Assuming throughout that each firm chooses positive emissions, a firm's objective is

$$\begin{aligned} &\min_{(e_i, l_i)} c_i(e_i) + p(l_i - l_i^0) + \pi\left(\phi(e_i - l_i) + \gamma(e_i - l_i)^2/2\right) \\ &\text{subject to } e_i - l_i \geq 0,\ l_i \geq 0. \end{aligned} \qquad [1]$$

[5] Noncompliance in the SO_2 Allowance trading program is penalized with a constant penalty that has always been many times higher than going allowance prices. The penalty was set at \$2,000 per ton of emissions in excess of allowances in 1990 dollars, while allowance prices have rarely risen above \$200. The penalty for the 2002 compliance year was about \$2,850 per ton of excess emissions, while permits traded at prices that were generally less than \$170 per allowance (US EPA 2003b). Analysts and policymakers alike stress the importance of making sure that marginal penalties exceed the price of permits (US EPA 2003c).

Restricting the firm to $e_i - l_i \geq 0$ follows from the fact that a firm will never have the incentive to be over-compliant.[6] The Lagrange equation for [1] is

$$\Lambda = c_i(e_i) + p(l_i - l_i^0) + \pi\left(\phi(e_i - l_i) + \gamma(e_i - l_i)^2/2\right) - \beta_i(e_i - l_i),$$

and the first-order conditions are:

$$\Lambda_e = c_i'(e_i) + \pi[\phi + \gamma(e_i - l_i)] - \beta_i = 0; \qquad [2]$$

$$\Lambda_l = p - \pi[\phi + \gamma(e_i - l_i)] + \beta_i \geq 0,\ l_i \geq 0,\ \Lambda_l l_i = 0; \qquad [3]$$

$$\Lambda_\beta = -(e_i - l_i) \leq 0,\ \beta_i \geq 0,\ \beta_i(e_i - l_i) = 0.^7 \qquad [4]$$

The following Lemma characterizes firms' compliance choices. It is proved in the appendix.

***Lemma* 1:** If $p < \pi\phi$, then $v_i = 0$. If $p \geq \pi\phi$, then $v_i = (p - \pi\phi)/\pi\gamma \geq 0.$ Furthermore, regardless of the relationship between p and $\pi\phi$, individual violations are uniform across firms.

The lemma reveals that each firm is compliant if and only if the permit price is not greater than the intercept of the expected marginal penalty schedule (i.e., $p \leq \pi\phi$). For $p \geq \pi\phi$, each firm's violation is $v_i = (p - \pi\phi)/\pi\gamma.$ Since all the variables that determine violation levels are constants that do not vary across firms, individual violations are all the same. Stranlund and Dhanda (1999) have argued that differences in the size of individual violations of risk-neutral firms that trade emissions permits competitively should be independent of differences in their abatement costs. This result further suggests that if penalties are applied uniformly, there is no reason for an enforcer to consider monitoring some firms more closely than others. This result justifies our assumption that monitoring is uniform across firms.[8]

---

[6] If $e_i < l_i$, then the firm could reduce its abatement costs by allowing its emissions to increase to l_i without incurring any costs.

[7] Because the firm's objective is strictly convex and the constraint $e_i - l_i \geq 0$ is linear, these conditions are necessary and sufficient to identify optimal choices.

[8] Uniform monitoring is not justified when firms face emissions standards. Garvie and Keeler (1994) show that firms with higher marginal abatement costs should be monitored more

Combining equations [2] and [3] yields $c_i'(e_i) + p = 0$, which uniquely determines a firm's choice of emissions. Since each firm chooses its emissions so that its marginal abatement cost is equal to the going permit price, the permit market will equalize firms' marginal abatement costs. Consequently, whatever level of aggregate emissions results in equilibrium, the aggregate abatement costs of reaching that level of emissions are minimized. Furthermore, in equilibrium the permit price is equal to aggregate marginal abatement costs at the resulting level of aggregate emissions. These results are contained in the following lemma. Since the results are well known, the lemma is offered without proof.

***Lemma* 2:** Each firm chooses its emissions so that $c_i'(e_i) + p = 0$. Consequently, in equilibrium, $p = -C'(E)$, where E is aggregate emissions and

$$C(E) = \min_{\{e_i\}} \sum_{i=1}^{n} c_i(e_i) \quad \text{s.t.} \quad \sum_{i=1}^{n} e_i = E. \qquad [5]$$

Furthermore, p and E are inversely related by $E_p = -1/C''(E) < 0$.[9]

Now suppose that a fixed number L permits are put into circulation. Let the equilibrium price given full compliance by all firms be $p(L)$, and note by Lemma 2 that $p(L) = -C'(L)$. With the possibility of noncompliance the equilibrium permit price is $\overline{p}$, and equilibrium aggregate emissions are $\overline{E} = E \mid \overline{p} = -C'(E)$. Proposition 1 fully characterizes the equilibrium price and aggregate emissions as functions of the supply of permits, the monitoring probability, and the parameters of the marginal penalty function. It is proved in the appendix.

---

closely, because their incentives for noncompliance are greater than firms with lower marginal abatement costs.

[9] $C''(E) > 0$ follows from the strict convexity of individual marginal abatement costs. The assumption of a perfectly competitive permit market is critical for this analysis, because it implies the minimization of aggregate abatement costs, and firms' marginal compliance incentives are equal in equilibrium. These conclusions do not hold in the presence of market imperfections like market power and transaction costs. See van Egteren and Weber (1996), Malik (2002), and Chavez and Stranlund (2003) for analyses of compliance and enforcement of emissions trading programs in the presence of market power. Chavez and Stranlund (2004) analyze compliance and enforcement in the presence of transaction costs.

***Proposition* 1:** Given an aggregate supply of permits, a monitoring probability, and a linearly increasing marginal penalty function, the equilibrium permit price and aggregate emissions must satisfy:

***i*)** $\overline{p} \leq p(L)$ and $\overline{E} \geq L$.

***ii*)** If $p(L) \leq \pi\phi$, then $\overline{p} = p(L)$ and $\overline{E} = L$.

***iii*)** If $p(L) \geq \pi\phi$, then $p(L) \geq \overline{p} \geq \pi\phi$. Furthermore, $\overline{p}$ and $\overline{E}$ are uniquely determined by

$$\overline{p} = -C'(\overline{E}); \qquad [6a]$$

$$\overline{p} = \pi\left[\phi + \gamma(\overline{E} - L)/n\right]. \qquad [6b]$$

Part *i*) of Proposition 1 says that the equilibrium permit price cannot exceed the full compliance price, and equilibrium emissions cannot fall below the aggregate supply of permits. This result follows because $\overline{p} > p(L)$ would imply an excess supply of permits. Parts *ii*) and *iii*) reveal that the relationship between the full compliance price and the intercept of the expected marginal penalty function, $\pi\phi$, determines whether there will be noncompliance in equilibrium. Indeed, there will complete compliance if and only if $p(L) \leq \pi\phi$. In this case, the permit price will be equal to the full compliance price.

Part *iii*) of Proposition 1 says that when the full compliance price is not less than the intercept of the expected marginal penalty function, the equilibrium price and aggregate emissions are uniquely determined by [6a] and [6b]. Since individual violations are equal across firms (Lemma 1), they are all equal to the average violation, $(\overline{E} - L)/n$. Therefore, the right-hand side of [6b] is the equilibrium expected marginal penalty for each firm. Graphically, the equilibrium permit price, aggregate emissions, and aggregate violations are determined by the intersection of the aggregate marginal abatement cost function and the uniform expected marginal penalty function.

3. Optimal Noncompliance with an Increasing Marginal Penalty

Having characterized equilibrium noncompliance given a permit supply, monitoring, and a linearly increasing marginal penalty function, we can now turn to the determination of the optimal level of noncompliance. Now suppose that $\overline{E}$ is an aggregate emissions target. The policy objective is to choose the supply of permits and monitoring to minimize the total costs of achieving the target, given a linearly increasing marginal penalty. The total costs of achieving the emissions target consist of aggregate abatement costs and enforcement costs. Enforcement involves two costs, monitoring costs and the administrative and related costs of collecting penalties.

We should note here that an optimal monitoring strategy will never involve $p(L) < \pi\phi$. In this case, all firms are compliant so there are no costs of collecting penalties. Furthermore, monitoring costs could be reduced without affecting the firms' compliance or emissions decisions by simply lowering π so that $p(L) = \pi\phi$. Since the determination of the optimal level noncompliance is naturally restricted by $p(L) \geq \pi\phi$, only part *iii*) of Proposition 1 is relevant for the following analysis.

From [6b] in Proposition 1, achieving $\overline{E}$ requires $\overline{p} = \pi\left[\phi + \gamma(\overline{E} - L)/n\right]$. Rearranging this provides the level of monitoring that is required to maintain the target, given a supply of permits:

$$\pi = \overline{p}\Big/\left[\phi + \gamma(\overline{E} - L)/n\right]. \qquad [7]$$

Since $\overline{p} = -C'(\overline{E})$, $\overline{p}$ is fixed because $\overline{E}$ is now fixed. Note that choosing the appropriate level of monitoring requires complete information about the aggregate marginal abatement cost function. Without this information it will be difficult to guarantee that the emissions target is met. However, a demonstration of how this problem can be overcome is provided at the end of this section.

To minimize enforcement costs it is analytically more convenient to choose individual violations than the supply of permits. Since these violations are uniform across firms, each firm's violation is $v = (\overline{E} - L)/n$, and the required monitoring probability is $\pi = \overline{p}/(\phi + \gamma v)$. Suppose that $\overline{n} \leq n$ firms are monitored at random so that $\pi = \overline{n}/n$. Let μ be the cost of monitoring a firm so that total monitoring costs are $\mu\overline{n} = \mu n\pi$. Substitute the required monitoring probability to obtain total monitoring costs in terms of individual violations:

$$M(v) = \mu n\overline{p}/(\phi + \gamma v). \qquad [8]$$

Note that $M(v)$ is decreasing in v. Therefore, if penalties were collected without cost, maximizing noncompliance would minimize total enforcement costs. This is not a very helpful result, because it literally means setting $L = 0$ and eliminating the permit market altogether.

Now turn to the costs of penalizing firms. Let λ be the per-dollar cost of collecting penalties from noncompliant firms. Since the expected penalty for every firm is $\pi(\phi v + \gamma v^2/2)$, expected aggregate costs of collecting penalties are $\lambda n\pi(\phi v + \gamma v^2/2)$. Upon substitution of the required level of monitoring, $\pi = \overline{p}/(\phi + \gamma v)$, expected penalty costs are

$$P(v) = \lambda n\overline{p}(\phi v + \gamma v^2/2)\big/(\phi + \gamma v)\,. \qquad [9]$$

It is easy to show that $P(v)$ is increasing in v. Therefore, noncompliance has countervailing effects on expected enforcement costs—monitoring costs decrease with noncompliance while expected penalty costs increase. This suggests that the optimal level of noncompliance balances the incentive to make noncompliance large to reduce monitoring costs against the incentive to make noncompliance small to reduce the expected costs of collecting penalties.

***Proposition* 2**: Given an aggregate emissions target $\overline{E}$, and penalty parameters ϕ and γ, both of which are positive constants, the optimal level of noncompliance is zero if and only if $\lambda\phi^2 \geq \gamma\mu$.

Proof: Since the permit market minimizes aggregate abatement costs (Lemma 2), we only need to focus on minimizing the enforcement costs of reaching the target, which are $TE(v) = M(v) + P(v)$. Use [8] and [9] to calculate

$$\begin{aligned} TE_v(v) &= \frac{-\gamma\mu n\overline{p}}{(\phi + \gamma v)^2} + \frac{\lambda n\overline{p}}{(\phi + \gamma v)^2}\left(\phi^2 + \phi\gamma v + (\gamma v)^2/2\right) \\ &= \frac{n\overline{p}}{(\phi + \gamma v)^2}\left(-\gamma\mu + \lambda\phi^2 + \lambda(\phi\gamma v + (\gamma v)^2/2)\right). \end{aligned}$$

If $\lambda\phi^2 > \gamma\mu$, then $TE_v(v) > 0$ for $v \geq 0$. Therefore, $TE(v)$ is minimized by choosing $v = 0$. If $\lambda\phi^2 = \gamma\mu$, then $TE_v(v) > 0$ for $v > 0$ and $TE_v(v) = 0$ for $v = 0$, indicating that $TE(v)$ is again minimized by choosing $v = 0$. However, if $\lambda\phi^2 < \gamma\mu$, then

$$TE_v(v=0)=\frac{n\overline{p}}{\phi^2}(-\gamma\mu+\lambda\phi^2)<0,$$

which implies that some $v > 0$ minimizes $TE(v)$. QED.

Proposition 2 reveals that with an increasing marginal penalty function, whether the optimal level of noncompliance is zero or positive depends on the costs of collecting penalties, monitoring costs and the parameters of the penalty function; that is on the relationship between $\lambda\phi^2$ and $\gamma\mu.$ If the per-dollar cost of collecting penalties, $\lambda,$ is relatively high, or the marginal cost of monitoring, μ, is low, then the optimal level of noncompliance is set to zero to avoid collecting penalties. Furthermore, it is easy to show that given a supply of permits, a monitoring probability, and holding aggregate emissions to $\overline{E}$, increasing the intercept of the marginal penalty schedule, $\phi,$ requires reducing the slope parameter, $\gamma.$ Doing so implies that total expected penalties would be higher. Thus, higher ϕ and lower γ are associated with higher penalty costs, thereby increasing the likelihood that the optimal level of noncompliance is zero.

To illustrate Proposition 2, consider two polices such that:

P1 $L^1 < \overline{E}$ and $\pi^1=\overline{p}\big/\left[\phi+\gamma(\overline{E}-L^1)/n\right];$

P2 $L^2 = \overline{E}$ and $\pi^2=\overline{p}/\phi.$[10] [10]

Both of these policies achieve the aggregate emissions target $\overline{E}$ as illustrated in Figure 1. P1 involves a lower amount of monitoring (note that $\pi^1<\pi^2$) and a positive amount of noncompliance. On the other hand, P2 involves more intense monitoring and zero noncompliance. If P1 is optimal, then monitoring costs are high relative to penalty costs, and optimality requires relatively light monitoring and a certain amount of noncompliance. In contrast, if P2 is optimal, then penalty costs are high relative to monitoring costs and monitoring should be intense enough to induce full compliance. Note that monitoring under P2 is higher because the equilibrium marginal penalty under this policy is simply the intercept of the marginal penalty function. Under P1 the equilibrium marginal penalty is further up the marginal penalty schedule.

---

[10] Note that since $\pi\le 1,$ P2 requires $\phi\ge\overline{p}.$

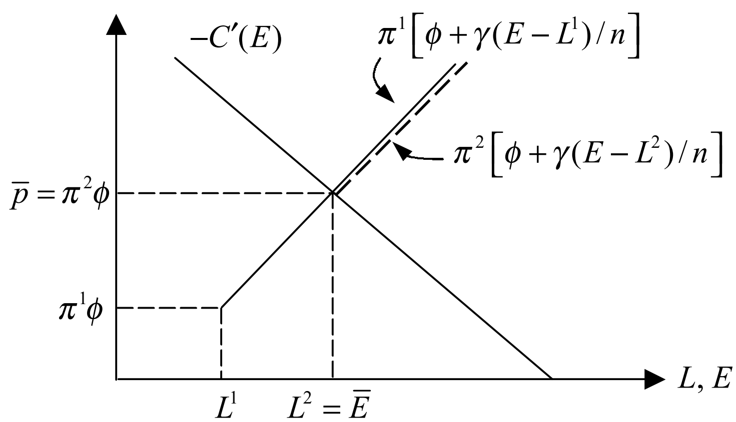

Figure 1: Optimal noncompliance with an increasing marginal penalty.

Before the analysis turns to a constant marginal penalty in the next section, it is worth making a couple of points now that apply to the entire work. First, as long as regulators have complete information about the aggregate marginal abatement cost function, all of the results of this paper hold if the emissions target is chosen optimally to minimize the sum of environmental damage, abatement costs, and expected enforcement costs. This is true because monitoring and noncompliance can always be chosen to minimize the expected costs of reaching some emissions target. Optimal aggregate emissions can then be determined, given cost-minimizing monitoring and noncompliance. In the absence of complete information about abatement costs, the results about optimal noncompliance in this paper may not hold if aggregate emissions are chosen optimally. This point is discussed further in section 6.

The other point worth making here concerns the one made earlier about the difficulty of meeting the emissions target when there is uncertainty about the aggregate marginal abatement cost function. This difficulty is easily overcome if the marginal penalty function is constructed to adjust with the equilibrium permit price.[11] Consider tying the marginal penalty to the permit price by letting $\phi = fp$ and $\gamma = gp$, where f and g are positive constants. Recall from [6b] that the equilibrium price and aggregate emissions are determined in part by $\overline{p} = \pi[\phi + \gamma(\overline{E} - L)/n]$.

[11] This is in line with how penalties are to be determined in the EPA's Clear Skies Initiative, which calls for a unit penalty that is three times the clearing price in the most recent auction of permits (US EPA 2003a).

With $\phi = fp$ and $\gamma = gp$, in equilibrium we have $\phi = f\overline{p}$, $\gamma = g\overline{p}$, and $\overline{p} = \pi\overline{p}[f + g(\overline{E} - L)/n]$. Note how the marginal penalty moves with the equilibrium permit price. Now, given that $\overline{E}$ is the aggregate emissions target, monitoring with probability $\pi = [f + g(\overline{E} - L)/n]^{-1}$ will guarantee that the target is achieved. Since the monitoring probability is now independent of $\overline{p} = -C'(\overline{E})$, the target will be achieved without complete knowledge of the aggregate marginal abatement cost function. This approach works to guarantee the target, because the marginal penalty adjusts to aggregate marginal abatement costs through the equilibrium permit price.

4. Perfect Compliance is Optimal with a Constant Marginal Penalty

While allowing imperfect compliance can be a way to reduce the costs of achieving an environmental target when an increasing marginal penalty is employed, this is not so with a constant marginal penalty. This section demonstrates that perfect compliance is optimal with a constant marginal penalty.

Take the penalty function, $\phi v_i + \gamma v_i^2/2$, and set $\gamma = 0$ so that the marginal penalty is the constant ϕ. Toward characterizing an equilibrium with this penalty, modify the first-order conditions for a firm's choices of emissions and permit demand, equations [2-4]:

$$\Lambda_e = c_i'(e_i) + \pi\phi - \beta_i = 0; \qquad [10]$$
$$\Lambda_l = p - \pi\phi + \beta_i \geq 0,\ l_i \geq 0,\ \Lambda_l l_i = 0; \qquad [11]$$
$$\Lambda_\beta = -(e_i - l_i) \leq 0,\ \beta_i \geq 0,\ \beta_i(e_i - l_i) = 0. \qquad [12]$$

Equation [11] reveals immediately that the equilibrium permit price cannot exceed the expected marginal penalty. If it did, then $p > \pi\phi$ implies $p - \pi\phi + \beta^i > 0$, and $l_i = 0$ for every firm. Since aggregate permit demand would then be zero, the permit market would not clear. Given the equilibrium requirement that $p \leq \pi\phi$, the following lemma characterizes a firm's choices when it faces a constant marginal penalty. It is proved in the appendix.

***Lemma* 3:** With a constant marginal penalty, a firm's permit demand and emissions are characterized by the following:

$$\text{If } \begin{cases} p = \pi\phi \\ p < \pi\phi \end{cases}, \text{ then } \begin{cases} l_i \in [0, e_i] & \text{and } e_i = e_i \mid c_i'(e_i) + p = 0; \\ l_i = e_i & \text{and } e_i = e_i \mid c_i'(e_i) + p = 0. \end{cases} \qquad [13]$$

Note that each firm chooses it emissions so that its marginal abatement cost is equal to the permit price. As in Lemma 2 of section 2, the permit market will minimize aggregate abatement costs, and aggregate emissions are determined from the equality of the price and the aggregate marginal abatement cost function; i.e., $p = -C'(E)$. Equation [13] also reveals that when $p = \pi\phi$, a firm's permit demand is indeterminate, which implies that individual violations will be indeterminate as well.[12] However, since aggregate emissions will be determined uniquely, aggregate violations will also be unique.

As before, let L be the number of permits in circulation and let $p(L) = -C'(L)$ be the full compliance price. The equilibrium permit price with the possibility of noncompliance is $\overline{p}$, and aggregate emissions are $\overline{E} = E \mid \overline{p} = -C'(E)$. Proposition 3 fully characterizes the equilibrium permit price and equilibrium emissions when firms face a constant marginal penalty. It is proved in the appendix

***Proposition* 3:** Given an aggregate supply of permits, a monitoring probability, and a constant marginal penalty function, the equilibrium permit price and aggregate emissions must satisfy:

i) $\overline{p} \leq p(L)$ and $\overline{E} \geq L$.

ii) If $p(L) \leq \pi\phi$, then $\overline{p} = p(L)$ and $\overline{E} = L$.

iii) If $p(L) \geq \pi\phi$, then $\overline{p} = \pi\phi = -C'(\overline{E})$ and $\overline{E} \geq L$.

[12] Recall Stranlund and Dhanda's (1999) result that differences in individual violations of risk-neutral firms that trade emissions permits competitively are independent of differences in their abatement costs. This is true with a constant marginal penalty as well, because the choice of violation only involves a comparison of the permit price and the constant expected marginal penalty. Nothing having to do with a firm's characteristics goes into this choice. However, the more specific result that individual violations are uniform with an increasing marginal penalty does not hold in case of a constant marginal penalty.

There will be noncompliance in equilibrium if and only if the marginal expected penalty is less than the full compliance price. In this case, as well as in the case in which the marginal expected penalty is equal to the full compliance price, the equilibrium permit price is exactly equal to the expected marginal penalty.

Toward determining optimal noncompliance, again let $\overline{E}$ be an aggregate emissions target that is to be achieved at minimum cost by choosing monitoring and the supply of permits, given a constant marginal penalty ϕ. As in the case of an increasing marginal penalty, it is only necessary to minimize monitoring costs plus the expected costs of collecting penalties, because the permit market minimizes aggregate abatement costs.

Using Proposition 3, the minimum amount of monitoring necessary to maintain $\overline{E}$ is $\pi = \overline{p}/\phi$, where $\overline{p}$ is now a constant equal to $-C'(\overline{E})$.[13] Total monitoring costs are $\mu n\pi = \mu n\overline{p}/\phi$. In contrast to the case of an increasing marginal penalty in which monitoring costs are decreasing in the amount of noncompliance, monitoring costs with a constant marginal penalty are independent of the amount of noncompliance. Therefore, to minimize expected enforcement costs we only need to minimize the expected costs of collecting penalties. Given the monitoring required to reach the target, this is accomplished simply by setting the supply of permits equal to the aggregate emissions target to eliminate noncompliance and the costs of collecting penalties.[14] This proves the following proposition.

***Proposition* 4:** Given an aggregate emissions target $\overline{E}$, and a constant marginal penalty ϕ, the optimal level of noncompliance is zero.[15]

[13] Clearly, $\pi = \overline{p}/\phi$ requires $\phi \geq \overline{p}$.

[14] To be specific about expected penalty costs, note that expected aggregate penalties are $\pi\phi(\overline{E} - L)$, for $L \leq \overline{E}$. The expected costs of collecting these penalties are $\lambda\pi\phi(\overline{E} - L)$. Substitute the required monitoring, $\pi = \overline{p}/\phi$, to obtain $\lambda\overline{p}(\overline{E} - L)$, which is monotonically decreasing in the supply of permits.

[15] As in the case of an increasing marginal penalty, setting the correct level of monitoring to induce $\overline{E}$ with a fixed constant marginal penalty requires knowledge of the aggregate marginal abatement cost function. However, as before, tying the marginal penalty to the prevailing permit price eliminates this problem. Setting $L = \overline{E}$, $\phi = fp \geq p$, and monitoring with probability $\pi = 1/f$ will guarantee the achievement of $\overline{E}$ at least cost without complete information about abatement costs.

It is interesting to note that the optimality of full compliance when a constant marginal penalty is employed does not depend at all on the costs of monitoring.[16] This is true because achieving the emissions target requires a fixed amount of monitoring. In the case of an increasing marginal penalty we identified a tradeoff between light monitoring and noncompliance to conserve monitoring costs and more intense monitoring and zero noncompliance to conserve penalty costs. No such tradeoff exists when a constant marginal penalty is employed—the only way to minimize enforcement costs in this case is to eliminate the expected costs of collecting penalties by guaranteeing perfect compliance.

5. The Optimality of a Constant Marginal Penalty and Perfect Compliance.

Armed with characterizations of optimal noncompliance for linearly increasing and constant marginal penalties, we are now ready to establish the dominance of a constant marginal penalty and perfect compliance in achieving an aggregate emissions target at least cost. Toward that end, return to the policies involving a given linearly increasing marginal penalty specified by [10] and illustrated in Figure 1. Recall that one policy, P1, involves relatively light monitoring and a positive amount of noncompliance, while the other, P2, eliminates noncompliance with more intense monitoring. For the purposes of this section, these two policies are arbitrary in the sense that there is no requirement that one or the other is optimal, although this possibility is not precluded either. A third policy is constructed that combines elements of the other two, but that employs a constant marginal penalty and induces perfect compliance. The following proposition reveals that this policy achieves the emissions target at lower cost than the other two.

***Proposition* 5:** Consider two permit supply/monitoring policies that achieve the aggregate emissions target $\overline{E}$ with the same linearly increasing marginal penalty function. The first policy involves relatively light monitoring and some noncompliance:

P1 $\quad L^1 < \overline{E}$ and $\pi^1 = \overline{p} \Big/ \left[\phi + \gamma(\overline{E} - L^1)/n \right]$.

[16] If the emissions target was chosen optimally as well, monitoring costs would affect this choice—high monitoring costs would imply higher aggregate emissions. However, the result that noncompliance should be zero would still hold, at least as long as there is no uncertainty about aggregate abatement costs.

The second policy involves more intense monitoring and zero noncompliance:

P2 $L^2 = \overline{E}$ and $\pi^2 = \overline{p}/\phi$.

Consider a third policy with the following characteristics:

P3 $L^3 = \overline{E},\ \pi^3 = \pi^1,\ \phi^3 = \phi + \gamma(\overline{E} - L^1)/n,$ and $\gamma^3 = 0,$

This policy employs the monitoring of P1, a constant marginal penalty equal to the equilibrium marginal penalty of P1, and induces perfect compliance like P2. This policy achieves the aggregate emissions target with lower expected costs than both P1 and P2.

***Proof*:** It is necessary at the outset to prove that P3 achieves the aggregate emissions target. Note for P1 that

$$\overline{p} = \pi^1\left[\phi + \gamma(\overline{E} - L^1)/n\right]. \qquad [14]$$

From Proposition 3, the equilibrium price for P3 is $\hat{p} = \pi^3\phi^3$. Upon substitution of $\pi^3 = \pi^1$ and $\phi^3 = \phi + \gamma(\overline{E} - L^1)/n$, as specified by P3, we have $\hat{p} = \pi^1\left[\phi + \gamma(\overline{E} - L^1)/n\right]$. Since the right-hand-side of this equation is identical to the right-hand-side of [14] we have $\hat{p} = \overline{p}$, which indicates that P3 also achieves the aggregate emissions target.

Since all three policies achieve the same aggregate emissions target at minimum aggregate abatement costs, comparing the total costs of the policies simply involves comparing expected enforcement costs. In comparing P3 and P1, note first that since monitoring is the same under both policies their monitoring costs are equal. However, since P1 involves some noncompliance while noncompliance is zero under P3, expected penalty costs are positive under P1 and zero under P3. Since aggregate abatement costs and monitoring costs are the same under both policies, but P1 incurs penalty costs and P3 does not, P3 achieves the aggregate emissions target with lower expected costs than P1.

Now compare P3 and P2. Both policies induce perfect compliance so expected penalty costs are zero under both policies. However, P2 requires higher monitoring than P3. Since aggregate abatement costs and expected penalty costs are the same under both polices, but P3 has lower monitoring costs, P3 achieves the aggregate emissions target with lower costs than P2. QED.

Note carefully that the dominance of a constant marginal penalty and complete compliance does not depend on having complete freedom in the choice of the penalty. Of course, it would have been a trivial matter to demonstrate this dominance with an arbitrarily high marginal penalty. Instead, the proposition specifies that the constant marginal penalty, ϕ^3, be set equal to the equilibrium marginal penalty under policy P1. This marginal penalty is higher than the equilibrium marginal penalty under policy P2. However, ϕ^3 is an *available* marginal penalty under P2—it just isn't reached because of the elimination of noncompliance under this policy. It should be clear that the restriction placed on the constant marginal penalty in Proposition 5 is only meant to demonstrate the result with a relatively conservative penalty. Of course, choosing a higher constant marginal penalty would allow less intense monitoring and lower enforcement costs.

Somewhat loosely, Proposition 5 reveals that *any* policy that achieves an aggregate emissions target with a linearly increasing marginal penalty can be improved upon by a policy with a constant marginal penalty and perfect compliance. The policy recommendation of this paper is therefore quite clear: To achieve an aggregate emissions target at least cost, an emissions trading program should include a constant marginal penalty that is set as high as is feasible, a supply of permits equal to the emissions target, and sufficient monitoring to guarantee perfect compliance.

6. Discussion

The cost-effectiveness of perfect compliance and a constant marginal penalty can be easily incorporated into the choice of the optimal environmental target, at least when there is complete information about abatement costs. Suppose that emissions are uniformly mixed and cause damage that is characterized by an increasing and convex damage function $D(E)$. The first-best environmental target is chosen to equate marginal damage to aggregate marginal abatement costs. If this is to be achieved with a competitive permit program, the equilibrium permit price is equal to marginal damage at the first-best level of emissions. That is, the permit market generates the optimal Pigovian price on emissions.

With the necessity of enforcing an emissions trading program, the efficient environmental target is no longer first-best. Suppose that a constant marginal penalty is employed. Proposition 4 then implies that whatever level of aggregate emissions is chosen, it will be optimal to issue that number of permits and monitor firms to guarantee perfect compliance. The monitoring probability is $\pi = p/\phi$, where $p = -C'(E)$. If the per-firm cost of monitoring is μ, total monitoring costs are $\mu n\pi = -\mu nC'(E)/\phi$. The efficient level of emissions then minimizes the social cost function $D(E) + C(E) - \mu nC'(E)/\phi$. Assuming that this is strictly convex, the efficient level of emissions is the solution to $D'(E) + C'(E) = \mu nC''(E)/\phi$. Since this implies $D'(E) > -C'(E)$ at the efficient outcome, the efficient level of emissions is greater than the first-best level. Moreover, the price of emissions, $p = -C'(E)$, is less than the optimal price. Less control than first-best is efficient, because the efficient choice of emissions internalizes the enforcement costs of inducing this level of emissions. Finally it is easy to show that social costs are decreasing in the marginal penalty, ϕ, yielding the conventional result that this penalty should be set as high as possible.

Two problems arise with uncertainty about abatement costs. The first problem has been mentioned and dealt with: namely, that it will be difficult to achieve a fixed emissions target with a fixed penalty schedule. It has been shown, however, that tying marginal penalties to going permit prices can eliminate this problem. With a marginal penalty function that moves with the permit price, if aggregate marginal abatement costs turn out to be higher or lower than expected, this will be reflected in the equilibrium permit price and the marginal penalty will adjust accordingly.

It is not so straightforward to deal with incomplete information if the emissions target is chosen optimally. Furthermore, the main result of this paper about the optimality of perfect compliance and a constant marginal penalty may not hold in this setting. With incomplete information about abatement costs, optimality may actually call for allowing noncompliance and employing an increasing marginal penalty. Roberts and Spence (1976) have proposed a modification of an emissions trading program to include a “safety valve” tax. This tax reduces the welfare loss of a permit-trading program when aggregate marginal abatement costs are underestimated, because it allows firms to pay a price to escape the burden of unexpectedly high abatement costs. The expected marginal penalty could be constructed to serve this purpose, and could be chosen to allow noncompliance if marginal abatement costs turn out to be higher than expected. Furthermore, an increasing marginal penalty schedule may be preferred to a constant marginal penalty. The main benefit of Roberts and Spence's mixed permit/tax policy is that it can be constructed to approximate the marginal damage schedule. If marginal damages are increasing, then an increasing expected marginal penalty could provide a better approximation to marginal damage than a constant expected marginal penalty.

Montero (2002) has provided an interesting paper that links the insights of Roberts and Spence to imperfect compliance. He reexamined Weitzman's (1974) comparison of price (emissions tax) and quantity instruments (tradable permits) under uncertainty and imperfect compliance, and found that the potential advantage of an emissions tax over a transferable permit system is significantly reduced by incomplete enforcement. The main reason is that under incomplete enforcement, a transferable permit system resembles the mixed transferable permit/tax scheme of Roberts and Spence. However, Montero assumes that perfect compliance is not possible, which, of course, runs contrary to the results of this paper. A potentially fruitful investigation for the future would reconsider the results of this paper when there is incomplete information about abatement costs, and then use the results to reexamine the prices versus quantity debate.

Finally, we should recognize that the way monitoring is modeled in this paper, although very common in the literature, includes several inherent assumptions that if relaxed might affect the results. Perhaps the most important assumption is that the monitoring technology perfectly reveals a firm's compliance status. This, of course, means that it perfectly reveals a firm's emissions. The monitoring technology that comes closest to that envisioned in this paper, as well as the rest of the literature in this area, are the continuous emissions monitoring systems that are required of all firms in the SO_2 Allowance and other emissions trading programs. This technology is evidently very accurate, yet requiring its installation imposes a significant cost on regulated firms. One may wonder, therefore, whether requiring the gold standard in monitoring technologies is optimal. Furthermore, a continuous emission monitoring technology is not available for every situation in which an emissions trading program might be adopted (e.g., relatively small, perhaps mobile, sources that may or may not have distinct points of discharge). There is wide recognition among analysts that it is important to gain a better understanding of how emissions trading programs should be designed and enforced in situations in which a direct and continuous measure of emissions is either not optimal or non-existent. Along this line of inquiry, it is also important to analyze optimal noncompliance in these settings.

7. Conclusion

We have characterized optimal noncompliance and the basic structure of penalties for an emissions trading program that seeks to achieve an aggregate emissions target at least cost. This paper offers several new results, but the most important is the cost-effectiveness of a constant marginal penalty and perfect compliance. Although this result is quite strong, it should be accompanied by certain caveats concerning incomplete information about abatement costs and the monitorability of emissions. Despite these caveats, there are several important benefits of this work. First, it provides a clear baseline to which the results of extensions into more complicated settings can be compared. Second, since its conclusions runs contrary to the most common assumptions made in the literature about compliance levels and the structure of penalties, this work should cause analysts in this area to more carefully evaluate their assumptions about the sort of penalty they are using and whether their focus should be on perfect or imperfect compliance. Finally, because it employs a conventional model of emissions trading and a reasonable policy objective to reach a strong conclusion, it provides policymakers with an important insight into the design of emissions trading programs—if the policy goal is to reach an environmental target cost-effectively, an emissions trading program should be enforced with a constant marginal penalty and should be designed to induce perfect compliance.

Appendix

***Proof* of *Lemma* 1:** Using [3], $p < \pi\phi$ implies $p - \pi[\phi + \gamma(e_i - l_i)] < 0,\ \forall l_i \in [0, e_i]$. Therefore, $\Lambda_l \geq 0$ requires $\beta_i > 0$. From [4], $\beta_i > 0$ implies $l_i = e_i$ and therefore $v_i = 0$. Note that this must be true for every firm.

Now if $p > \pi\phi$, then $v_i = (p - \pi\phi)/\pi\gamma > 0$. To establish this result suppose that $p > \pi\phi$ and $v_i = 0$. Then, $v_i = 0$ implies $l_i = e_i > 0$. Furthermore, from [3] $\Lambda_l = p - \pi\phi + \beta_i > 0$. However, $\Lambda_l > 0$ requires $l_i = 0$; hence, we have a contradiction. Note from [4] that $v_i > 0$ implies $\beta_i = 0$, and [3] becomes $\Lambda_l = p - \pi[\phi + \gamma(e_i - l_i)] = 0$. Solving for $v_i = e_i - l_i$ yields $v_i = (p - \pi\phi)/\pi\gamma > 0$. Note that if $p = \pi\phi$, then $v_i = (p - \pi\phi)/\pi\gamma = 0$. Finally, since p, π, ϕ, and γ are constants that do not vary across firms, v_i is the same for all i. QED.

***Proof of Proposition* 1:**

Part *i*) Suppose instead that $\bar{p} > p(L)$. Then since $E_p < 0$ by Lemma 2, $\bar{E} < L$. This implies however, that the constraint $e_i \geq l_i$ in [1] would be violated for at least some i. Since $\bar{p} \leq p(L)$, $E_p < 0$ implies $\bar{E} \geq L$.

Part *ii*) Toward a contradiction suppose that $p(L) \leq \pi\phi$ and $\bar{p} < p(L)$. Then, since $\bar{p} < \pi\phi$, Lemma 1 implies $v_i = 0$ for all i, and $\bar{E} = L$. However, $\bar{p} < p(L)$ implies $\bar{E} > L$, because $E_p < 0$. This contradiction establishes $\bar{p} = p(L)$ and $\bar{E} = L$.

Part *iii*) Part *i*) establishes $p(L) \geq \pi\phi$. Now, instead of $\bar{p} \geq \pi\phi$, suppose that $\bar{p} < \pi\phi$. Lemma 1 then implies $v_i = 0$ for all i, as well as $\bar{E} = L$. However, $\bar{p} < \pi\phi$ implies $\bar{p} < p(L)$, which in turn implies $\bar{E} > L$. This contradiction establishes $\bar{p} \geq \pi\phi$.

To establish [6], note first that equilibrium requires $\bar{p} = -C'(\bar{E})$ by Lemma 2. Furthermore, since $p \geq \pi\phi$, Lemma 1 implies $v_i = (\bar{p} - \pi\phi)/\pi\gamma \geq 0$. Rearrange this to obtain $\bar{p} = \pi[\phi + \gamma v_i]$. Also from Lemma 1, $v_i = v_j$ for each i and j, which implies that each firm's violation is equal to the average violation; that is, $v_i = (\bar{E} - L)/n$, for each i. Substituting $v_i = (\bar{E} - L)/n$ into $\bar{p} = \pi[\phi + \gamma v_i]$ yields $\bar{p} = \pi\left[\phi + \gamma(\bar{E} - L)/n\right]$. QED.

***Proof of Lemma* 3:** Given a choice of $e_i > 0$, if $p = \pi\phi$, then it is clear from [11] that the firm is indifferent about how many permits it will hold. If $l_i \in (0, e_i]$, then $\Lambda_l = p - \pi\phi + \beta_i = 0$, and hence, $\beta_i = 0$. If $l_i = 0$, then [12] implies $\beta_i = 0$. Substitute $\beta_i = 0$ and $p = \pi\phi$ into [10] to obtain $c_i'(e_i) + p = 0$. Now if $p < \pi\phi$, then [11] implies $\beta_i > 0$, and [12] implies $l_i = e_i$. Since $l_i > 0$, [11] implies $\Lambda_l = p - \pi\phi + \beta_i = 0$. Combine this with [10] to obtain $c_i'(e_i) + p = 0$. QED.

***Proof of Proposition* 3:**

Part *i*) Suppose instead that $\bar{p} > p(L)$. Then since $E_p < 0$, $\bar{E} < L$. This implies, however, that the constraint $e_i \geq l_i$ would be violated for at least some i. Since $\bar{p} \leq p(L)$, $E_p < 0$ implies $\bar{E} \geq L$.

Part *ii*) Toward a contradiction suppose that $p(L) \leq \pi\phi$ and $\overline{p} < p(L)$. Then, since $\overline{p} < \pi\phi$, Lemma 3 implies $v_i = 0$ for all i, and $\overline{E} = L$. However, $\overline{p} < p(L)$ implies $\overline{E} > L$, because $E_p < 0$. This contradiction establishes $\overline{p} = p(L)$, which in turn implies $\overline{E} = L$.

Part *iii*) Since $\overline{p} = -C'(\overline{E})$ and $\overline{E} \geq L$ have already been established, we only need to prove $\overline{p} = \pi\phi$. As stated in the text, [11] implies $\overline{p} \ngtr \pi\phi$. Suppose that $\overline{p} < \pi\phi$. Then, since $p(L) \geq \pi\phi$, $p(L) \geq \pi\phi > \overline{p}$. From Lemma 3, $\pi\phi > \overline{p}$ implies $e_i = l_i$, and hence, $\overline{E} = L$. However, since $\overline{p}$ and $\overline{E}$ are inversely related by $\overline{p} = -C'(\overline{E})$, $p(L) > \overline{p}$ implies $\overline{E} > L$. This contradiction establishes $\overline{p} = \pi\phi$. QED.

References

Becker, Gary. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76(2), 169-217.

Boemare, Catherine and Phillipe Quirion. 2002. "Implementing Greenhouse Gas Trading in Europe: Lessons from Economic Theory and International Experiences." Working paper 35.2202. Foudazione Eni Enrico Mattei.

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, May.

Chavez, Carlos A. and John K. Stranlund. 2004. "Enforcing Transferable Permit Systems in the Presence of Transaction Costs." Working paper, Department of Resource Economics, University of Massachusetts, Amherst

Garvie, Devon and Andrew Keeler. 1994. "Incomplete Enforcement with Endogenous Regulatory Choice." *Journal of Public Economics* 55: 141-162.

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

Malik, Arun S. 2002. "Further Results on Permit Markets with Market Power and Cheating." *Journal of Environmental Economics and Management* 44(3), 371-90.

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

Malik, Arun. 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.

Montero, Juan-Pablo, Jose Miguel Sanchez, and Ricardo Katz. 2002. A Market-Based Environmental Policy Experiment in Chile." *Journal of Law and Economics* 45 (1), 267-287.

Polinsky*,* A. Mitchell and Steven Shavell. 2000."The Economics Theory of Public Enforcement of Law." *Journal of Economic Literature* 38(1), 45-76

Roberts, M.J. and M. Spence. 1976. "Effluent Charges and Licenses Under Uncertainty," *Journal of Public Economics* 5, 193-208.

Stranlund, John 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 and Carlos 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.; Carlos A. Chavez and Barry Field. 2002. "Enforcing Emissions Trading Programs: Theory, Practice and Performance." *Policy Studies Journal* 30(3), 343-361.
U.S. Environmental Protection Agency. 2003a. *Section-by-Section Summary of the Clear Skies Act of 2003*. Washington DC: U.S. EPA Office of Air and Radiation. http://www.epa.gov/air/clearskies/legis.html.
U.S. Environmental Protection Agency. 2003b. *Acid Rain Program 2002 Compliance Report*. Washington DC: U.S. EPA Acid Rain Program.
U.S. Environmental Protection Agency. 2003c. *Tools of the Trade: A Guide to Designing and Operating a Cap and Trade Program for Pollution Control.* Washington DC: U.S. EPA Office of Air and Radiation.
van Egteren, Henry and Marian Weber. 1996. "Marketable Permits, Market Power, and Cheating." *Journal of Environmental Economics and Management* 30: 161-173.
Weitzman, Martin. 1974. "Prices vs. Quantities," *Review of Economic Studies* 41(4), 477-91.