[redacted]

THE SOCIAL COST OF UNIFORM REGULATORY STANDARDS

Carol Adaire Jones*
and
Suzanne Scotchmer**

July 27, 1988

Abstract: Uniform regulatory standards, frequently employed in environmental health and safety, are widely criticized as inefficient on the grounds that all firms are required to comply, regardless of compliance cost. Since firms will self-select to comply only if their compliance costs exceed the expected penalty for non-compliance, the inefficiency *could* be avoided by an enforcement policy chosen to maximize social welfare. But we argue that the enforcement agency goal is likely to place a larger weight on the benefits of compliance than on the costs of compliance, which will produce distortions. We show that the legislature can reduce the resulting distortions by limiting the enforcement agency budget and by permitting the agency partially to self-finance, by retaining a portion of its noncompliance penalties. Finally, if the enforcement agency has a good "signal" of firms' compliance costs, the distortions can be made very small by appropriate choice of the enforcement budget.

Acknowledgments: We thank Evelyn Brodkin, Edward Gramlich, Jon Harford, and members of the Berkeley and Stanford Industrial Organization Workshops, the University of Michigan Public Finance Seminar, the Carnegie-Mellon Political Economy Workshop, and the Eastern Michigan University Economics Departmental Seminar for useful comments. This research was supported by a University of Michigan Rackham Fellowship and Grant, and by the National Science Foundation, Grant SES-86-10021.

*University of Michigan
**University of California, Berkeley

THE SOCIAL COST OF UNIFORM REGULATORY STANDARDS

Abstract: Uniform regulatory standards are widely criticized as inefficient on the grounds that all firms are required to comply, regardless of compliance cost. Since firms will self-select to comply only if their compliance costs exceed the expected penalty for non-compliance, the inefficiency *could* be avoided by an enforcement policy chosen to maximize social welfare. But we argue that the enforcement agency goal is likely to place a larger weight on the benefits of compliance than on the costs of compliance, which will produce distortions. We show that the legislature can reduce the resulting distortions by limiting the enforcement agency budget and by permitting the agency partially to self-finance, by retaining a portion of its noncompliance penalties. Finally, if the enforcement agency has a good "signal" of firms' compliance costs, the distortions can be made very small by appropriate choice of the enforcement budget.

SOCIAL COST OF UNIFORM REGULATORY STANDARDS

Regulatory standards promulgated by the Environmental Protection Agency [EPA], the Occupational Safety and Health Administration [OSHA], and other agencies charged with public safety are typically uniform: All firms must comply, regardless of their compliance costs. The uniform standards are widely criticized as inefficient, on the grounds that efficient standards would apply only to firms for which the social benefit exceeds the social cost of compliance.[1]

This criticism of uniform standards is potentially shortsighted because it fails to distinguish between the standards *stipulated* by the Congress or executive agencies and the *effective* standards, represented by the pattern of compliance induced by enforcement. Firms will self-select to comply with the standard only if the cost of compliance is less than the expected fine for non-compliance. It is *possible* to achieve an efficient pattern of compliance, despite the uniform standard, simply by setting an expected fine equal to the benefit of a firm's compliance.[2] The effectiveness of such a policy depends on how well the enforcement agency can observe the benefits of each firm's compliance. If different inspection policies can be applied to firms with different benefits of compliance, then the socially efficient pattern of compliance could emerge. Although this argument requires that the enforcement agency can observe the benefits of compliance, it is unnecessary for the enforcement agency to observe firms' private compliance costs, since firms will self-select to comply only if their private costs of compliance are less than the expected fine.

---

[1] See, for example, Kneese and Bower [1968], Kneese and Schultze [1975], Nichols and Zeckhauser [1977], Nichols [1984], and the wide-ranging U.S. Senate Study on Federal Regulations [1978].

[2] Furthermore, as has often been pointed out, the cost of inspections theoretically can be made arbitrarily small by choosing fine rates arbitrarily large, while maintaining the required expected fine.

The efficient pattern of compliance will emerge only if the regulatory agency chooses its enforcement policy to maximize social welfare, as measured by the benefits of compliance net of all costs, including inspection costs and firms' private compliance costs. To assume that the agency ignores firms' compliance costs in setting the uniform standard, but respects firms' costs in enforcing it, seems inconsistent. The same political pressures that constrain agency enforcement most likely constrain agency standard-setting. The Congressional committees which craft enabling legislation for agencies such as EPA and OSHA also oversee standard-setting *and* enforcement. These committees are generally thought to attract advocates for their specific programmatic areas, rather than being representative of Congress as a whole.[3] The advocacy orientation is reflected, for example, in the language of OSHA's and the EPA's enabling statutes, which focus on the benefits of regulation, without reference to an efficient trade-off between the benefits and costs of regulation.[4] Any references to cost considerations or feasibility constraints generally revolve around the political criterion of preventing plant closings.[5]

Another important influence on the agency (and Congressional committees) is interest groups who lobby about environmental safety and health issues. Whereas lobbies

---

[3]See, for example, Stigler [1971], Peltzmann [1976, 1984], and Becker [1983], who describe the regulatory process as one in which organized interest groups divert the policy focus from the "public interest" in order to serve their own more narrow interests.

[4]For example, the Occupational Safety and Health Administration (OSHA) statute directs the agency to implement standards which attain the "highest degree of health and safety protection for the employee" [section 6(b)(5)]. The Environmental Protection Agency (EPA) is mandated to set standards incorporating "an adequate margin of safety" for all pollutants that "endanger the public health or welfare," (section 108). Between 1968 and 1978, Congress promulgated a wide range of "new" social legislation employing similar regulatory strategies, including the Consumer Product Safety Act, Traffic Safety Act, Child Protection and Toy Safety Act, Coal Mine Health and Safety Act, Surface Mining Control an Reclamation Act, Truth-in-Lending Act, and the Toxic Substance Control Act, among others.

[5]For example, at OSHA a consensus standard is considered economically infeasible at a particular plant if implementation would seriously jeopardize the cited employer's long-term financial profitability and competitiveness. For standards promulgated through rule-making [section 6(b)(5) standards], compliance with the standard would have to threaten the whole industry's long-term financial profitability and competitiveness before any single plant in the industry could be exempted. See Mintz, pp. 518-519.

exist both for and against stringent environmental safety and health policies, the lobbyists for stringent policies and enforcement have the strong statutory language to support their side of the argument. The various environmental laws also provide for citizen law suits, a threat which puts pressure on agencies to implement the statutory goals. Courts have cited the statutory language in upholding the focus on benefits of compliance, with limited reference to firms costs except when plants may shut down.[6] An enforcement policy that maximizes social welfare (benefits of compliance, net of firms' costs) typically does *not* maximize compliance or the benefits of compliance. If the agency were to use social welfare as its enforcement goal, groups lobbying for environmental health and safety issues could complain that the agency was ineffective, arguing that it would be possible to increase compliance (or the benefits of compliance) without increasing the enforcement budget.

For these reasons, it is reasonable to assume that an agency that has set uniform standards[7] will not necessarily enforce the standards in a way that maximizes social welfare (benefits net of firms' costs). If the agency does not place equal weight on the benefits and costs of compliance, uniform standards will lead to inefficient patterns of

---

[6]See Melnick (1984). In the legal challenges to OSHA standards, the Supreme Court has never supported use of the principle of economic efficiency in agency standard-setting. When the Court vacated the benzene standard, it directed the agency instead to find "significant risk," a benefit-based concept, before promulgating standards. (Industrial Union Department v. American Petroleum Institute, 448 US 607 (1980).) In the cotton dust decision, the Court explicitly rejected the use of cost-benefit analysis in the development of OSHA health standards. (American Textile Manufacturers Institute v. Donovan, 452 US 490 (1981).)

[7]For simplicity, we study regulations that are "either/or": Either an automobile in California has an emissions control device, or it does not. Either a coal-burning electricity plant installs smoke scrubbers or it does not. While a model with variable compliance levels would be richer, this simple model allows us to make our main points. OSHA generally establishes standards limiting allowable exposures to toxic substances that are uniform across all plants, but several EPA laws distinguish between old and new sources and among industries--imposing uniform standards within categories, but not necessarily across categories. The criteria for the distinctions are usually based on technological feasibility, not net benefits. The condition critical to our argument is that the variation in compliance costs across firms for a particular standard be sufficiently large that the uniform standard is inefficient. C. James Koch and Robert A. Leone [1978] reported the substantial variation in the costs of complying with uniform water emission standards within the tissue paper industry. Albert L. Nichols [1984] documented differences in costs of meeting uniform benzene air emission standards by manufacturers of maleic anhydride.

compliance. We examine the extreme case, in which the agency objective function has a weight of one on the benefits of compliance and a weight of zero on the firms' costs of compliance;[8] its enforcement goal then is to maximize the benefits of compliance subject to its enforcement budget. If we can show that the inefficiency is small in this extreme case, then it will be even smaller if the agency places some weight on firms' costs.

If an agency seeking to maximize the benefits of compliance could set its own fine level, as well as allocate its inspection budget, it (theoretically) could achieve full compliance with a vanishingly small budget (with "infinite" fines). Full compliance is inefficient if some firms' costs exceed the benefits of compliance. In order to curb the agency's overzealousness, it is necessary that the power to set the maximum fine rate and the power to set the inspection budget be vested outside the agency. This is precisely how the American system works: The appropriations committees of the Congress (*not* the agency's oversight committee) set the agency's budget, and finite maximum fines are set by the agency's enabling legislation.[9] As a consequence, there is power in the budget to curb the agency. If this power is exercised, the agency seeking to maximize compliance will consider itself under-financed.

Lower fines require higher inspection budgets to achieve a fixed level of compliance. By setting the fine rate higher, the social loss due to enforcement costs can be reduced. In this paper, we take the fine rates to be fixed exogenously in the enabling legislation. We use the term "first-best" to describe the pattern of compliance when the agency maximizes benefits net of firms' private costs rather than of benefits of compliance. But since we also assume fine rates are fixed, the policy is not truly first-best.

---

[8] Alternatively, we could explore the case in which firms with high costs of compliance have some influence in setting goals for enforcement, resulting in a non-zero weight on compliance costs in the agency goal. See, for example, Baron [1985].

[9] Jones (1988) summarizes the enabling legislation that sets various enforcement agencies' fine rates.

We investigate the divergence between the first-best pattern of compliance and the pattern of compliance that emerges if the appropriations committee distributes the budget to maximize social welfare,[10] taking into account that each enforcement agency allocates its budget inefficiently (maximizing the benefits of compliance rather than benefits net of firms' private compliance costs.) The appropriations committees cannot set salaries or fire agency administrators according to whether administrators choose enforcement policies in the public interest; if they could, the appropriations committees could induce the agency to maximize social welfare. The appropriations committees have available only two weak budgetary instruments: the budget level itself and the power to set a rebate policy in which the agency keeps a share of the fine revenue it collects. We explore the power of these budgetary instruments to undo the inefficiencies that result from inefficient enforcement by regulatory agencies.

Our main conclusions are: (i) The optimal budgetary policy is to limit the enforcement budget, but also to require the agency partially to self-finance by retaining a share of the fines it collects. (ii) Typically the first-best pattern of compliance cannot be achieved by feasible budget instruments. However, the second-best will be close to the first-best if each agency, which cannot observe firms' private costs of compliance directly, has a good "signal"[11] of each firm's private cost, and the inspection probability can depend on this signal. (iii) Budgetary instruments may not be powerful in getting close to the first-best if the signal of compliance cost is relatively uninformative.

---

[10] Appropriations committees are generally viewed as representative of Congress as a whole. See Fenno, [1966] for the classic discussion of the Congressional budgetary process, or Allen Schick, [1980], Congress and Money, for coverage of more recent budget process history. Becker [1983] argues that, within a representative body, even the "special interest" model of government yields predictions that policies will tend to correct market failures while at the same time favoring the politically powerful.

[11] By a "signal", we mean an observable aspect of the firm that is correlated with its compliance cost. In addition to informal sources of information about cost available through the enforcement network, the federal government collects regulatory cost data in the process of individual rule-making. See, for example, Research Triangle Institute, "Regulatory Analysis of the Proposed OSHA Standards on Asbestos," prepared for OSHA and the USDOL, May 1984. Also, the U.S. Census Bureau reports data at the 4-digit SIC level on the costs of different regulations based on an annual survey of 20,000 firms. See Evans (1986) for summary statistics.

Section I describes the enforcement policy that maximizes social welfare (which we call first-best), assuming the fine rate is fixed. Section II studies the second-best enforcement policy, in which the Congressional appropriations committees choose the optimal enforcement budget, taking into account that the enforcement agency maximizes the benefits of aggregate compliance, rather than benefits net of firms' compliance costs.

The inspection policy can depend on the observable cost signals, which are correlated with firms' true costs. When a signal is available, two types of inefficiency occur in the second-best policy.[12] First, in both the first-best and second-best policies, some inspection classes with high cost signals (and high costs, on average) will not be inspected at all, and will have no compliance. But the cutoff signals that separate inspected classes from uninspected classes may differ in the first- and second-best. Second, the patterns of compliance among inspected classes differ: In the second-best, the compliance rate is uniform across all inspected sectors rather than varying inversely with the cost signal as in the first-best.

In Section III, we show that by requiring the agency partially to self-finance from its non-compliance fines, the appropriations committee can reduce distortions. Rebates provide incentives for the agency to shift compliance from high-cost sectors to low cost sectors. A different explanation that has been offered for why agencies should be allowed to keep a share of their revenues is that rebates reduce moral hazard. When the effort level of bureaucrats cannot be observed, rebates can induce them to work harder. We show here that another useful incentive effect of rebates is to induce bureaucrats to allocate the inspection budget more efficiently.

---

[12] We do not consider a third possible inefficiency which could occur in both the first-best and the second-best contexts. In a one-period model where the penalty amount is limited, the expected non-compliance penalty may be too small to induce compliance where socially efficient even if the probability of inspection were one. [See Viscusi and Zeckhauser (1979), and Jones (1988)]. Penalties can be compounded if the agency's treatment of a firm can depend on the firm's history of compliance. Since this is true with both the distorted agency goal and the "public interest" agency goal, it is peripheral to our concerns.

The rebate mechanism we have in mind is an implicit understanding between the appropriations committee and the agency that last year's fine revenues will be considered in this year's budget. The more direct mechanism of off-line budgeting also exists, though it appears to be rarely used for non-compliance penalties received from violators.[13]

In Section IV, we show that if each noncompliant firm is costly to prosecute (effectively imposing a negative rebate), the inefficiencies in the pattern of second-best compliance among inspected classes is exacerbated. With no rebates, the compliance rate would rise with the cost signal, contrary to the optimal pattern of compliance.

In Section V, we show that if the signal of firms' compliance cost is very informative, budgetary control can make the allocative distortion small. Unfortunately, the same is not true when the signal is only slightly informative. In that case, the agency's induced pattern of compliance can be very far from optimal, and there is no power in the budget to undo this problem. Social welfare might be enhanced if there were no cost signal whatsoever.

For simplicity, the body of this paper discusses atemporal inspection policies: The probability of inspecting any particular firm does not depend on its history of inspections or compliance. Since there is no opportunity to bring noncompliant firms into compliance as a consequence of the inspection, the only benefits of enforcement with an atemporal policy arise through ex ante deterrence. In Appendix B we show that our main result, that the social loss in the second-best is small if the cost signal is a good predictor of cost, still holds when the enforcement agency can re-inspect and thereby accrue ex post benefits from bringing firms into compliance, and firms know ex ante that this will occur only when it is ex post reasonable for the inspection agency to reinspect.

---

[13] In 1987, $106B (est.) in collections from non-federal sources for user fees, loan repayments, and penalty assessments went directly into agency budgets [USGAO, 1987].

I. First-best Regulation and Enforcement

We assume that firms have different unobservable costs, g, for complying with the regulation. However, each firm has an observable cost signal, y, such as its industry or product line. Since the signal is observable before inspection, the enforcement agency can choose a different frequency of inspection for each cost signal, y. The cost of an inspection is c and the fine imposed on a noncompliant firm is f. We assume the allowable fine is fixed and not chosen as part of the enforcement policy. Budgetary control depends on the fixed fine rate.

We assume that compliance costs in class y have mean y and are distributed symmetrically around y according to H(g-y), on a support contained in (y-m,y+m).[14] Thus, each inspection class has the same distribution of costs, except for location. For each y, H is differentiable and positive on the interior of its support.

If a firm in class y is inspected with probability p(y), it will self-select to comply if $g < p(y)f$, and therefore the compliance rate in inspection class y will be H(p(y)f-y). This self-selection by cost is why the pattern of compliance can be close to efficient even though the compliance costs of individual firms are unobservable. There is no analogous mechanism to make firms self-select by benefits, but if the agency wants to maximize benefits of compliance, it will set higher expected fines in sectors where the benefits of compliance are high. Since our arguments would be neither enriched nor undermined by assuming that the benefits of compliance differ across firms, we assume the social benefit of each firm's compliance is one. Aggregate benefits are $\Sigma_i H(p(y_i)f-y_i)$. (With little loss of generality and no loss of insight, we assume that all inspection classes are the same size.)

---

[14] The constraint that compliance costs are non-negative means that cost signals are bounded away from zero. We have parameterized the size of the support of h with one parameter, m, for convenience, but if the distribution were asymmetric, the density h might be zero on part of the domain (y-m,y+m).

Since firms in any inspection class all face the same expected fine, there will be a "cutoff" cost of compliance, $G^*(y)$, achieved by inspection probability $p^*(y)=G^*(y)/f$, which separates compliant from noncompliant firms. $G^*(y)$ maximizes social welfare (1):

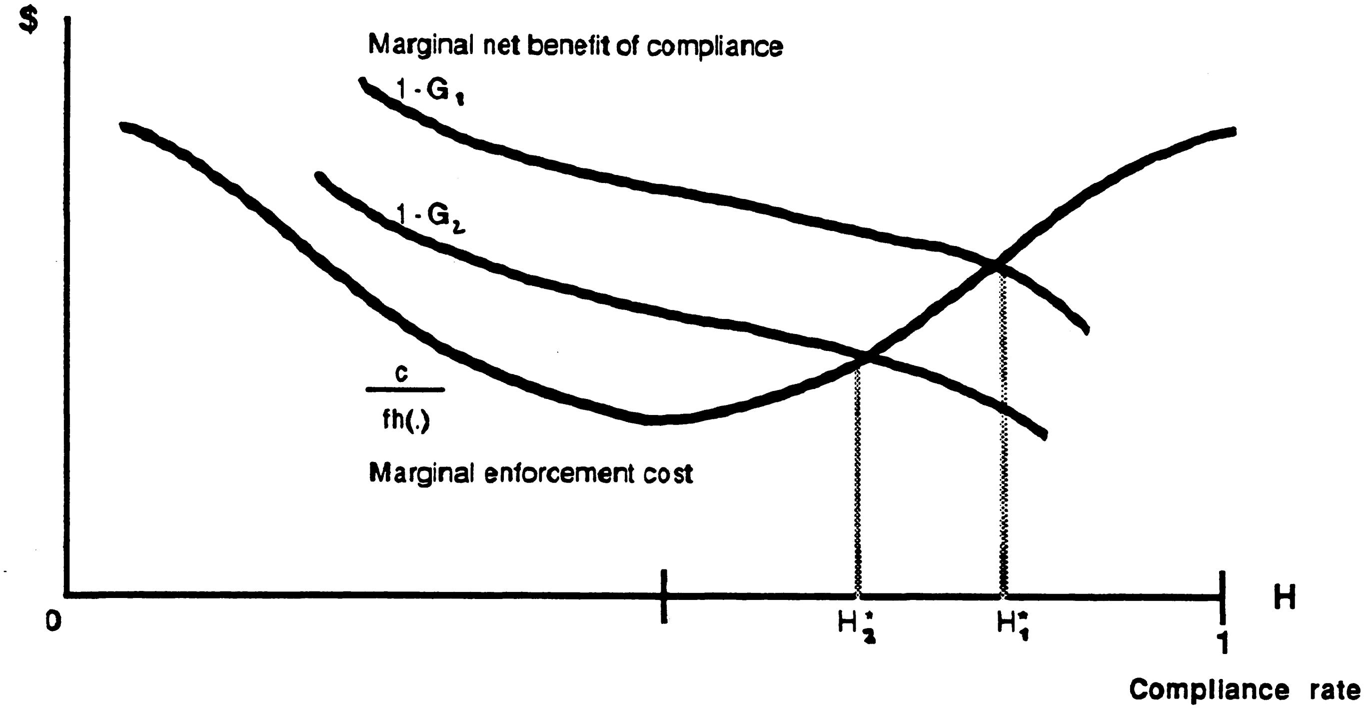
$$(1)\quad SW(G,y) = H(G-y) - \int_{y-m}^{G} g\, h(g-y)dg - cG/f$$

The first term is the benefit of compliance, the second term is the compliance cost borne by firms, and the third term is the cost of inspections. If inspection class y is inspected at all, the first-order condition describing the optimal $G^*(y)$ or, equivalently, the optimal $p^*(y)$, is

$$(2)\quad [1-G] - c/fh(G-y) \geq 0, \quad \text{with equality if} \quad H(G-y)<1.$$

We can differentiate (2) implicitly to see how the compliance rate optimally changes with the cost signal within inspected classes. Assuming the objective function is strictly concave at the optimum, and $0<H(G^*(y)-y)<1$, it follows that $dG^*(y)/dy < 1$.

The optimal pattern of compliance is shown in Figure 1. The downward sloping lines, $1-G_i$, show the net benefit of bringing a marginal firm in sector i into compliance, as the compliance rate H_i is increased in a low-cost sector, y_1, and a high-cost sector, y_2. At a given compliance rate, H, the net benefit of bringing the marginal firm into compliance is lower in the high-cost sector because the marginal firm at that compliance rate has a higher compliance cost. The marginal inspection cost of increasing the compliance rate in each sector is $c/fh(G-y)$. If the density h is single peaked and symmetric, the marginal cost of inspections will be as shown for both inspection classes, y_1 and y_2. [If $H(G_1-y_1)=H(G_2-y_2)$, then $h(G_1-y_1)= h(G_2-y_2)$.] The second-order condition only requires that the firms' net benefits curve, 1-G, cross the marginal cost curve for inspections from above; $h'(.)$ can be positive or negative at the optimum.



i = 1 low compliance cost sector

i = 2 high compliance cost sector

c = cost of an inspection

f = non-compliance fine

G (= pf) = compliance cost of marginal firm in compliance

$H^* = H(G^*(y) - y)$ first best compliance rate in sector i

h = compliance density

Figure 1. First best enforcement strategy

Since $dG^*(y)/dy < 1$, the compliance rate declines with y: The high-cost class has a lower compliance rate than the low-cost class at the social optimum.

High-cost inspection classes should not be inspected at all if the maximum possible social welfare is negative.[15] There is a "cutoff" cost signal, say Y^*, above which inspection classes should escape scrutiny. Compare two inspection classes, $y_1 < y_2$. By definition, $SW(G^*(y_1),y_1) > SW(G^*(y_2),y_1)$. But, also, $SW(G^*(y_2),y_1) > SW(G^*(y_2),y_2)$, since $\partial SW(G,y)/\partial y < 0$. It follows that if class y_1 escapes scrutiny, class y_2 does also.

II. Distortions in the Second-Best Agency's Inspection Policy

We now investigate the second-best pattern of compliance when the enforcement policy maximizes benefits of compliance subject to an enforcement budget constraint. Unlike in the social optimum above, we cannot separate the enforcement problems for different inspection classes because, with a fixed budget, an increase in inspections in one inspection class requires a decrease in another class. The enforcement agency chooses G(y,E) (for each y) to maximize:

$$(3) \quad \Sigma_i \, H(G(y_i,E)-y_i) \qquad \text{subject to} \quad E = c\,\Sigma_i \, [G(y_i,E)/f]$$

where E is the enforcement budget. Here we have again substituted G/f for p. Provided a class is inspected, the first-order condition describing the optimum is

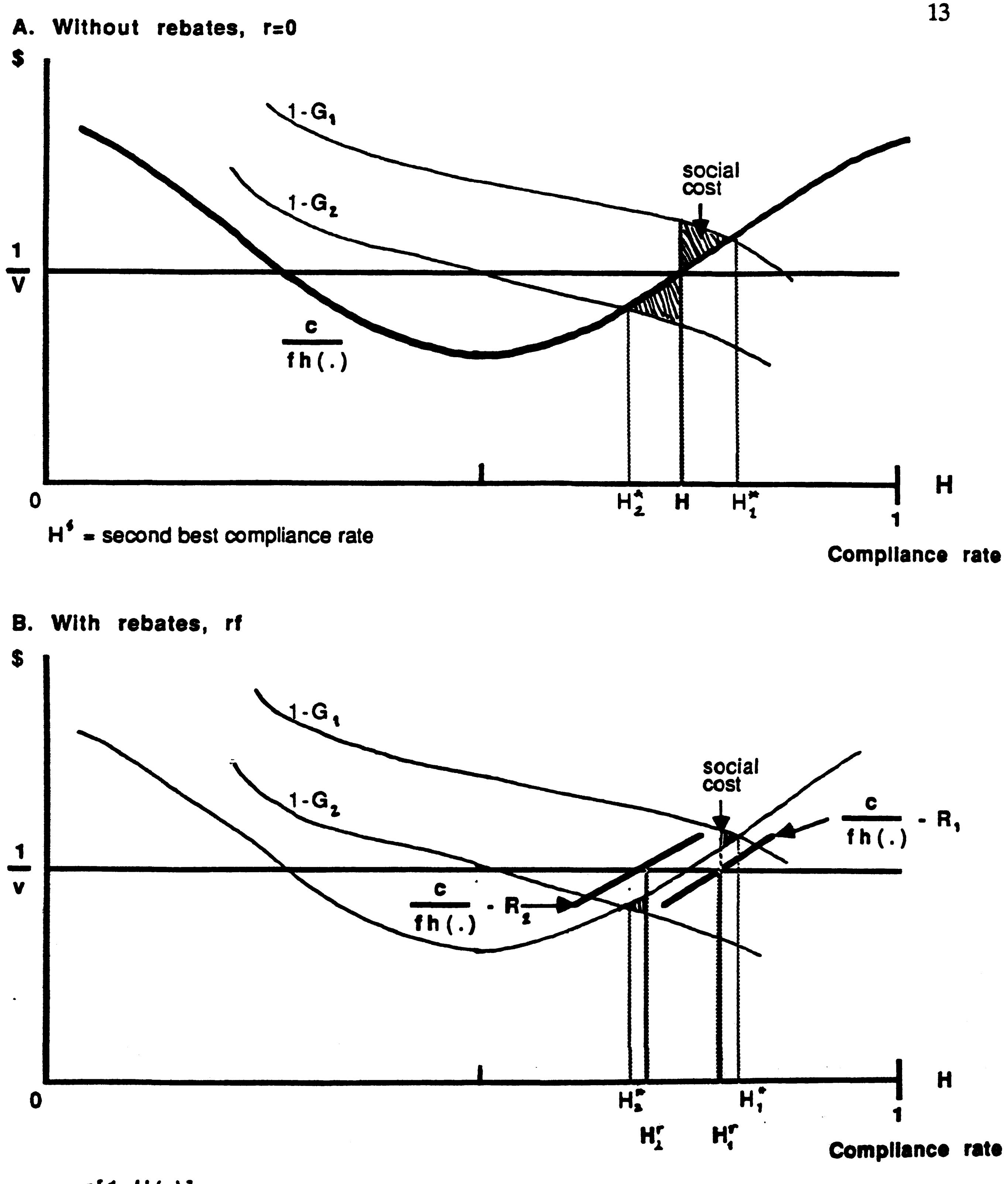
$$(4) \quad [1/v(E)] - c / [fh(G(y,E)-y)] \geq 0, \quad \text{with equality if } H(G(y,E)-y) < 1$$

[15] If the regulation applies to a prominent industry, selective non-enforcement may be politically untenable. For example, the corporate average fuel economy (cafe) standard applies to the automobile industry, which has three very large domestic participants. The costs of compliance vary substantially across the firms, but it is generally considered to be politically infeasible to enforce the standard selectively within the industry. However, the consequence has not been full compliance: The industry is engaged in negotiations with the government to revise the standard.

where v(E) is the shadow value of an additional dollar in producing compliance. Since the second-best agency does not incorporate firms' private compliance costs in its goal, the marginal benefit perceived by the agency of increasing the compliance by one firm is one. The marginal benefit $1/v(E)$ is the dollar value of marginal compliance, given that the budget is constrained. In Figure 2, the perceived marginal benefit curve is horizontal, therefore the marginal cost curve in panel 2A (with no rebates) must be increasing at the optimum for an interior solution; that is, the second-order condition requires $h'(.)<0$.

An important feature of both the first- and second-best inspection patterns is that some high-cost (high y) inspection classes may not be inspected at all, though the cutoff may be different in the two cases. It takes a higher frequency of inspection to elicit a given compliance rate from a high-cost inspection class than from a low-cost inspection class. As in the first-best, eliciting compliance from firms in a high-cost inspection class may cost more than its value to the enforcement agency. For y larger than some cutoff, say Y, the net value $H(G(0,E))-cv(E)[G(0,E)+y]/f$ will be negative, so $p(y,E)=0$ and the compliance rate is zero. In other words, though the agency does not care directly about private costs of compliance, it nevertheless exempts high-cost inspection classes because the budget is limited and it is costly to induce compliance from high-cost firms.

Within inspected classes, the compliance rate will be uniform in the second-best. It may be uniformly too high in inspected classes (when $h'>0$ in the first-best and $h'<0$ in the second-best) or it may be too high in high-cost classes but too low in low-cost classes (when $h'<0$ in both the first- and second-best). Figure 2A illustrates the social cost of uniform standards with second best enforcement for the latter case.



$R_i = \dfrac{r[1-H(.)]}{h(.)} - rG$, H_i^r = second best compliance rate with rebates in sector i

Figure 2. Social cost of uniform standards with second best enforcement : Diffuse cost signals.

III. The Incentive Effects of Rebated Fine Revenues

We now consider how rebated fines affect the pattern of compliance achieved by the compliance-maximizing enforcement agency. With probability 1-H(.), an inspected firm is noncompliant, and will be prosecuted and fined. We assume the appropriations committees permit the agency to keep a fraction r of the fine revenues to finance other inspections. As before, we can substitute G(y,E,r)/f for p(y,E,r) to describe the enforcement agency's objective function. For each y, the enforcement agency chooses G(y,E,r) to solve:

(5) Maximize $\Sigma_i H(G(y_i,E,r)-y_i)$ subject to

$$E = \Sigma_i\, c\, [G(y_i,E,r)/f] - rf\, \Sigma_i\, [G(y_i,E,r)/f]\, [1-H(G(y_i,E,r)-y_i)]$$

If compliance in class y is positive, the first-order condition describing the optimum is

$$(6) \qquad \frac{1}{v(E,r)} - \frac{c}{f\,h(.)} + \frac{r[1-H(.)]}{h(.)} - rG(.) \geq 0$$

where, again, v(E,r) is the shadow value of an enforcement dollar in producing compliance. The condition holds with equality if the compliance rate in class y is less than one, and with inequality if all firms in class y comply.

When there are no fine rebates, the marginal cost of inspections, c/fh(.), is the same in each inspection class at each compliance level H(.), as in Figure 2A. But with rebates, the marginal cost curves in different inspection classes differ because the values of fine rebates differ. With equal compliance rates, H, there are more inspections in a high-cost class than in a low cost class, since it takes a higher probability of inspection to make the same share of higher-cost firms comply. When the compliance rate increases at the margin, rebates are lost on more inframarginal inspections in a high-cost class than in a low-cost class, and this makes the agency's effective marginal cost of inspections higher in the high-cost sector.

Figure 2B illustrates that with fine rebates, the marginal cost curve for the high-cost class lies above the marginal cost curve for the low-cost class. We have not shown that the height of the line $1/v(E,r)$ will also change with r if E is held constant.) As a result, compliance in the low-cost class will be larger due to the fine rebate, while compliance in the high-cost class will be smaller. This shift will enhance efficiency, as Figure 2B illustrates.

PROPOSITION 1: *The optimal budget policy requires rebates as well as direct budget, since rebates encourage the second-best agency to shift inspections from high-cost classes to low-cost classes.*

Proof: There are many combinations of rebates r and direct budgets E(r) that induce the same fixed total expenditure on inspections, say E^*. Starting with no rebates, $r = 0$, and the direct budget $E(0)=E^*$, we will show that a marginal increase in r [and the corresponding decrease in E(r) required to hold expenditures on enforcement fixed] increases social welfare. We will simplify notation by writing G(y,r) and v(r) instead of G(y,E(r),r) and v(E(r),r).

Since total expenditures on inspections are fixed at $c\ \Sigma_i\ p(y_i) = (c/f)\ \Sigma_i\ G(y_i,r)$, constant total expenditures imply that $\Sigma_i\ [\partial G(.)/\partial r] = 0$. Social welfare is

$$(7) \quad SW = \Sigma_i \int_{y_i-m}^{G(y_i,r)} (1-g)\ h(g-y_i)\ dg \ - c/f\ \Sigma_i\ G(y_i,r).$$

Since the derivative of the last term with respect to r is zero, the change in social welfare when r increases marginally is the derivative of the first term, or

(8) $\partial SW / \partial r = \Sigma_i [1-G(y_i,r)] h(G(y_i,r)-y_i) [\partial G(y_i,r)/ \partial r]$

We will evaluate the derivative at the initial point, $r = 0$. Since the agency choice of $G(y,0)-y$ is constant for all y (from equation (4) above), $h(G(y,0)-y)$ has the same value, say h, for all y for which $0<H(.)<1$. Therefore the derivative of social welfare at $r = 0$ has value

(9) $h \Sigma_i [\partial G(y_i,0)/ \partial r] - h \Sigma_i G(y_i,0) [\partial G(y_i,0)/ \partial r]$

Since the first term is zero, we only need to show the last term is positive. Differentiating (6) implicitly and then setting $r=0$, we discover that, for inspection classes with positive compliance, but not full compliance,

(10) $$\frac{\partial G(y,0)}{\partial r} = \frac{-v(0) [1-H(G(y,0)-y) -G(y,0)h(G(y,0)-y)] + [dv(0)/dr] (c/f)}{h'(G(y,0)-y)}$$

The second order condition for (3) (or (5) at $r=0$) requires $h'(.)\leq 0$. For the following reason, it will not be cost-effective for the agency to have $h'(.)=0$. By symmetry and single-peakedness of h, $h'(.)=0$ would imply the compliance rate was 50% in all inspected classes. Consider any two such classes. By symmetry of h, the cost-savings of decreasing compliance in one class to zero is the same as the additional cost of increasing compliance in the other class to one, and the same aggregate compliance is preserved. But now there is an additional lump-sum cost saving in the sector whose compliance rate has been reduced to zero, since it takes a non-trivial inspection cost $(y-m)/f)$ to achieve the lowest-cost firms' compliance in that class. This lump-sum saving means that the compliance rates zero and one are less costly to achieve than compliance rate 50% in both sectors, and therefore the second-best optimum will not have $h'(.)=0$.

Since h(.), h′(.) and H(.) respectively have fixed values for all y when r=0, and since h′(.)<0 by the second order condition, we see that $\partial G(y,0)/\partial r = a - bG(y,0)$ for appropriate constants a and b>0.

Since G(y,0) is nonincreasing with y, it follows that there exists $\hat{y}$ such that $\partial G(y,0)/\partial r < 0$ if $y > \hat{y}$ and $\partial G(y,0)/\partial r > 0$ if $y < \hat{y}$. It also follows that:

$$- \Sigma_i G(y_i,0) [\partial G(y_i,0)/\partial r] \geq - \Sigma_i G(\hat{y},0) [\partial G(y_i,0)/\partial r]$$

$$= - G(\hat{y},0) \Sigma_i [\partial G(y_i,0)/\partial r] = 0$$

Thus, within the inspection classes with positive but not complete compliance, social welfare increases with a marginal increase in r, from r=0. There will be no change in compliance of the inspection classes for which compliance is zero. The marginal adjustment to r could generate at most a marginal saving in enforcement costs to be applied to inspection classes with no compliance. The smallest probability of inspection that will elicit positive compliance from such an inspection class is c(y-m)/f, which is nonmarginal. Q.E.D.

We have implicitly assumed throughout the analysis that full compliance across all sectors was inefficient. If alternatively, full compliance is efficient, but insufficient budget were allocated to the enforcement agency to achieve it, rebates would still increase efficiency.

IV. The Incentive Effects of Prosecution Costs

Our argument in Section I implies that, provided the enforcement agency maximizes social welfare rather than aggregate compliance, there is no social loss *due to the uniform standard*. Enforcement is costly because fine rates are finite, but conditional on the finite fine rates, a benevolent social planner could not increase social welfare by choosing a nonuniform standard. The result does not hold if the agency must incur prosecution costs to bring noncompliant firms into compliance. Suppose the enforcement agency incurs prosecution costs, e, to document or litigate noncompliance. If the regulatory standard could be ***nonuniform***, the agency operating in the public-interest could avoid prosecution costs (without changing incentives to comply) by stipulating that any firm with compliance cost greater than $G^*(y) = p^*(y)f$ is exempt. Since the only noncompliant inspected firms would be exempt from compliance, no firms would be prosecuted and the costs would be avoided. But with a *uniform* standard, prosecution costs cannot be avoided when non-compliant firms are detected in inspected sectors. For the second-best enforcement agency, a *de jure* obligation to prosecute firms detected to be non-compliant would exacerbate the inefficient distribution of compliance among inspected classes.

The enforcement problem with costly prosecution can again be described by (5), except that we must substitute +e for -rf. Instead of getting a rebate rf for every noncompliant firm it inspects, the enforcement agency pays a prosecution cost, e. The incentives due to rebated fines are then reversed. Prosecution costs inefficiently shift inspections from low-cost inspection classes to high-cost inspection classes. Initially, it might seem surprising that the enforcement agency wants to shift inspections to high-cost classes that are already heavily inspected, and therefore bear heavy prosecution costs. But the intuition is analogous to the rebate case above: Frequent inspections of high-cost classes increase compliance, thereby decreasing the expected prosecution costs. The reduction in prosecution costs as compliance increases is greater in high cost sectors because they require far more inspections to achieve a given compliance rate.

V. The Costliness of the Second-Best Distortion and the Power of the Budget

To evaluate the severity of the efficiency distortions when the enforcement agency maximizes aggregate compliance, rather than social welfare, it is instructive to consider the two extreme cases, (i) there is no cost signal, and (ii) the cost signal is totally informative. With no cost signal, the enforcement agency can observe nothing prior to inspection, as in the first-best, and will inspect all firms with the same probability. There will be a uniform "cutoff" cost level below which firms comply. The appropriations committee can elicit the first-best pattern of compliance from the second-best enforcement agency simply by providing the inspection budget that the first-best agency would use.

At the other extreme, if y is a *perfect* signal of cost, all firms with signal y have the same cost $g=y$. As in the social optimum, the probabilities of inspection will be $p(y,E)=y/f$ for y less than the cutoff determined by the enforcement agency's budget. Therefore, by giving the second-best enforcement agency the same budget that would be used by the first-best enforcement agency, the Congress can ensure that the resulting pattern of compliance will be first-best. We conclude:

<u>PROPOSITION 2:</u> *When (i) no cost signal is available, or (ii) the cost signal is perfectly correlated with compliance cost, the Congress can achieve the first-best by giving the second-best enforcement agency the budget that would be used in the first-best policy.*

Except in the two extreme cases mentioned, the first-best cannot be achieved merely by manipulating the second-best enforcement agency's budget. And even though fine rebates help, they are not powerful enough to make the second-best pattern of compliance coincide exactly with the first-best. It is therefore of interest to ask whether the social loss of the second-best inevitably becomes small as the extremes are approached.

We say that cost signals become more informative if the distribution of compliance cost, H, becomes more compressed around the mean, but the distribution of signals y is fixed.[16] Hence, the distributions of compliance cost in different classes overlap less as the signal becomes more informative. We index a sequence of distributions by n = 1,2..., and let $H^n(g-y) = H(n(g-y))$. The support of H^n is contained in $[y-(m/n), y+(m/n)]$ and the density, $h^n(g-y) = nh(n(g-y))$, becomes arbitrarily large on the interior of the support, where h is positive.

We show in Appendix A that as the density h^n becomes very large everywhere on the interior of its support, the compliance rates in inspected classes approach one, in both the first- and second-best. (A small increase in the probability of audit would otherwise generate a huge increase in compliance.) Furthermore, the appropriations committee can ensure that the high-cost inspection classes that escape scrutiny in the second-best are close to those that escape scrutiny in the first-best, simply by providing the first-best budget. Thus, since the first-best and second-best patterns of compliance become very similar as n becomes large, the social cost vanishes. This result occurs whether or not rebates are employed. Therefore:

<u>**PROPOSITION 3:**</u> (Appendix A) *As the distribution of compliance cost in each class becomes more concentrated around its mean y (n becomes large), the social loss in the second-best converges to zero.*

At the other extreme, we say that cost signals become less informative if the distribution of signals y becomes more compressed around its mean, while the distribution of compliance costs, H, remains fixed within each compliance class.[17] Hence, the distributions of compliance cost in different classes overlap more as the signal becomes

---

[16]For example, if $h(g-y)=(n/2m)$ on the support $(y-(m/n), y+(m/n))$, the cost signal becomes more informative if n becomes larger. The distribution of signals remains fixed.

[17]For example, if there are only two signals, y_1 and y_2, the cost signals become less informative as these two signals move toward their mean, $y_m = (1/2)(y_1 + y_2)$.

less informative. Unfortunately, it is *not* always true that the social loss in the second-best becomes small as the distribution of y becomes concentrated at its mean. A condition under which the social loss does become small is that the compliance-cost density h is strictly declining at the first-best. This condition is violated when h is uniform and, if the first-best compliance cutoff is less than the mean, when it is normal. The following example illustrates what can go wrong.

Example: Suppose h is uniform on $[y-(1/4), y+(1/4)]$, with density 1/2, and there are two inspection classes, $y_1 < y_2$. If $c=f=1$, and the benefit of a firm's compliance is one, $G^*(y)=1/2$ in both inspection classes (provided neither cost signal is so high that optimality requires zero inspections). Optimal compliance rates in the two classes differ, but they become close as y_2 and y_1 become close. However, if the second-best enforcement agency inspects an inspection class at all, it elicits full compliance from that inspection class.[18] [In (4), h is constant.] Thus, the enforcement agency will elicit full compliance from class y_1, and no (or partial) compliance from y_2 (depending upon its budget). As the distributions of compliance cost become very close (as one signal becomes close to the other), this pattern persists. As a result, a large social loss persists because high-cost firms in the low-signal class will comply, but low-cost firms in the high-signal class will not comply. This pathology cannot be avoided by letting the agency self-finance through fine rebates. **End of Example.**

The interpretation of this example is that a little information might be a bad thing. If the mean compliance costs in two industries differ only slightly, the second-best enforcement agency may enforce one of them heavily and the other not at all, which may be very far from optimal. In this example, the inefficiency cannot be avoided by controlling the budget or by rebating fines.

---

[18] The one possible exception is the highest-cost inspection class it inspects: the agency may run out of budget before it can induce full compliance from this class.

Nevertheless, Proposition 2 (the social loss vanishes as the signal becomes close to perfect) survives. As the signal becomes perfect, compliance rates in inspected classes become close to one, and by controlling the budget, the appropriations committee can ensure that the number of uninspected classes becomes close to the number in the first-best.

VI. Conclusions

This paper is motivated by the observation that when regulatory agencies set uniform standards, it is reasonable to assume that they enforce those standards in a way that places more weight on the benefits of compliance than on firms' private compliance costs. Such an enforcement goal introduces distortions from the socially efficient outcome. If such an agency could choose its own budget or the fine rate, it would be overzealous in its enforcement efforts, relative to the efficient level of enforcement. We have discussed the power of feasible budget instruments to control the overzealous agency. By restricting the agency budget, the Congressional appropriations committees can induce the agency to avoid inspecting the higher cost inspection classes for which compliance is inefficient. By requiring the agency partially to self-finance from its non-compliance penalties, Congress can mitigate the inefficient distribution of compliance among inspected classes. On the other hand, if prosecution is costly (negative fine rebates), the inefficient distribution of compliance will be exacerbated.

If the agency has a good signal of firms' unobserved compliance costs, setting the second-best budget level can substantially undo the distortions in patterns of compliance. With a good cost signal, the inefficient distribution of compliance within inspected classes is secondary in importance to the fact that both the first- and second-best policies concentrate attention on inspection classes with low cost.

The pattern of compliance can be close to efficient because firms self-select on the basis of compliance cost whether to comply. There is no analogous mechanism to make them self-select on the basis of benefits. We assumed throughout that benefits were equal across all sectors. However, if benefits of compliance varied across sectors the conclusions would be unchanged, so long as the enforcement agency incorporates the benefits of compliance in its enforcement goal and therefore enforces high-benefit sectors more intensively than low-benefit sectors.

Alternatively, an agency may simply maximize compliance, rather than benefits of compliance. In this case, if benefits vary across firms the welfare results about rebates may be compromised. For example, if benefits are closely correlated with the costs of compliance, then uniform compliance rates in the second-best may be very close to optimal. In the case where net benefits of compliance increase with the costs of compliance, rebates would induce inefficiencies.

Although we have assumed throughout that fine rates are fixed by the enabling legislation that created the enforcement agency, our discussion also suggests a partial explanation for why the enforcement agency is not allowed to choose fine rates. By giving agencies the power to set fines, Congress would forfeit the power of the enforcement budget, since full compliance could then be achieved costlessly. For any enforcement budget, the second-best enforcement agency could choose penalties sufficiently high to induce full compliance. The appropriations committee can prevent this inefficient outcome by reserving the power to set both fine rates and the enforcement budget. In contrast to other explanations in the literature for why optimal fine rates are bounded,[19] this explanation rests on the hierarchical agency problem in which the higher-level principal seeks budgetary control over the lower-level principal with inefficient goals.

---

[19] For example, infinite fines may lead to an inefficiently large loss in utility if consumers are risk averse and consumers sometimes fail to comply because their benefits of noncompliance are high (Polinsky and Shavell (1979)) or because there is randomness in whether they are convicted (Snyder (1987).

References

Baron, D. P. 1985. "Noncooperative Regulation of a Nonlocalized Externality," *Rand Journal of Economics*, Vol. 16, No.4.

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

____________. 1983. "A Theory of Competition among Pressure Groups for Political Influence." *Quarterly Journal of Economics,"* Vol. 98, pp. 371-400.

Demski, Joel S. and David E. M. Sappington. 1987. "Hierarchical Regulatory Control," *Rand Journal of Economics*, Vol 18, pp. 369-383.

Evans, David. 1986. "The Differential Effect of Regulation Across Plant Size: Comment on Pashigian." *Journal of Law and Economics*, Vol. XXIX, 1, pp. 187-200.

Fenno, Richard F. Jr. 1966. *The Power of the Purse: Appropriations Politics in Congress.* Boston: Little, Brown and Company.

Jones, Carol A. 1982. "Models of regulatory enforcement and compliance."

______________. 1988. "Standard-setting with Incomplete Enforcement Revisited", *Journal of Policy Analysis and Management.* Vol. 8, No. 1.

Kneese, A. V. and B. T. Bower. 1968. *Managing Water Quality: Economics, Technology, Institutions*. Baltimore: John Hopkins Press.

Kneese, A. V. and C. T. Schultze. 1975. *Pollution, Prices and Public Policy*. Washington, D.C.: Brookings Institution.

Koch, C. James and Robert A. Leone. 1979. "The Clean Water Act: Unexpected Impacts on Industry," *Harvard Environmental Law Review*, 3, pp. 84-111.

Melnick, R. Shep. 1984. "The Politics of Partnership: Institutional Coalitions and Statutory Rights," Center for American Political Studies, Harvard University.

Mintz, Benjamin. 1984. *OSHA: History, Law, and Policy*. BNA Books. Washington D. C.

Nichols, Albert L. 1984. *Targeting Economic Incentives for Environmental Protection*. Cambridge: MIT Press.

Nichols, Albert L. and R. J. Zeckhauser. "Government comes to the workplace: An assessment of OSHA." *The Public Interest*, Number 49, Fall 1977, 39-69.

Peltzman, S. 1976. "Toward a more general theory of regulation." *Journal of Law and Economics.* Vol. 19, pp. 211-240.

Peltzman, S. 1984. "Constituent interest and Congressional voting." *Journal of Law and Economics*. Vol. 27, pp. 181-210.

Polinsky, A. Mitchell and Steven Shavell, "The optimal trade-off between the probability and magnitude of fines." *AER*, Vol. 69, No. 5, Dec. 1979, pp. 880-891.

Research Triangle Institute. "Regulatory Analysis of the Proposed OSHA Standards on Asbestos." For OSHA/DOL, May 1984.

Sanchez, Isabel. 1987. "Hierarchical Design and Enforcement of Income Tax Policies", Essay from Ph.D. Dissertation, U.C., San Diego.

Schick, Allen. 1980. *Congress and Money*. Urban Institute, Washington D.C.

Scotchmer, Suzanne and Joel Slemrod. 1987. "Optimal Obfuscation in Tax Enforcement," U.C. Berkeley, Graduate School of Public Policy Working Paper 126.

Scotchmer, Suzanne. 1986. "Equity in Tax Enforcement", Harvard Institute of Economic Research Working Paper 1233.

________________. 1987. "Audit classes and tax enforcement policy," *AER*, Vol. 77, No. 2, pp. 229-233.

Snyder, Edward. 1987. "Enforcement and Efforts to Influence Verification: Theory and Evidence from Anti-trust." University of Michigan Working Paper.

Stigler, George. 1971. "The Theory of Economic Regulation." *Bell Journal of Economics and Management Science*. Vol. 2, pp. 3-21.

USGAO. *1987. Use of Spending Authority and Permanent Appropriations is Widespread.* Report AFMD-87-44.

U.S. Senate Committee on Government Affairs. 1978. *Study on Federal Regulations.*

Viscusi, W. Kip and Richard J. Zeckhauser, 1979. "Optimal Standard-setting with Incomplete Enforcement." *Public Policy*, 27, 437-56.

Weingast, B. R. and M. J. Moran 1983. "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *Journal of Political Economy*. Vol. 91, pp. 765-800.

APPENDIX A
PROPOSITION 2

To streamline the argument, we assume there is a continuous distribution of y, F(y). We first define some notation in order to state Proposition 2 more precisely. Let $SW(G^*(y,n),y,H^n)$ be social welfare, analogous to (1), when the distribution of compliance cost is H^n, and $G^*(y,n)$ is the optimal "cutoff" in class y. Suppose the first-best optimal inspection budget is E^{*n}. Let $SB(E^{*n},H^n)$ be the social surplus in the second-best enforcement policy with budget E^{*n}.[20] To show that the social loss in the second-best converges to zero as n becomes large, it is enough to show that it converges to zero when the budget is E^{*n}. The true social loss in the second-best, with the optimal budget, is even less and also converges to zero. Therefore, we show:

<u>Proposition 2:</u> For each $\delta>0$, there exists $N(\delta)$ such that for all $n>N(\delta)$, $\left| \int SW(G^*(y,n),y,H^n)\, dF(y) - SB(E^{*n},H^n) \right| < \delta$.

<u>Proof of Proposition 2:</u>

<u>Lemma 1:</u> For any $\varepsilon>0$, there exists $n_1(\varepsilon)$ such that, if $n>n_1(\varepsilon)$ and $y \le [f-c\varepsilon]/[f+c]$, then $H^n(G^*(y,n)-y)>1-\varepsilon$.

<u>Proof:</u> Since we have assumed that h is positive on the interior of its support, therefore h^n becomes unbounded for large enough n at any point interior to its support, and it follows that a solution to the first-order condition (2) requires that $H^n(G^*(y,n)-y)$ is close to zero or one, or that $G^*(y,n)$ is close to one. The latter is ruled out, since y is strictly less than one and there is no firm with compliance cost one in the support of H^n, provided n is large enough. Let $G^{**}(y,n)$ represent the solution to (2) for which $H^n(G^{**}(y,n)-y)$ is close to or equal to one. Choose $n_1(\varepsilon)$ large enough that, for $n>n_1(\varepsilon)$, $H^n(G^{**}(y,n)-y)>1-\varepsilon$ and $m/n < \varepsilon$. The net social benefit of inspecting class y with compliance rate $H^n(G^{**}(y,n)-y)$ is greater than $[H^n(G^{**}(y,n)-y)[1-y]-[c/f][f/[f+c]]$, since the average compliance cost of complying firms

[20] $SB(E^{*n},H^n)=\int H^n(G(y,n)-y)dF(y)-\int\int^{G(y,n)} g\,dH^n(G(y,n)-y)dF(y) - c\int [G(y,n)/f]\,dF(y)$.

is no greater than the mean compliance cost y, and the cost of inspections is no greater than the last term, since the maximum compliance cost incurred in class y is less than $f/[f+c]$. Since $H^n(G^{**}(y,n)-y)>1-\varepsilon$, the net social benefit is greater than $[1-\varepsilon][1-y]-[c/f][f/[f+c]]$, which is positive for y as chosen.

On the other hand, if H^n is close to zero, the value of compliance is close to zero, while the cost of inspections is no less than $c[y-(m/n)]/f$ (since the minimum compliance cost in class y is $y-(m/n)$), and therefore, if H^n approached zero, the net social benefit would be negative. It follows that $G^*(y,n)=G^{**}(y,n)$ (that is, $H^n(G^*(y,n)-y)$ is close to one, rather than zero), and the result follows. QED

<u>Lemma 2</u>: For any $\varepsilon>0$, there exists $n_2(\varepsilon)$ such that, if $n>n_2(\varepsilon)$ and if $y>\varepsilon+f/[f+c]$, then $H^n(G^*(y,n)-y)=0$.

<u>Proof:</u> Take $n_2(\varepsilon)=m/\varepsilon$. Then the minimum compliance cost g in such a class exceeds $f/[f+c]$. The net social benefit of compliance in class y is less than $(H^n(G^*(y,n)-y)[1-f/[f+c]]-c[\varepsilon+f/[f+c]]/f)$, which is negative even if $H^n(G^*(y,n)-y)=1$. QED

Turning to the second-best, let G(y,n) be the second-best compliance rate in class y, when the enforcement budget is E^{*n}.

<u>Lemma 3:</u> For any $\varepsilon>0$, there exists $n_3(\varepsilon)$ such that, if $n>n_3(\varepsilon)$, $H^n(G(y,n)-y)>1-\varepsilon$ in inspected classes.

<u>Proof</u>: It follows from (5) that since h^n becomes unbounded everywhere interior to the support, $H^n(G(y,n)-y)$ approaches zero or one. We pointed out in the text that the compliance rate in all inspected classes will be the same. Suppose this uniform compliance rate in inspected classes approached zero. Even though the compliance rate in high-cost classes is close to zero, the frequency of inspections p(y,n) is not; it is larger than $[y-(m/n)]/f$ (which is the inspection rate that gets the lowest-cost firm in class y to comply), and for n large enough, this is larger than $2m/n$. The maximum increment to the inspection rate required to get full compliance in any one inspection class is $2m/n$. Therefore, if compliance rates were close to

zero in all inspected classes, the enforcement agency could reduce the inspection rate in a high-cost class by 2m/n and increase the inspection rate by that amount in another class, thus giving up a very small amount (close to zero) of compliance in the high-cost class, but getting full compliance from the other class to which those inspections were added. We conclude it cannot be optimal for the uniform compliance rate in all inspected classes to approach zero as n becomes large. QED

Let Y^{*n} represent the maximum cost signal in the first-best policy for which the compliance rate is positive. Let $Y^n(E^{*n},n)$ represent the maximum cost signal in the second-best policy (with budget E^{*n}) for which the compliance rate is positive.

<u>Lemma 4</u>: For any $\delta>0$, there exists $n_4(\delta)$ such that, if $n>n_4(\delta)$, $|F(Y^{*n})-F(f/[f+c])|<\delta$ and $|F(Y^n(E^{*n},n))-F(f/[f+c])|<\delta$.

<u>Proof</u>: It follows from Lemmas 1 and 2 that Y^{*n} converges to $f/[f+c]$, and the first result then follows by continuity of F (since we have assumed the distribution of cost signals is atomless). Since the compliance rate in inspected classes in the second-best converges to one, as it does in the first-best, $Y^n(E^{*n},n)$ must converge to the same thing as Y^{*n}, namely $f/[f+c]$. The second result also follows by continuity of F. QED.

<u>Lemma 5:</u> For any $\varepsilon>0$, there exists $n_5(\varepsilon)$ such that, if $n>n_5(\varepsilon)$, $H^n(G(y,n)-y)>1-\varepsilon$ in all classes $y\leq[f-\varepsilon c]/(f+c)$.

<u>Proof:</u> This adds to Lemma 3 that all classes less than $[f-\varepsilon c]/[f+c]$ will be inspected for large enough n. This follows from Lemma 4, since the cutoff $Y(E^{*n},n)$ converges to $f/[f+c]$. QED

We will add up the social loss of the second-best from three nonintersecting groups of inspection classes: (a) a group for which compliance is at least $1-\varepsilon$ in both the first- and second-best, (b) a group with no compliance in either the first- or second-best, and (c) the remaining classes. For $\varepsilon>0$, group (a) will consist of those classes $y\leq[f-\varepsilon c]/[f+c]$. For n larger than $\max\{n_1(\varepsilon),n_3(\varepsilon),n_5(\varepsilon)\}$, the compliance rates for those classes in the first- and second-best policies are greater than $1-\varepsilon$, and therefore the social loss from the difference in patterns of compliance is less than $\varepsilon F([f-\varepsilon c]/[f+c])$. There is no social loss to classes represented by (b). By continuity of F (when F is atomless), for every $\delta>0$, we can choose ε small enough so that $|F([f-\varepsilon c]/[f+c])-F(f/[f+c])|<\delta$. Because of this and Lemma 4, for every $\delta>0$, we can choose $\varepsilon>0$ so that the measure of classes (c), $\max\{|F(Y^n(E^{*n},n))-F([f-\varepsilon c]/[f+c])|, |F(Y^{*n})-F([f\varepsilon c]/[f+c])|\}$, is less than 2δ. The maximum social loss in each such class is at most one (the social value of compliance), so the total social loss in group (c) is less than 2δ. Define γ to be the maximum social loss in groups (a) and (c), $\gamma=\varepsilon F([f-\varepsilon c]/[f+c])+2\delta$. To complete the proof, we need to show that for any $\gamma>0$, there exists $N(\gamma)$, such that, if $n>N(\gamma)$, the social loss is less than γ. Choose $\delta<\gamma/2$, and choose $\varepsilon>0$ so that $\gamma\leq\varepsilon F([f-\varepsilon c]/[f+c])+2\delta$, and so that $|F([f-\varepsilon c]/[f+c])-F(f/[f+c])|<\delta$. Choose $N(\gamma) = \max\{n_1(\varepsilon),n_2(\varepsilon),n_3(\varepsilon),n_4(d),n_5(\varepsilon)\}$. The result follows. QED

APPENDIX B
PROPOSITION 2 WITH FOLLOW-UP INSPECTIONS

In the text we focused on the deterrent effect of inspections and fines. We assumed that firms cannot be brought into compliance during the inspection, because compliance requires a major change in policy or installation of capital equipment. Bringing firms into compliance would require a follow-up inspection, and thus leads us into the difficult area of intertemporal enforcement policies. Here we show that the major result in our paper, Proposition 2 (that the social loss in the second-best is small if the cost signal is a very good predictor of compliance cost), emerges intact if the enforcement agency has a rational intertemporal policy of follow-up inspections to bring firms into compliance.

We assume that, having discovered noncompliance and observed the firm's cost, the enforcement agency will choose ex post whether to promise reinspection.[21] The probability of inspection required to achieve compliance in the second round is g/f, provided the enforcement agency can commit to that probability. But, unlike the first round, in which many similar firms are subject to inspection, and therefore the frequency of inspection is observable, the firm observes in the second round only whether it gets inspected. Since this is a firm-specific matter, "probabilities" are unobservable. We therefore assume that the enforcement agency cannot credibly commit to "probabilities" in the second round. Rather, it commits to inspect or not, and if it commits to inspect, the firm complies if $g < f$.[22]

The argument underlying Proposition 2 had two parts. First, we showed that in both the first- and second-best, there is a set of high-cost inspection classes with no compliance. Second, we showed that as the signal becomes more informative, the compliance rates in

---

[21] An alternative would be that the inspection agency chooses ex ante the probabilities of follow-up inspections for noncompliant firms, and commits to them. Such a policy would not be "subgame perfect" since the agency would want to change its mind after the noncompliant firm is found.

[22] Enforcement costs are thus higher than if the agency could credibly commit to "probabilities" of inspection in the second round, but in both cases, Proposition 2 survives. This is essentially because, as the signal becomes a perfect predictor of cost, almost all inspected firms are compliant, and, since almost no inspected firms are found noncompliant, the difference between inspecting them with probability one or a lower probability does not matter.

inspected classes become arbitrarily close to one in both the first- and second-best. It follows from these two considerations that if the second-best enforcement agency has the optimal inspection budget used by the first-best enforcement agency, the pattern of compliance elicited becomes very close to the first-best pattern of compliance as the signal becomes more informative. Proposition 2 extends to the case of follow-up inspections if these two properties remain. When the budget is fixed by Congress, the second property follows from the first.

We consider first the first-best pattern of compliance elicited by the enforcement agency if it maximizes social welfare. We assume that all compliance costs, g, are less than the fine f, so the firm will comply if it was found noncompliant in the random inspection and expects a follow-up inspection with probability one. A firm will comply voluntarily before the random inspection if $g \le p[f+g]$, or $g/f+g] \le p$. Since $g/[f+g]$ increases with g, inspection frequency $p(G)=G/[f+G]$ will cause all firms $g \le G$ to comply. (We notice that p never needs to exceed 1/2, since all firms with cost g less than f would then comply.) It is optimal to promise reinspection if $1 \ge c+g$; that is, the benefit of compliance exceeds the cost of inspecting the firm to enforce compliance plus the compliance cost. The first-best objective function, for class y, can then be written (analogously to (1)):

(B1) $$H(G-y) + p(G) \int_G^{1-c} (1-c-g)dH(g-y) - cp(G)$$

where $p(G)=G/[f+G]$. If class y has positive compliance, the first-order condition describing $G^*(y)$ is[23]

(B2) $$h(G-y)\,[1-p(G)[1-c-G]] + \left[\int_G^{1-c}[1-c-g]dH(.) - c\right] dp(G)/dG \ge 0$$

with equality if $H(G^*(y)-y)<1$. Now substitute h^n for h. Since $p(G) \le 1/2$ and $dp(G)/dG$ is bounded, (B2) is positive for any G interior to the support of h^n, for large enough n, since h^n becomes unbounded. Hence, compliance in inspected classes approaches one as n becomes large.

23 This is provided $G^*(y)<1-c$. Otherwise, there will be no follow-up inspections, since the second term of (B1) would be negative.

Suppose that for all G, the value of (B1) is negative for class y_1. Since the derivative of (B1) with respect to y is negative for fixed G, the value of (B1) for $y_2>y_1$ is less than the value for y_1 for each G, and therefore, if class y_1 escapes scrutiny, y_2 does as well. There is a cutoff signal Y^* above which no compliance is elicited.

We turn now to the second-best.

In the second-best, the opportunity cost of eliciting compliance with a follow-up inspection is the foregone deterrence benefits from random inspections in the next time period. This cost depends on the budget, E, as set by Congress. We have already shown that with no follow-up inspections, the second-best pattern of compliance has the two properties required for Proposition 2. Therefore, we discuss the other case; that noncompliant firms are brought into compliance. Since the foregone deterrence is v(E)c when cost c is diverted to a follow-up inspection, and the benefit is one, follow-up inspections require $1-v(E)c>0$.

The second-best enforcement problem is then to maximize by choice of G, where $p(G)=G/[G+f]\leq 1/2$,

$$\text{(B3)} \quad \int H(G-y)dF(y) + \int p(G(y))[1-H(.)]dF(y)$$

$$\text{subject to } E = c\int p(G(y))dF(y) + c\int p(G(y))[1-H(.))]dF(y)$$

Class y will have zero compliance if the following expression is negative for all G and p(G), where v(E) is the shadow price of a marginal dollar of enforcement budget, E:

$$\text{(B4)} \quad H[G-y] + p(G)\,[1-H(.)] - v(E)\,c\,p(G)\,[1+[1-H(.)]]$$

Since this expression is declining in y for fixed G, it follows again that if class y_1 is allowed zero compliance, class $y_2>y_1$ is also allowed zero compliance. Again, there is a "cutoff" signal, say Y, above which $p(G(y))=0$, and no firms comply.

If compliance in class y is positive, the first-order condition describing the second-best compliance rate is

$$\text{(B5)} \quad h(.)\,[1 - p(G(y))\,[1-v(E)c]\,] + [1-H(.)]\,p'(G(y))-v(E)c\,[1+[1-H(.)]]\,p'(G(y)) \geq 0$$

with equality if $H[G(y)-y]<1$.

To consider the compliance rates in inspected classes as n becomes large, substitute h^n for h and H^n for H. For any G in the interior of the support of h^n, the value of (B5) is positive for large enough n, since h^n becomes unbounded. Therefore, the compliance rates in inspected classes must converge to one as n grows large. This completes our demonstration of the two conditions for Proposition 2 to hold.