Targeting Enforcement to Improve Compliance with Environmental Regulations

Lana Friesen*

Paper presented at AARES 2001
Adelaide, South Australia
23-25 January 2001

By targeting enforcement efforts on specific segments of the regulated community, greater compliance with environmental regulations can be achieved. In this paper, the inspection minimizing targeting scheme with two groups is derived. Firms are moved at random into the target group, while escape from the target group occurs only when an inspection reveals the firm is in compliance. The optimal targeting scheme reduces inspection costs compared with the strategy suggested by Harrington (1988), where firms are moved into the target group on the basis of compliance record. However, the range of parameter values for which the optimal solution is feasible is limited.

JEL classification: Q28, K32, K42

Keywords: law enforcement, compliance, environmental regulation

* Commerce Division, PO Box 84, Lincoln University, Canterbury, New Zealand. Phone: (643) 325-3627. Fax: (643) 325-3847. Email: friesenl@kea.lincoln.ac.nz.

1. Introduction

In the past twenty-five years, numerous laws and regulations have been passed to protect the natural environment and alleviate the effects of pollution on human health and ecosystems. To comply with these requirements, regulated firms must often incur significant expenditures. Effective monitoring and enforcement will therefore be necessary to bring about improvements in environmental quality.

Standard theory predicts that a firm will comply with a regulation when its compliance cost is less than the expected penalty associated with violation. Frequent monitoring and relatively high fines will be necessary to deter firms from violating regulations. Observation, however, suggests otherwise. Compliance is generally considered to be high, despite low inspection probabilities and small fines being imposed, if at all.[1]

Harrington (1988) provides one possible explanation for these seemingly contradictory observations. He shows how an enforcement agency can enhance deterrence by dividing regulated firms into two groups according to their past compliance record. Inspection resources are concentrated on firms in one of the groups, the target group, where surveillance is more frequent and the penalty is larger than in the non-target group. Firms inspected and found in violation are moved into the target group. Once there, being moved back into the non-target group rewards firms found in compliance. The "stick" of stricter enforcement and "carrot" for compliance combine to create stronger incentives to comply than a simple random auditing framework. As a result, a firm may comply even when their compliance cost exceeds the expected current penalty. Alternatively, the same level of deterrence can be achieved with the expenditure of fewer monitoring resources.

Subsequent papers have considered the robustness of Harrington’s results under asymmetric information (Raymond (1999)), the social optimality implications (Harford and Harrington (1991), Harford (1991)), and alternative explanations for high compliance rates such as self-reporting and enforcement power (Livernois & McKenna(1999)).

The innovation in this paper is to propose an improved transition structure for the two-group audit framework. The goal of the enforcement agency, as in

---

[1] Harrington (1988), for instance, surveys evidence that surveillance frequency is low, fines or penalties are rarely imposed, but compliance is nonetheless quite high in the United States. Livernois and McKenna (1999) provide similar evidence for Canada.

Harrington (1988), is to minimize the number of (costly) inspections required to achieve a desired compliance rate. Harrington however, assumes that targeting occurs on the basis of a firm's past compliance record, rather than solving for the optimal targeting strategy. In contrast, the optimal targeting strategy is to move firms randomly into the target group, but still maintain escape on the basis of observed compliance behaviour. This new transition design maximizes deterrence using the two-group structure and provides additional cost savings for the agency compared with the past compliance targeting framework suggested by Harrington.

A drawback of the optimal targeting scheme proposed here is its limited range of applicability. In particular, optimal targeting will be infeasible for high desired compliance rates or large compliance costs. To achieve its goal in this parameter range, the agency will need to use past compliance targeting despite incurring higher inspection costs compared with optimal targeting.

In the next section, the two-group targeting model is described along with the firm's compliance choice. Section 3 derives the agency's optimal transition structure and the feasible parameter range. Optimal targeting and Harrington's past compliance targeting scheme are compared in Section 4. Section 5 provides a discussion and concludes the paper.

2. Model Structure and Firm Choice

Consider the following dynamic game between a single firm, which is subject to an environmental regulation, and an enforcement agency.[2] The firm is assumed to have already installed the capital equipment necessary to meet the regulation ("initial compliance"), however, ongoing costs of $c per period must be incurred if the firm is to achieve "continuing compliance". The only other option available to the firm is violation. The agency is assumed to know the firm's compliance cost.

The goal of the agency is to minimize the enforcement costs associated with achieving any compliance rate, $0 \leq Z \leq 1$, where Z represents the percentage of time the firm is in compliance.[3] It is assumed that there are no costs associated with fine imposition and collection, and therefore the agency's goal is to minimize the number

---

[2] Alternatively, there could be a number of identical firms, with the compliance rate interpreted as the percentage of firms in compliance at any one time.

[3] Harrington (1988), Garvie and Keeler (1994), and Livernois and McKenna (1999) use the same regulator goal, for example.

of inspections required to achieve the compliance goal. Inspections perfectly reveal the firm's compliance status.[4] The maximum fine that can be imposed for violations of the regulation is K.

The agency adopts a strategy of targeting its limited enforcement resources by allocating the firm to one of two groups: the target group (G2), where scrutiny is high, or the non-target group (G1). The agency can choose different inspection probabilities and fines for the two groups. Let p_i be the inspection probability in each group and F_i the fine (i=1,2). The expected penalty in G2 is assumed to exceed the expected penalty in G1 or $p_2F_2 > p_1F_1$. The agency can also choose the basis of transition between the two groups. For example, the agency could condition the transition process on the result of inspections or allow random movement.

In each period, the agency's interaction with the firm will take one of three possible forms:

(i) no inspection,

(ii) an inspection that reveals the firm is in compliance, or

(iii) an inspection that reveals violation.

The agency can condition the transition probabilities between the groups on any or all of these observations. These transition parameters are given in Table 1. For example, if the agency wants to move violators directly to G2, then it will set d=0. If firms found complying in G2 are to be moved to G1 with certainty, then u=1, and so on. This framework is general enough to encompass both Harrington's past compliance targeting (a=1, b=1, d=0, w=0, y=0) and the standard static model where firms face a constant enforcement scheme over time (a=b=d=1 and/or w=u=y=0).

The firm's probability of moving between the two groups depends on its compliance decision. For example, following an inspection, the firm remains in G1 with probability *b* if in compliance or with probability *d* if in violation. If not

[4] This assumption is made to simplify the analysis. The results still hold when inspections are imperfect, so long as inspections have some power to discriminate between complying and violating firms.

inspected, the firm remains in G1 with probability *a* regardless of its compliance status. These one-period transition probabilities given in Table 2, describe a Markov process, where t^{A}_{mn} gives the firm's likelihood of moving from group *m* to group *n* when action A is chosen in the group *m*. For example, t^{C}_{21} is the probability that the firm will be moved from G2 to G1 when it complies. If a=b=d=1, the firm always remains in G1. Similarly, if w=u=y=0, the firm can never escape from G2. The agency can choose the transition structure to punish violators and reward complying firms in the following manner. When b>d, violators are punished by being sent to G2 more often than complying firms, i.e. $t^{V}_{12} > t^{C}_{12}$. When u>y, complying firms are rewarded by escaping G2 more quickly, i.e. $t^{C}_{21} > t^{V}_{21}$.

The firm chooses among four possible strategies, denoted by f_{ij}, where i describes the action in G1 and j the action in G2:

f_{cc} is the strategy of complying in both groups,

f_{cv} is compliance in G1 only,

f_{vc} is compliance in G2 only, and

f_{vv} is the strategy of violating in both groups.

Given the agency's enforcement scheme, the goal of the firm is to choose the strategy that minimizes the present value of its expected costs over the infinite horizon. Let $0<\delta<1$ be the firm's discount factor. Because a Markov process governs transitions, the optimal strategy for the firm is independent of the initial state of the system and is stationary.[5]

The expected cost, in present value terms, of following strategy f_{ij} when initially in group *m* (where m=1,2) is denoted by $E^{ij}(m)$ and can be computed by adding the expected cost discounted one period to the single period cost. Following strategy f_{cc}, for example, costs the firm, starting in G1 and G2 respectively,

$$E^{CC}(1) = c + \delta(t^{C}_{11} E^{CC}(1) + t^{C}_{12} E^{CC}(2))$$

$$E^{CC}(2) = c + \delta(t^{C}_{21} E^{CC}(1) + t^{C}_{22} E^{CC}(2)).$$

Solving these two equations yields $E^{CC}(1)$ and $E^{CC}(2)$. The expected costs for the other three strategies are summarized in Table 3. "To evaluate [strategy] f_{ij} solve the system of equations formed by taking the i^{th} equation of column 1 and the j^{th} equation

[5] Kohlas (1982) Chapter 5.

of column 2."[6] The solutions to these four sets of simultaneous equations are given in Table 4.

Except in the case of full compliance (f_{cc}), the expected cost comprises two terms. The first term represents the expected cost if the firm remains in the initial group forever. The second term is an adjustment factor that reflects the likelihood of the firm being moved to the other group. This adjustment factor will be positive if the expected costs are greater in the other group. Consider the case of a firm following strategy f_{VV}, which leads to an expected cost each period equal to the expected fine for the group to which the firm currently belongs. Suppose that the firm starts in G1, the first component is simply the infinite discounted sum of the single period costs, p_1F_1. This amount must be adjusted for the probability that the firm will end up in G2 and incur a greater expected fine. The adjustment takes account of the time the firm will spend in G2 relative to G1. Conversely, if the firm starts in G2, single-period costs in G2 must be adjusted down to reflect the likelihood of the firm being in G1 some of the time.

The firm's optimal strategy will depend on its compliance cost in the manner described in the following proposition, the proof of which is contained in the Appendix.

Proposition 1. *The optimal strategy for a firm with compliance cost c is f_{CC} if $c \leq Q_1$; f_{VC} if $Q_1 < c \leq Q_2$; and f_{VV} if $Q_2 < c$; where $Q_1 = p_1F_1$ and*

$$Q_2 = p_2F_2 + \frac{\delta(t_{21}^C - t_{21}^V)(p_2F_2 - p_1F_1)}{1 - \delta t_{11}^V + \delta t_{21}^V}.$$

When compliance costs are low, the firm chooses to comply in both groups (f_{CC}), and when compliance is costly, the firm will violate in both groups (f_{VV}). For intermediate values of *c*, the firm complies only in G2 and violates in G1 (f_{VC}). The firm's expected costs are illustrated in Figure 1, where the darkened line indicates the firm's optimal strategy.

If the agency's goal is full compliance (Z=1), targeting provides no advantage. To ensure the firm complies in both groups (f_{cc}), the expected penalty in G1 must exceed the firm's compliance cost. The required inspection probability in this case is

[6] Harrington (1988) p. 36.

identical to the simple static or one-group case. In this case, the firm need only compare the single-period returns of complying and violating in G1. Being moved to G2 is no threat to a firm that is always in compliance. The firm's likelihood of being in G1 is affected by its compliance decision in G1, however, since the firm always complies, the expected penalty in G2 is irrelevant to its decision.

Targeting is beneficial only when the agency's goal is partial compliance (Z<1).[7] An examination of the Q_2 term reveals that a firm with compliance cost greater than the expected fine in G2 can be induced to comply, at least some of the time, if $t_{21}^{C} > t_{21}^{V}$. This generates cost savings for the agency by reducing the inspection frequency required in G2 to induce compliance compared with the simple static or one-group model.

These cost savings over the simple one-group model, generated by increased deterrence, arise in two ways. First, the incentive to comply in G2 is increased when the firm has a greater probability of escape when complying as opposed to violating. If u=y, the firm has the same probability of escaping G2 whether it chooses compliance or violation. There is no additional incentive to comply in G2, and the decision comes down to a simple comparison of single-period returns. However, when u>y, complying in G2 has the additional benefit of increasing the probability of escape from G2. The gain from escape depends on both the expected fine in G1 and the amount of time the firm expects to remain in G1, which is reflected in the transition probabilities. This incentive is maximized by setting u=1 and y=0. By contrast, in Harrington (1988) this component was smaller, because while y=0, the escape probability *u* typically took on a value less than one. The second factor that affects deterrence is the differential expected penalty between the two groups. The greater this difference, the greater the reward to compliance in G2. This incentive is maximized by setting F_1=0 and F_2=K as in Harrington.

In addition, note that Q_2 is increasing in *a* and *d* since a longer time spent in G1 increases the reward to complying in G2. On the other hand, increases in *w* and *y*

[7] Enforcement agencies typically have a stated goal of achieving full compliance. However, in practise, agencies operate under quite restrictive budgets that make it necessary to allocate resources in such a way that only partial compliance is achieved for many regulations.

have the opposite effect. These terms enter through the transition probabilities in the denominator.

3. The Optimal Targeting Scheme

Consider the situation where the agency has a given compliance goal of Z<1. The agency's problem is to minimize the number of inspections required, on average, to achieve this goal. Let π_2^{VC} be the steady state probability that an f_{VC} firm is in G2. The agency's goal then is to minimize $I_{VC} = p_1 + (p_2 - p_1)\pi_2^{VC}$. The agency has ten parameters to choose: the inspection probability and fine in each group and the six transition probabilities. The parameters must be chosen to ensure that the firm *both* follows the desired f_{VC} strategy ($c \le Q_2$), *and* spends long enough in G2 to meet the compliance goal $(\pi_2^{VC} \ge Z)$. Formally, the agency's problem is:

minimize I_{VC}
$F_1, F_2, p_1, p_2, a, b, d, u, w, y$

subject to: $c \le Q_2$

$\pi_2^{VC} \ge Z$

$0 \le F_1 \le K$	$0 \le F_2 \le K$	
$0 \le p_1 \le 1$	$0 \le p_2 \le 1$	
$0 \le a \le 1$	$0 \le b \le 1$	$0 \le d \le 1$
$0 \le u \le 1$	$0 \le w \le 1$	$0 \le y \le 1$

Because inspections are costly, the compliance rate constraint $(\pi_2^{VC} \ge Z)$ will be met exactly. The parameters must be chosen to ensure that the firm spends Z% of the time in G2, on average, given the firm will only comply in G2. The steady state probability that the firm is in G2 is given by $\pi_2^{VC} = \dfrac{1 - [a(1-p_1) + p_1 d]}{1 - [a(1-p_1) + p_1 d] + w(1-p_2) + p_2 u}$. At the margin, any parameter change that increases escape from G2 (for example, an increase in *u* or *w*) must be exactly offset by a parameter change that increases the firm's chance of being sent back to G2 from G1 (for example, by decreasing *a* or *d*), so that, on average, the firm still spends the same proportion of time in each group.[8]

To ensure that the firm complies in G2, the firm choice constraint ($c \leq Q_2$) must also be met. The agency will choose the fines and G2 transition probabilities to maximize the value of Q_2 for any given inspection probabilities. Such adjustment of these non-costly parameters will lower the inspection probability required in G2 to meet the firm choice constraint and thus provide cost savings for the agency. The G1 transition probabilities will be adjusted to meet the compliance rate constraint. From the earlier discussion, deterrence will be maximized by setting u=1, y=0, w=0, $F_1=0$ and $F_2=K$. The firm must comply in order to escape from G2. Even with these choices, the agency must still inspect the firm sufficiently often in G2 to create a credible deterrent. The inspection frequency in G2 (p_2) will be chosen to meet the firm choice constraint exactly.

With the firm choice constraint met, the G1 transition parameters *a* and *d* are adjusted to ensure the Z target is met. More frequent inspection in G2 increases the incentive to comply by increasing the expected penalty *and* increasing the firm's chance of escape from G2. Accordingly, either *a* or *d* must be decreased to offset this effect. As it turns out, however, the agency will never inspect firms in G1. By randomizing the movement of firms into the targeted group the agency avoids inspection costs in G1, without affecting the firm’s incentive to comply in G2. Inspections are not needed in G1 because the firm’s compliance choice depends on the average length of stay in the two groups, and this interval is already fixed by the compliance goal Z. The firm must spend Z% of its time in G2 and the remainder in G1, on average, for the compliance goal to be met. Any change in the inspection frequency in G1 will be exactly offset by a change in the G1 transition probabilities, *a* and *d*, that keeps the length of stay in G1 constant. Costly inspections in G1 have no additional deterrent effect in G2 and therefore the optimal choice is to set the inspection frequency in G1 equal to zero.

The agency's optimal transition structure is described in Proposition 2. A formal proof can be found in the Appendix.

[8] Parameters *b* and *y* are irrelevant because the firm violates in G1 and complies in G2, and monitoring is perfect.

Proposition 2. *The optimal two-group targeting scheme is* $F_1^* = 0, F_2^* = K$, $a^* = 1 - \frac{Zp_2}{(1-Z)}$, *any value for b and d, w*=0, u*=1, y*=0,* $p_1^* = 0$, *and*

$$p_2^* = \frac{\sqrt{(1-\delta)^2(1-Z)^2K^2 + Z^2\delta^2c^2 + 2\delta Kc(1-\delta)(1-Z)(2-Z)} - ((1-\delta)(1-Z)K - Z\delta c)}{2\delta K}.$$

The optimal transition structure is illustrated in Figure 2. If the firm is in G1, it will be moved to G2 for the next period with probability (1-*a*). Remaining in G1 means not being inspected, and hence the firm violates. On the other hand, if the firm is moved to G2, inspection occurs with probability p_2 and if the firm is found in compliance it is immediately returned to G1. Demonstrating compliance is the only way for the firm to escape G2.

The optimal targeting scheme however will not always be feasible. As the firm's compliance cost increases relative to the maximum fine, more frequent inspection is required in G2 to induce compliance. For sufficiently large compliance costs even certain inspection in G2 will be inadequate to induce compliance, and the value for p_2^* given in Proposition 2 will exceed one. The range of feasible compliance costs will be stricter for large values of Z. As Z increases, the firm has to spend longer in G2 on average, if the compliance goal is to be met. However, this implies less time spent in G1, thus lowering the reward from compliance in G2. Accordingly, p_2 must be increased to induce the firm to comply in G2 and meet the Q_2 constraint. The required value for p_2 will therefore exceed one for smaller values of the compliance cost.

Specifically, the solution for p_2^* in Proposition 2 is valid only when $c \le K\Lambda$, where $\Lambda = \frac{(1-\delta)(1-Z)+\delta}{(1-\delta)(1-Z)+\delta-\delta(1-Z)} > 1$. Otherwise, $p_2^* > 1$. The term Λ represents the increase in deterrence that can be achieved using the optimal transition structure, as opposed to the standard, static, one-group model. The range for valid solutions as a function of Z is shown in Figure 3. Above the $p_2^* = 1$ locus, compliance costs are too high relative to the maximum fine for the desired compliance goal and $p_2^*>1$. The feasible parameter range lies below the $p_2^*=1$ locus.[9]

A second feasibility constraint arises with respect to the G1 transition parameter *a*, which may take on a negative value when either p_2 or Z is large. The role of *a* is to ensure that the firm spends Z% of the time in G2 for any given G2 transition parameters. If p_2 is large, as it will be when compliance costs are high, the firm leaves G2 fairly frequently, making it impossible to find an *a*>0 that meets the compliance goal. The solution for *a** is valid only when $c \leq K\Phi$, where $\Phi = \frac{(1-Z)(\delta(1-Z)+Z)}{Z^2}$. Otherwise, $a^*<0$. The dividing line between invalid and valid values of the parameter *a** is also shown in Figure 3. Above the $a^*=0$ locus, $a^*<0$ and below the locus $a^*>0$. As shown in Figure 3, the constraint on a* is only binding for Z>0.5, as for lower compliance goals a*>0 even if $p_2^*=1$.[10]

Combining these two feasibility conditions yields Proposition 3.

Proposition 3. *Optimal targeting is feasible only if* $Z \leq \frac{1}{2}$ *and* $c \leq K\Lambda$*; or* $Z > \frac{1}{2}$ *and* $c \leq K\Phi$.

For large compliance goals (Z), the agency will be unable to use optimal targeting. Harrington's (1988) past compliance targeting strategy may still be feasible, as discussed in the following section.

4. Comparison with Past Compliance Targeting

In the optimal targeting scheme, the agency randomly moves the firm *into* the target group, G2. Escape from G2 occurs only when an inspection reveals the firm is complying. In this sense, escape from G2 depends on the firm's compliance record. In contrast, Harrington (1988) considered a targeting scheme where movement *into* G2 also depends on the firm's compliance history. In particular, a firm inspected in G1 and found in violation will be moved to G2 for the following period. As in the

[9] The term Λ is decreasing in Z for δ>0. The $p_2^*=1$ locus is convex to the origin if δ>0.5, linear if δ=0.5 and concave to the origin if δ<0.5.

[10] The a*=0 locus is decreasing in Z and convex to the origin for 0<Z≤1.

optimal scheme, escape from G2 only follows discovered compliance (w=y=0); however u<1 is typical.

Past compliance targeting is more costly than optimal targeting for the agency for two reasons. Firstly, positive inspections are needed in G1 ($p_1>0$). Secondly, more frequent inspection is required in G2 due to the decrease in deterrence that results from setting u<1. However, as Figure 3 demonstrates past compliance targeting can be used for higher values of Z than for optimal targeting. The upper boundary for feasibility is shown as PC.[11] Because of decreased deterrence, this lies below the $p_2^*=1$ locus in the optimal model for Z<1. However, large values of Z are still feasible. The shaded area in Figure 3 shows the parameter range where past compliance targeting extends the range of applicability beyond that which can be achieved with optimal targeting.

5. Discussion

The key result of this paper is that adopting the optimal two-group targeting scheme can provide additional cost savings for an enforcement agency even over the scheme suggested by Harrington (1988). By randomly selecting firms for the target group the agency can save on inspections in the non-target group. The incentives for compliance are unaffected because an appropriate adjustment is made to the transition structure. In fact, deterrence is enhanced, providing further cost reductions, by allowing escape from the target group to occur whenever compliance is demonstrated.

The range of applicability of optimal targeting is however, limited, especially for "large" compliance goals. Past compliance targeting may still be feasible in this range. Ultimately, with Z=1 the agency can do no better than adopt the standard static one-group model.

The agency's compliance goal (Z) is a key determinant of the type of targeting the agency should adopt. While the stated goal of most enforcement agencies is full compliance, in practice, resource constraints require decisions to be made about where enforcement dollars will be spent, resulting in differential target compliance rates across industries and regulations. For example, one of the long-term strategic goals of

[11] The exact condition is given in the appendix.

the United States Environmental Protection Agency (EPA) is to create "a credible deterrent to pollution and greater compliance with the law: EPA will ensure *full compliance* with laws intended to protect human health and the environment."[12]. In recent years, recognizing its limited resources and the ever-increasing universe of regulated firms, the EPA has begun adopting innovative, non-traditional, approaches to achieving compliance.

> Effective compliance and enforcement is … dependent on effective targeting of the most significant public health and environmental risks. Because of this and a recognition that government resources are finite, EPA has worked since the reorganization to improve our ability to target our efforts to the areas of greatest need.[13]

This suggests, for instance, that the compliance goal Z will likely be large, even Z=1, for toxic substances, where the risk to the public and the environment is high, and smaller where the risk associated with violation is less.

Harrington's original work has been criticized on two main grounds. Firstly, as Harford and Harrington (1991) point out, the targeting scheme results in firms with identical compliance costs controlling pollution by different amounts. As a result, control costs are not minimized and social welfare may be reduced compared with the standard static model. Secondly, as demonstrated by Raymond (1999), Harrington's results may not be robust in the presence of asymmetric information and uncertainty. Both criticisms could be equally applied to this model, nevertheless, optimal targeting does provide additional cost savings over Harrington's model. In addition, as Harford and Harrington (1991, p.394) conclude "[o]nce a standard has been selected, a state-dependent enforcement strategy is the most cost-effective way to achieve a given level of compliance with the standard."

An interesting extension of the model would consider the agency's choice of the compliance goal (Z) along with the enforcement scheme. This would allow incorporation of other factors such as environmental risk and sector size, with the possibility of differential treatment based on these factors in addition to a firm's compliance record. A much broader model would consider these issues in the context of heterogenous compliance costs. Building on existing work by Polinsky & Rubinfeld (1991), the information revealing aspects of repeat offending could also be further considered.

---

[12] EPA (1997), p. 56. This is goal nine of the EPA's ten strategic goals. Emphasis mine.
[13] EPA (1999), p. 20.

Appendix

Proof of Proposition 1

Proposition 1 follows directly from the following observations, which describe the relationships among the expected costs of the four strategies. In the case of indifference, the firm is assumed to choose the policy favouring compliance.

(1) Regardless of the group a firm starts in,

$$\left.\begin{array}{l} E^{CC} < E^{VC} \text{ if } c < Q_1 \\ E^{CC} = E^{VC} \leq E^{VV} \text{ if } c = Q_1 \\ E^{CC} > E^{VC} \text{ if } c > Q_1 \end{array}\right\} \text{ where } Q_1 = p_1 F_1 .$$

(2) Regardless of the group a firm starts in,

$$\left.\begin{array}{l} E^{VC} < E^{VV} \text{ if } c < Q_2 \\ E^{VC} = E^{VV} < E^{CC} \text{ if } c = Q_2 \\ E^{VC} > E^{VV} \text{ if } c > Q_2 \end{array}\right\} \text{ where } Q_2 = p_2 F_2 + \frac{\delta(t_{21}^{C} - t_{21}^{V})(p_2 F_2 - p_1 F_1)}{1 - \delta t_{11}^{V} + \delta t_{21}^{V}} .$$

(3) The firm will never choose strategy f_{CV}. To see this, note that strategy f_{CV} is dominated by f_{CC} when $c \leq X_1$ and by f_{VV} when $c \geq X_2$, where $X_1 = p_2 F_2 + \frac{\delta(t_{21}^{C} - t_{21}^{V})(p_2 F_2 - p_1 F_1)}{1 - \delta t_{11}^{C} + \delta t_{21}^{C}}$ and $X_2 = p_1 F_1 + \frac{\delta(t_{11}^{C} - t_{11}^{V})(p_2 F_2 - p_1 F_1)}{1 - \delta t_{11}^{V} + \delta t_{21}^{V}}$. Since $X_1 > X_2$ for all values of the transition parameters, there is no compliance cost where following a strategy of f_{CV} is preferred. This result holds regardless of the group the firm starts in.

(4) $Q_2 > Q_1$. This result can be established by comparing the expressions for Q_1 and Q_2.

Proof of Proposition 2

The agency's problem is solved using the Lagrangian function, $L = I_{VC} + \lambda_1(c - Q_2) + \lambda_2(Z - \pi_2^{VC})$. The first order conditions comprise boundary conditions for each of the choice variables, plus a complementary slackness condition for each of the two inequalities. *One* of the following two boundary conditions must hold for *each* of the choice variables,

(i) $\frac{\partial L}{\partial v}(v^*) \geq 0$ and $v^* \frac{\partial L}{\partial v}(v^*) = 0$ or

(ii) $\frac{\partial L}{\partial v}(v^*) \leq 0$ and $(\bar{v} - v^*)\frac{\partial L}{\partial v}(v^*) = 0$

where v ={F_1, F_2, p_1, p_2, a, b, d, u, w, y}, $\bar{v}$ is the maximum value the variable can take, and v^* is the optimal choice. The complementary slackness conditions for the two inequalities are $\frac{\partial L}{\partial \lambda_i} \leq 0$, $\lambda_i \geq 0$, and $\frac{\partial L}{\partial \lambda_i}\lambda_i = 0$, where i=1,2.

The solution is found by proceeding in the following steps.

(1) The complementary slackness conditions imply that the two constraints are strictly binding, i.e. $c=Q_2$ and $\pi_2^{VC}=Z$. There are four cases to consider.

Case 1: $\lambda_1=0$ and $\lambda_2=0$. The agency's problem is simply to minimize I_{VC}. The solution in this case is to never inspect ($p_1=p_2=0$), however this implies $Q_2=0$. But since c>0 this gives $c>Q_2$, and the firm never complies.

Case 2: $\lambda_1>0$ and $\lambda_2=0$. The agency's problem is to minimize I_{VC} while ensuring that $c=Q_2$. With $p_2>p_1$, the optimal solution involves u=1, because this choice both reduces I_{VC} and maximizes Q_2, by making the firm escape G2 more often. Since u≥y, Q_2 is non-decreasing in the G1 transition probabilities a, b, and d. The optimal choice is a=b=d=1, since this also reduces I_{VC}. With this choice however, $\pi_2^{VC}=0$; the firm is never in G2.

Case 3: $\lambda_1=0$ and $\lambda_2>0$. The agency's problem is to minimize I_{VC} while ensuring the firm is in G2 Z% of the time. The first order condition for each variable *v* is $\frac{\partial L}{\partial v} = (p_2 - p_1 - \lambda_2)\frac{\partial \pi_2^{VC}}{\partial v}$. If $p_2-p_1>\lambda_2$, then $\frac{\partial L}{\partial v}$ and $\frac{\partial \pi_2^{VC}}{\partial v}$ have the same sign. In this case, the optimal solution is a=b=d=1, because this maximizes the firm's stay in G1 and thus minimizes I_{VC}. However, this implies that $\pi_2^{VC}=0$. On the other hand, if p_2-p_1<λ_2 , then the optimal choice is a=b=d=0 and w=u=y=0, implying that π_2^{VC}=1. The firm never escapes G2 once there (Q_2=p_2F_2). The required inspection rate is the same as in the static one-group model.

Case 4: λ_1>0 and λ_2>0. This is the only remaining possibility and is shown below to yield a valid solution for at least some parameter values.

(2) Use the constraints to eliminate choice variables, leaving an unconstrained problem to solve.

Step 1. Solving the compliance rate constraint (π_2^{VC}=Z) for *a* yields, $a = \dfrac{(1-Z)(1-p_1 d) - Zt_{21}^C}{(1-Z)(1-p_1)}$.

Step 2. Substitute for *a* from Step 1 in the firm choice constraint (c=Q_2) and rearrange to get the following expression

$$\left[p_2F_2 - c\right]\left[(1-\delta)(1-Z) + \delta w(1-p_2) + \delta p_2 u\right] = \delta p_2(y-u)(1-Z)\left[c - p_1F_1\right]. \quad \text{(A1)}$$

If the agency is to increase deterrence compared with the static model, so that the firm complies even when p_2F_2<c, it must choose u>y. If y>u, then p_1F_1>c is required, violating the assumption that the expected fine is greater in G2 than in G1.

Step 3. Solve (A1) for p_2, the inspection frequency required in G2 for the firm to choose f_{VC}.

$$p_2 = \frac{-S + \sqrt{S^2 - 4RT}}{2R}$$

where

$$R = (-w+u)\delta F_2 \quad \text{(A2)}$$

$$S = ((1-\delta)(1-Z) + \delta w)F_2 + (y-u)(1-Z)\delta p_1 F_1 - (y - w + (u-y)Z)\delta c$$

$$T = -c((1-\delta)(1-Z) + \delta w)$$

Note that the second solution to the quadratic equation yields values for p_2 that are either negative or exceed one, and are hence invalid and are not reported here. The standard quadratic formula terms “a”, “b” and “c” have been renamed “R”, “S”, and “T” in order to avoid any confusion with other parameters in the model.

(3) Solve the remaining *unconstrained* minimization problem $I_{VC} = p_1 + (p_2 - p_1)Z$, where p_2 is given in (A2).

Step 1. The agency will maximize the fine differential between the two groups, i.e. $F_1^* = 0; F_2^* = K$. To see this consider the following first order conditions.

$\frac{\partial I_{VC}}{\partial F_1} = Z\frac{\partial p_2}{\partial F_1} = Z\frac{p_2(1-Z)(u-y)\delta p_1}{\sqrt{S^2 - 4RT}} > 0$ since $u>y$ as argued above.

$$\frac{\partial I_{VC}}{\partial F_2} = Z\frac{\partial p_2}{\partial F_2} = -Z\frac{\left[c[(1-\delta)(1-Z)+\delta w(1-p_2)+\delta p_2 u]+(1-Z)(y-u)\delta p_2(c-p_1F_1)\right]}{F_2\sqrt{S^2 - 4RT}} < 0.$$

Use (A1) to show the bracketed term in the numerator is positive.

Step 2. No inspections will be conducted in G1, i.e. $p_1^* = 0$. The relevant first order condition is $\frac{\partial I_{VC}}{\partial p_1} = (1-Z) + Z\frac{\partial p_2}{\partial p_1} = (1-Z) + Z\frac{p_2(1-Z)(u-y)\delta F_1}{\sqrt{S^2 - 4RT}} > 0.$

Notice that with this choice, the parameters *b*, *d*, and F_1 become irrelevant to the firm's decision.

Step 3. The agency will maximize the differential escape probability from G2 for a complying firm over a violating firm, i.e. u*=1; y*=0. A firm can only escape G2 when it is inspected, i.e. w*=0. Making use of the results $F_1^* = 0; F_2^* = K$, $\frac{\partial p_2}{\partial y} = \frac{p_2\delta c(1-Z)}{\sqrt{S^2 - 4RT}} > 0$. Imposing the solutions found so far yields $\frac{\partial p_2}{\partial w} = \frac{-p_2\delta c(1-Z)u + (c-p_2K)((1-\delta)(1-Z)+\delta u)}{(-w+u)\sqrt{S^2 - 4RT}} > 0$. From (A1), substitute for the first term as follows $-p_2\delta c(1-Z)u = -[c-p_2K][(1-\delta)(1-Z)+\delta w(1-p_2)+\delta p_2 u]$, which allows the derivative to be written as $\frac{\partial p_2}{\partial w} = \frac{(c-p_2K)\delta(1-p_2)}{\sqrt{S^2 - 4RT}} > 0$. Making use of the solutions found so far, $\frac{\partial p_2}{\partial u} = -\frac{(c-p_2K)(1-\delta)(1-Z)}{\sqrt{S^2 - 4RT}} < 0$.

Solutions in Harrington's Past Compliance Targeting Model

Although Harrington (1988) does not provide explicit solutions for p_1, p_2 and u, they are easy to find. Once the transition structure a=1, b=1, d=0, w=0, and y=0, is imposed, only F_1, F_2, p_1, p_2 and u remain as choice variables. The fines affect only Q_2, so $F_1=0$ and $F_2=K$ is optimal. The compliance rate and firm choice constraints are solved for p_1 and u respectively, leaving I_{vc} as a function of only p_2. The following solutions result.

$$p_1 = \frac{Z(1-\delta)(c\delta(1-Z)-J)}{\delta J}, \quad p_2 = \frac{c\delta Z + J}{K\delta}, \quad \text{and} \quad u = \frac{K(1-Z)(1-\delta)(c\delta(1-Z)-J)}{(c\delta Z + J)J},$$

where $J = \sqrt{K\delta c(1-Z)^2(1-\delta)} > 0$.

When the compliance cost, c, or the desired compliance goal Z, is large, the inspection probability in G2 required for compliance will exceed one. In particular, the solution given above is valid only when $c \leq K\Gamma$, where

$$\Gamma = \frac{(1-Z)^2(1-\delta) + 2\delta Z - \sqrt{((1-Z)^2(1-\delta) + 2\delta Z)^2 - 4\delta^2 Z^2}}{2\delta Z^2} > 0. \quad \text{Otherwise } p_2 > 1.$$

This dividing line is shown in Figure 3. Note that if $\delta<0.5$, the locus is increasing in Z, but always remains below the c=K line.

A second condition for a valid solution is that $c<K(1-\delta)/\delta$, otherwise both p_1 and u take on negative values. This condition will be violated only when the maximum fine size is very large relative to the compliance cost, in which case even full compliance will require only very infrequent inspections. For example, suppose $\delta=0.992$, reflecting a monthly discount rate that is equivalent to an annual rate of 10%, then K must be more than 124 times larger than c for this condition to be violated

Bibliography

Garvie, D., Keeler, A., 1994. Incomplete enforcement with endogenous regulatory choice. Journal of Public Economics 55, 141-162.

Harford, J., 1991. Measurement error and state-dependent pollution control enforcement. Journal of Environmental Economics and Management 21, 67-81.

Harford, J., Harrington, W., 1991. A reconsideration of enforcement leverage when penalties are restricted. Journal of Public Economics 45, 391-395.

Harrington, W., 1988. Enforcement leverage when penalties are restricted. Journal of Public Economics 37, 29-53.

Kohlas, J., 1982. Stochastic methods of operations research. Cambridge University Press, New York.

Livernois, J., McKenna, C., 1999. Truth or consequences: enforcing pollution standards with self-reporting. Journal of Public Economics 71, 415-440.

Polinsky, M., Rubinfeld, D., 1991. A model of optimal fines for repeat offenders. Journal of Public Economics 46, 291-306.

Raymond, M., 1999. Enforcement leverage when penalties are restricted: a reconsideration under asymmetric information. Journal of Public Economics 73, 289-295.

United States Environmental Protection Agency, 1999. Protecting your health and the environment through innovative approaches to compliance: highlights from the past 5 years. EPA/300-K-99-001.

United States Environmental Protection Agency, 1997. EPA strategic plan. EPA/190-R-97-002.

Table 1. Transition Parameters

		Probability of being moved to	
Initial Group	Observation	G1	G2
G1	None	a	1-a
	Compliance	b	1-b
	Violation	d	1-d
G2	None	w	1-w
	Compliance	u	1-u
	Violation	y	1-y

Table 2. Transition Probabilities

	Group in Second Period (n)			
	Action Taken in Initial Group			
	COMPLY		VIOLATE	
Initial Group (m)	G1	G2	G1	G2
G1	$t^{C}_{11} = (1-p_1)a + p_1b$	$t^{C}_{12} = (1-p_1)(1-a) + p_1(1-b)$	$t^{V}_{11} = (1-p_1)a + p_1d$	$t^{V}_{12} = (1-p_1)(1-a) + p_1(1-d)$
G2	$t^{C}_{21} = (1-p_2)w + p_2u$	$t^{C}_{22} = (1-p_2)(1-w) + p_2(1-u)$	$t^{V}_{21} = (1-p_2)w + p_2y$	$t^{V}_{22} = (1-p_2)(1-w) + p_2(1-y)$

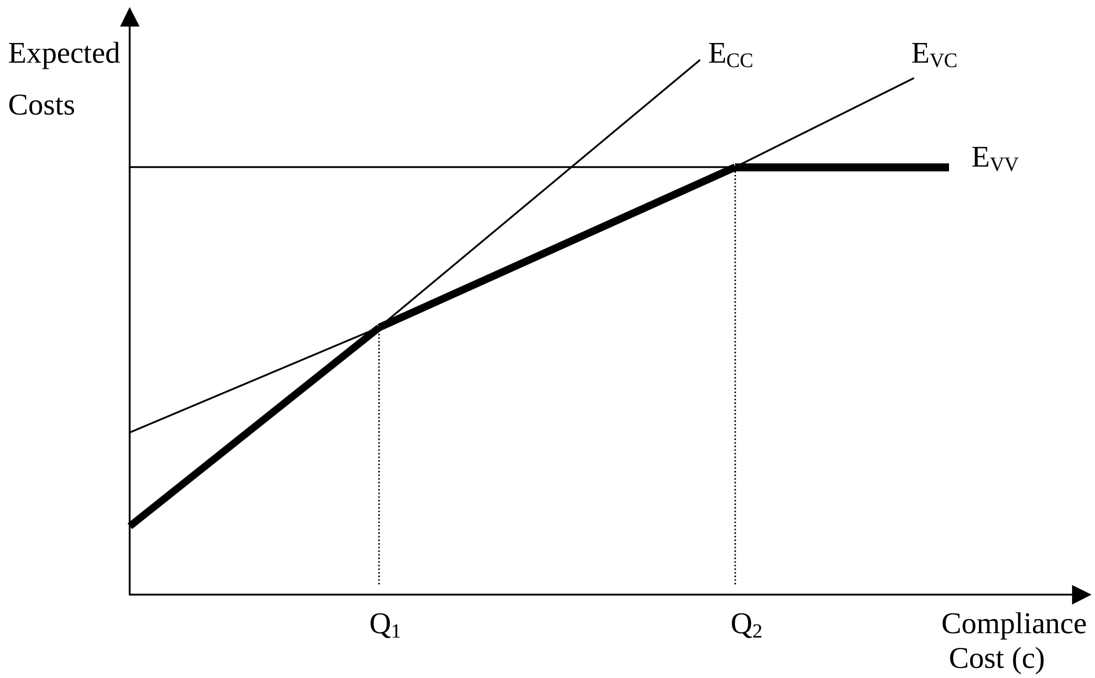
Table 3. Expected Costs for Each Strategy

Action	Cost Starting in Group 1	Cost Starting in Group 2
Comply	$E_1 = c + \delta(t_{11}^{C} E_1 + t_{12}^{C} E_2)$	$E_2 = c + \delta(t_{21}^{C} E_1 + t_{22}^{C} E_2)$
Violate	$E_1 = p_1 F_1 + \delta(t_{11}^{V} E_1 + t_{12}^{V} E_2)$	$E_2 = p_2 F_2 + \delta(t_{21}^{V} E_1 + t_{22}^{V} E_2)$

Table 4. Solutions for the Expected Costs

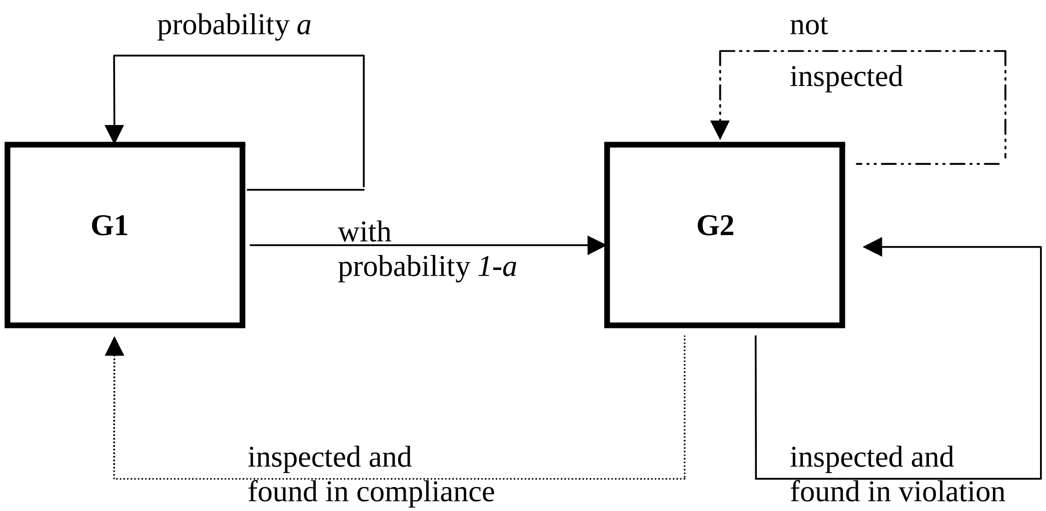
Strategy	$E^{ij}(1)$	$E^{ij}(2)$
f_{CC}	$\frac{c}{1-\delta}$	$\frac{c}{1-\delta}$
f_{VC}	$\frac{p_1F_1}{1-\delta}+\frac{\delta(1-t_{11}^1)(c-p_1F_1)}{(1-\delta)(1-\delta t_{11}^1+\delta t_{21}^0)}$	$\frac{c}{1-\delta}-\frac{\delta t_{21}^0(c-p_1F_1)}{(1-\delta)(1-\delta t_{11}^1+\delta t_{21}^0)}$
f_{VV}	$\frac{p_1F_1}{1-\delta}+\frac{\delta(1-t_{11}^1)(p_2F_2-p_1F_1)}{(1-\delta)(1-\delta t_{11}^1+\delta t_{21}^1)}$	$\frac{p_2F_2}{1-\delta}-\frac{\delta t_{21}^1(p_2F_2-p_1F_1)}{(1-\delta)(1-\delta t_{11}^1+\delta t_{21}^1)}$
f_{CV}	$\frac{c}{1-\delta}+\frac{\delta(1-t_{11}^0)(p_2F_2-c)}{(1-\delta)(1-\delta t_{11}^0+\delta t_{21}^1)}$	$\frac{p_2F_2}{1-\delta}-\frac{\delta t_{21}^1(p_2F_2-c)}{(1-\delta)(1-\delta t_{11}^0+\delta t_{21}^1)}$

Figure 1. The Firm's Optimal Strategy



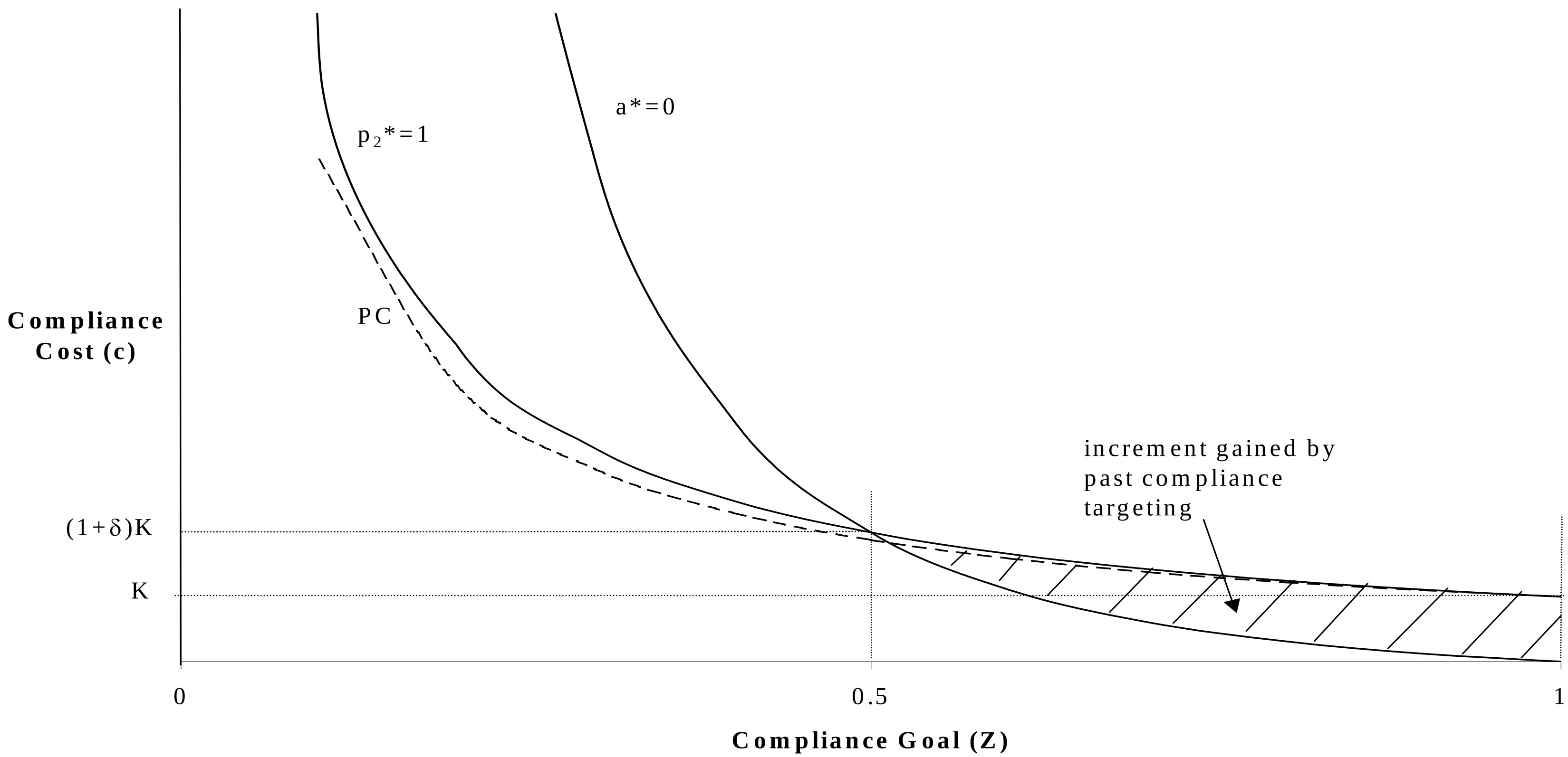
When compliance costs are low, the firm chooses to comply in both groups (f_{CC}), and when compliance is costly, the firm will violate in both groups (f_{VV}). For intermediate values of *c*, the firm complies in G2 and violates in G1 (f_{VC}).

Figure 2. The Optimal Transition Structure



Starting in G1, the firm is moved to G2 with probability 1-a. The only means of escape from G2 is for the firm to be found in compliance as a result of an agency inspection.

Figure 3. Valid Parameter Range



Optimal targeting is feasible only for compliance costs that lie below *both* the p_2*=1 locus and the a*=0 locus. Past compliance targeting is feasible anywhere below the PC line. The shaded area shows parameter values where past compliance targeting can increase the range of feasibility.