JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT 5, 26–43 (1978)

Firm Behavior Under Imperfectly Enforceable Pollution Standards and Taxes

JON D. HARFORD [1]

Department of Economics, State University of New York at Albany, New York 12222

Received November 2, 1976

Assuming expected profit maximization, the behavior of the firm under imperfectly enforceable pollution standards is examined. Among other results, it is found that cost subsidies can reduce the size of violation and amount of wastes, and that the shape of the expected penalty function determines the direction of the firm response to tighter standards. Under imperfectly enforceable pollution taxes, it is found, among other results, that the firm's actual level of wastes is independent of proportional changes in the expected penalty for pollution tax evasion, and that the marginal cost of actual waste reduction equals the unit tax on reported wastes. Some normative aspects of the results are discussed.

I. INTRODUCTION

In this paper we present simple theoretical models of firm behavior under imperfectly enforceable pollution standards and pollution taxes, respectively. Such models are motivated by the nature of the current and past pollution control efforts in this country and encouraged intellectually by coexisting literatures on approaches to externalities and the economics of crime. Although the model presented here has some strong simplifying assumptions, it presents a point of view distinctly different from most treatments of externalities and allows derivation of some interesting results.

The approach to pollution control adopted by the federal government has by and large been based upon the use of standards, both ambient and emission (or effluent). In the case of air pollution, the Clean Air Act of 1970 gives the Environmental Protection Agency the power to set emission standards on a wide variety of sources with the goal of meeting ambient air quality standards. In the case of water pollution, The Federal Water Pollution Control Act Amendments of 1972 give broad power to the EPA to set effluent standards on practically all industrial and municipal sources with the goal of meeting ambient water quality standards, which, according to legislative intent, are to become increasingly stringent.

In the case of both air and water pollution standards it has been the case that these standards have not always been complied with. For support of this understatement with respect to water pollution controls one should consult Kneese's

[1] The author gratefully acknowledges support from the Research Foundation of the State of New York for the researching and writing of this paper.

0095-0696/78/0051-0026$02.00/0

paper in Bain [2] and Zwick's [15] even less charitable view. In fact, major aspects of the recent water pollution legislation have been aimed at reducing the costs of enforcing pollution standards by making the EPA's power greater and more sharply defined. Both pieces of legislation mentioned have clauses setting penalties for violations of standards.

It may be that once full monitoring of pollution sources is effective, the probability of successfully violating standards without being caught and punished will be close to zero. However, past experience does not support the view that perfect enforcement of pollution standards would be of trivial costs in comparison with the benefits.

Because of the above considerations it is of interest to examine the behavior of a profit maximizing firm under imperfectly enforced pollution standards. However, it is also of interest to examine the behavior of the firm under imperfectly enforceable pollution taxes. One reason would be the substantial literature that has been devoted to proving that under perfect competition the tax per unit of pollution equal to marginal damages with no compensation of victims is the Pareto optimal policy (and that in general virtually all other solutions cannot attain Pareto optimality). Baumol and Oates [4] present a general proof of this proposition and discuss some of the substantial literature that led up to this conclusion. A second reason for studying the effect on the firm of an imperfectly enforceable pollution tax is that it provides us with a comparison set of results for the standards case. We find, in fact, that the pollution tax case does have interesting properties, the main one being a type of separability between the amount of pollution released and the amount of unreported pollution, or tax evasion, that occurs.

If there is a second major federal approach to pollution control it would be that of providing cost subsidies. For firms this is done mostly through the special form of accelerated depreciation allowed on pollution control capital as provided for in the Tax Reform Act of 1969. Such subsidies, by themselves, would seem to provide no positive incentive for pollution control. They merely reduce the cost of doing something which adds nothing to revenues. However, in a world of imperfectly enforceable controls, we can show that such subsidies reduce the tendency to violate standards.

Lastly, we will point out some implications of the analysis for the characterization of optimal levels of pollution control. Since firms that violate standards and evade taxes change the trade-offs in the economy one would expect a change in the set of efficient resource allocations in comparison with a world of perfectly costless enforcement or perfectly compliant firms.

In Section II we will examine the expected profit maximizing firm's response to imperfectly enforceable pollution standards. In Section III we examine the firm's behavior under evadeable pollution taxes. In Section IV we discuss some of the normative implications of the analysis, and Section V presents some conclusions.

II. THE FIRM RESPONSE TO STANDARDS

Let us consider a firm which produces a good x from which it receives revenue $R(x)$. The firm may be either a perfect competitor of some form of imperfect competitor. The costs C of producing the good x depend, not only on the quantity of x, but also on the amount of emitted wastes w. So by definition $C \equiv$

$C(x, w)$. This cost function is taken to reflect all of the process changes which would reduce created wastes and any efforts which would "neutralize" wastes already created. Emitted wastes are those which are created and not neutralized, and all references to firm wastes (or actual wastes) will mean emitted wastes as here defined.

We will indicate partial derivatives by subscripts. Thus, marginal output costs and marginal revenue are C_x and R_x, respectively. Under usual circumstances we would expect $|R_{xx} - C_{xx}| < 0$, where the subscripts indicate second partial derivatives, which we will assume exist for all of the relevant functions unless noted otherwise. In order for the firm to find it profitable to emit pollution in the absence of any external constraints it must be true that C_w is negative over a range from zero pollution up to some level w_o. "w_o" is the point where $C_w = 0$, and would be the amount of pollution released in the absence of any controls. We may assume that C_w becomes positive after this point; that is, it would increase costs to release any more pollution. Of course, the firm would never operate on this portion of its cost curve. Consistent with this description, we will assume that $C_{ww} > 0$. If costs were separable in pollution and output we would have $C_{xw} = 0$, but it is more reasonable to assume that $C_{xw} = C_{wx} < 0$, over the relevant range of the cost function.

Now assume that a particular pollution control board (PCB) or the EPA imposes a particular allowable level of released wastes, labeled w_s, which is less than what the firm would have otherwise chosen. The PCB wishes to see this standard complied with. Therefore it sets penalties in the form of fines (and perhaps prison terms) and expends resources on activities which create a probability of discovering and convicting standards violators greater than zero but less than one. For present purposes, the discovery of a violation, any adjudication process, and the receipt of a penalty for the violation will be treated as one event. For any size of violation, $v = (w - w_s)$, the firm sees the probability of discovery and punishment and the level of fine as parameters.

Generally we would expect that any fine f will be a positive function of the size of violation. It is also plausible that the probability p of detecting a violation will be a positive function of the size of violation. Thus we would suppose that f_v and p_v are ordinarily greater than zero. We will be assuming risk neutrality for the firm, so that the expected fine, $g = pf$, is all that matters to the firm. To facilitate the analysis and reduce the number of terms let us write $g = F_s PG(v)$, where F_s and P are shift parameters for the general size of fines under standards and the general level of the probability of detection and punishment of violators, respectively. Because of their nature, the terms F_s and P can be taken to be equal to one without loss of generality. Thus, $G(v) = f(v)p(v)$, and it is assumed to be continuous and twice differentiable; and

$$G' = f'p + p'f \tag{1a}$$

$$G'' = f''p + 2p'f' + p''f, \tag{1b}$$

where the primes indicate the order of differentiation.

It is conceivable that there could be a discontinuity at $v = 0$. This would require that there be a sudden jump from zero to some positive value in the probability of punishment as one goes from no violation to one of arbitrarily small size *and* that there be some minimum level of fine. We mention this as a possibility since such a discontinuity would tend to produce extreme solutions.

As a practical matter, however, laws are seldom enforced in such a way as to produce a significant discontinuity of this sort. Before going on it should be noted that both f'' and p'' may be negative and still have G'' positive.

The structure of the fines is something to be determined by the PCB and presents an optimization problem. It may be suggested that characterizing the optimal structure of fines is an interesting problem and quite possibly a very difficult one in general. The probability of detection as a function of the size of violation might be either something given in the nature of the production function of detection, or a result of conscious allocation of resources. For a situation involving different types of crimes, the latter interpretation is more plausible (and the concept of crime size less clear) while for crimes of the same type this writer would lean toward the former interpretation. In either case, we assume the firm views these functions as given, and, given risk neutrality, is concerned only with the product of these functions and its derivatives.

In order to complete our picture, let us assume that the firm receives a subsidy at a rate b, where b is between zero and one, for the costs of eliminating wastes. In reality subsidies usually apply only to explicit capital expenditures on pollution control and seldom or never to the costs of all potential ways of reducing pollution, which would include process changes and non-capital inputs into waste neutralization. Indeed, it would be difficult to divide costs between output and pollution control if costs are not additively separable in the technical sense. We will ignore this practical difficulty and assume the total subsidy is $b(C(x, w) - C(x, w_o))$ where w_o is taken to be the level of wastes that the firm would produce in the absence of pollution controls and cost subsidies. Since w_o is a constant, we may define $C(x, w_o) = C^o(x)$.

Given the previous definitions, expected profits, Π_s, is of the following form:

$$\Pi_s = R(x) - C(x, w) + b[C(x, w) - C^o(x)] - PF_sG(v). \tag{2}$$

The firm is assumed to choose output x and the level of waste w so as to maximize expected profits. The assumption of risk neutrality is not perfectly general, of course, but even for nonrisk neutral firms risk neutrality is a good approximation for relatively small risks, which is probably a fair description of most risks associated with violating pollution standards.[2]

The first order necessary conditions for an interior maximum are:

$$R_x - C_x + b(C_x - C_x^o) = 0 \tag{3a}$$

$$-(1 - b)C_w - PF_sG' = 0. \tag{3b}$$

Condition (3a) says that marginal revenue should equal marginal costs less the subsidized fraction of the difference between the marginal costs of output at the base level of waste and the marginal costs of output at the chosen level of

[2] Given the basic tenet of portfolio theory that one can evaluate the risk of any asset only in the context of its relationship to the array of all assets, it is difficult to see how one would make rigorous sense out of the idea of either risk averseness or risk preference for a firm that may be owned by thousands of stockholders with differing portfolios and preferences for risk. The present assumption can at least be defended on the grounds that in the long run (assuming survival of the firm and some version of the law of large numbers) the risk neutral firm will have had greater average profits than either the risk averse or risk preferring firm. To put it another way, if all profits were reinvested for all firms at some constant rate of return, then the risk neutral firms would eventually be the largest.

wastes. If the cost function is separable in x and w, the last term is zero and we get the conventional answer that marginal revenue should equal marginal cost. If $C_{xw} < 0$, we would expect $(C_x - C_x^o) > 0$, and therefore, that marginal revenue would be less than marginal costs. With such a cost function and subsidy one ends up subsidizing output as well as pollution control. This is no doubt part of the reason why most actual cost subsidies are for types of pollution control equipment that have no obvious connection to the production of output. This, of course, means that pollution control is not achieved in a least cost manner from society's viewpoint.

Condition (3b) tells us that the unsubsidized fraction of marginal costs of reducing wastes should equal the increase in the expected monetary value of the fine with a one unit increase in actual waste. Clearly it may often be the case that equality in (3b) does not hold. For example, if both the fine and the probability of receiving a penalty are fixed and independent of the violation size, then $G' = 0$, and $C_w = 0$ will appear to be the maximizing condition, which is what it would be in the absence of standards. However, if the expected fine is high enough one will get the other "corner" solution where $w = w_s$.

The two possibilities for the case just mentioned are illustrated in Fig. 1. We have drawn costs as a function of wastes while holding output constant. For simplicity we have assumed a zero subsidy in this and following figures. Two different fixed sizes of expected fine are indicated by G_1 and G_2. The level of costs under complete compliance is indicated by C_2, and under no effort at compliance the costs are C_1. Since $C_1 + G_1 < C_2$, the firm makes no effort at pollution control under expected fine G_1. Under expected fine G_2, the firm chooses zero violation since $C_1 + G_2 > C_2$.

In general, the possibilities for corner solutions depend upon the relative heights and shapes of the cost and expected fine functions, as well as the levels of w_s and b. Figure 2 shows a case where an interior solution, characterized by $w_o > w^* > w_s$, exists. The full compliance level of costs OM_2 is greater than the costs at the minimum point of the $C + G_3$ curve where wastes are w^* and costs are at height OM_1. It is due to the increasing nature of the expected fine and the decreasing marginal costs of waste prevention that w^* is to the left of w_o.

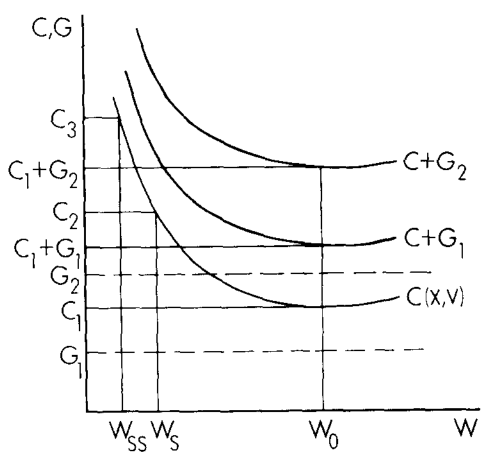

FIG. 1. Choice of violation size with constant expected fine.

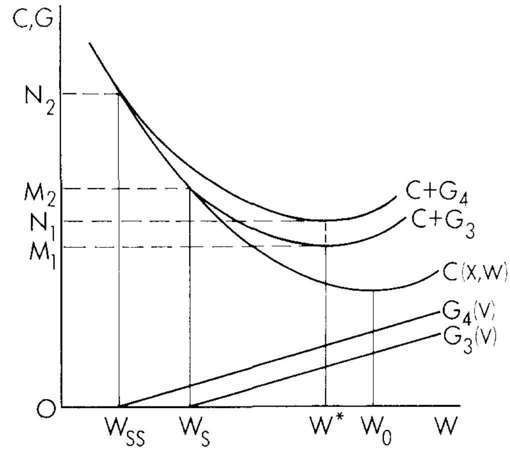

FIG. 2. Choice of violation size with a constant slope expected fine function.

Careful perusal of variations on Fig. 2 indicates that as long as $G'' \geqq 0$, a (further) requirement for a positive violation level is that

$$-C_w \big|_{w=w_s} > G' \big|_{v=0},$$

given a zero subsidy and the other assumptions on the cost function. Since the marginal costs of reducing wastes decrease as wastes increase, while the slope of the expected fine function is here assumed to be non-decreasing, we must have marginal costs of waste reduction at the allowed level of wastes greater than the rate of increase in the expected penalty with respect to the first unit of illegal wastes in order to have a positive violation.

To gain further insight into the possibilities of corner solutions we shall now consider the second order conditions for a maximum. These conditions are that the determinant of the Hessian matrix,

$$H_s = \begin{vmatrix} (R_{xx} - (1-b)C_{xx} - bC_{xx}{}^o) & -(1-b)C_{xw} \\ -(1-b)C_{xw} & -(1-b)C_{ww} - PF_sG'' \end{vmatrix} > 0, \quad (4)$$

and that the terms on the diagonal be negative. This expression has the common implication that the cross partials of the cost function should not be too large in comparison with the repeated partials in w and x. A further implication is that G'' cannot be too strongly negative or the second order conditions will not hold. One may note that the linear expected fine function in Fig. 2 implies that its second derivative is zero, and thus it is consistent with satisfaction of the second order conditions. The satisfaction of the second order conditions ensures us that there is at most one interior maximum, which will be the global maximum if there is no discontinuity in the expected fine function at the zero violation level.

We may now list and discuss a number of comparative static results showing rates of change of controlled variables with respect to parameter changes. In the following results we have taken advantage of the normalization $P = F_s = 1$.

$$\frac{\partial w}{\partial P} = \frac{\partial w}{\partial F_s} = \frac{[R_{xx} - (1-b)C_{xx} - bC_{xx}{}^o]G'}{H_s} < 0 \quad (5a)$$

$$\frac{\partial w}{\partial w_s} = \frac{-[R_{xx} - (1-b)C_{xx} - bC_{xx}{}^o]G''}{H_s} \gtrless 0, \quad \text{as} \quad G'' \gtrless 0. \tag{5b}$$

$$\frac{\partial w}{\partial b} = \frac{-[R_{xx} - (1-b)C_{xx} - bC_{xx}{}^o]C_w + (1-b)C_{xw}(C_x - C_x{}^o)}{H_s} < 0 \tag{5c}$$

$$\frac{\partial x}{\partial P} = \frac{\partial x}{\partial F_s} = \frac{(1-b)C_{xw}G'}{H_s} < 0 \tag{5d}$$

$$\frac{\partial x}{\partial w_s} = \frac{-(1-b)C_{xw}G''}{H_s} \gtrless 0 \tag{5e}$$

as $G'' \gtrless 0$, assuming $C_{xw} < 0$.

$$\frac{\partial x}{\partial b} = \frac{(C_x - C_x{}^o)[(1-b)C_{ww} + G''] - (1-b)C_wC_{xw}}{H_s} \gtrless 0. \tag{5f}$$

These results are based on the assumption that both the first and second order conditions for a maximum are satisfied. On this basis the relationships in (5a) tell us that wastes created and released will decrease with increases in the level of fines or increases in the probability of detection at all sizes of violation. In other words, a general increase in expected fines decreases the actual level of wastes and the size of violation which the firm chooses. We also find from (5b) that an increase in the cost subsidy will reduce the size of any standards violation, a result due basically to the fact that the relative cost of complying with the standard has been reduced. This provides a rationale for the commonly used cost subsidies, which economists have tended to ignore or presume to be useless in situations where none of the benefits of reducing pollution are internal to the firm.

On the output side we find that increased fines and probability of detection tend to reduce the expected profit maximizing level of output to the extent that marginal costs of output are increased by reductions in the emitted wastes. Thus, without separability we would tend to find that limitations on the amount of wastes would tend to reduce the profit maximizing level of output for the firm.

Interestingly enough, the direction of change in waste and output levels with respect to a change in the allowable wastes under the standard is dependent upon G'', the second derivative of the expected fine function with respect to the size of the violation. If the slope of G is increasing, then waste and output will move in the same direction as allowable wastes under the standard. This is the more intuitive result. One may conjecture that it is likely though not certain that this would be the most relevant case. If $G'' > 0$, it can be plausibly argued from the relevant expressions that the increase in wastes will be less than the allowable wastes under the standard, i.e., that the size of violation will decrease with a loosening of the standard.

However, there is nothing in the second order conditions which definitely rules out the case where the slope of the expected fine function is decreasing. In this case we find that making the standard marginally stricter will increase the actual amount of waste released by the firm. To give the reader a better feel for the situation we refer to Fig. 2, which illustrates what occurs in the border line case where the slope of the expected penalty function is constant. The shift in the wastes standard from w_s to w_{ss} has the effect of shifting upward the curve representing the sum of costs plus the expected penalty to $C + G_4$. Thus costs at w^* increase from a height of OM_1 to ON_1. However, due to the parallel nature of the shift, ON_1 now represents the new minimum of the sum of costs plus expected penalty. In other words, there is no change in actual wastes emitted. The size of the violation increases in the same magnitude as the reduction in the allowed level of wastes.

Finally, we note that the change in output with respect to a change in the cost subsidy rate is in general ambiguous in sign. This is due to the existence of two counteracting effects. The subsidy reduces the cost of pollution control which allows more output for the same pollution control efforts and total costs. However, because the subsidy encourages less pollution it indirectly increases the marginal costs of output. Which effect dominates depends upon the size of practically all of our parameters and the first and second partial derivatives of the cost function. In the special case of separability between output and wastes in the cost function, we can assert that the change in output with respect to a change in the subsidy rate will be zero.

All of these comparative static results are under the assumption that the first order conditions are satisfied with equality and the second order conditions hold. If, on the other hand, we do not have an interior solution, we would still expect that small changes in the parameter values would have either no effect, or effect waste and output levels in the direction indicated by the comparative statics results.

This expectation tends to hold up even with respect to variations in w_s in the case where complying with the standards maximizes expected profits. One might be tempted to expect that permitted and actual wastes would move in the same direction regardless of the shape of the expected fine function. That this may not be the case is illustrated within Fig. 1. When the allowable level of wastes is decreased from w_s to w_{ss} under the assumption that the expected fine is a constant G_2 we find that overall expected costs are now lowest at the level of w_o wastes, i.e., costs at full compliance with the stricter standard C_3 are greater than $(C_1 + G_2)$. Thus, in this case, increasing the strictness of the standard increases the actual level of wastes from w_s to w_o. To put it another way, the firm may go from full compliance to totally ignoring the standard as it increases in strictness if the expected fine does not increase rapidly enough with the size of violation. For the special case illustrated by Fig. 1, Anderson [1] preceded us in discovering this result.

All of the previous results are about individual firms and, therefore, one needs to know something about industry structure in order to discuss effects of pollution standards on aggregate of firms. For firms operating in an environment where expected long run economic profits should be zero, average costs become important for determining the amount of wastes which the industry will generate since average costs determine the size of the industry in terms of output. Thus, even in the seemingly perverse case where tightening standards increases the firm's waste creation, we would expect (and Figs. 1 and 2 confirm) that the firm's average costs would increase. This would tend to reduce the size of a competitive industry and work in the opposite direction from the "counterproductive" effect on the individual firm.

In the same vein, an increase in the cost subsidy rate for pollution control would be expected to reduce the average costs of output for the firm even though the marginal effect is uncertain. Thus the size of the industry would be larger with the subsidy, ceterus paribus, even though each individual firm may produce less pollution. It is possible that the larger industry size would increase pollution relatively more than any reduction in the pollution per unit of output that the cost subsidy might induce. The overall effect will depend upon a number of factors including the elasticity of the market demand for the output.

III. EVADEABLE POLLUTION TAXES

It is of interest to compare the kind of results we obtain with pollution standards with those we would get in the case where a tax per unit of waste emitted is imposed. To make the comparison reasonable, it is assumed that the pollution tax is evadeable. This may be interpreted as meaning that actual wastes released, w, are greater than the reported wastes, w_r, where the tax is directly on reported wastes only. The firm's costs C are a function of ouput x and the actual wastes released w in the same manner assumed in the case of standards. (We no longer assume any cost subsidy.) All of the remarks about the signs of the derivatives of the cost function still apply.

The unit tax t is applied to w_r so that total pollution tax paid equals t times w_r. In order to prevent the firm from evading all taxes by reporting zero released wastes, the pollution taxing agency imposes penalties for evasion of pollution taxes. The size of the evasion will be measured by $v = (w - w_r)$, where we re-use the letter v to reflect the size of violation in this different case. There is some justification for this procedure. In the cases of standards, the reported wastes would always be w_s since there is no incentive to have actual or reported wastes less than w_s, and reported wastes above w_s would be asking for a penalty. So even in the standards case the violation size v could be interpreted as the difference between actual and reported wastes.

Again we will use the function $G(v) = pf$ to reflect the shape of the expected fine as a function of the violation size where p and f have the same meaning as before. The shift parameter of the probability of detection and punishment will again be labeled P, which will be taken to be equal to one. We shall distinguish between two components of what we shall call the expected penalty function. The first component consists of a fine shift parameter F, also taken to be equal to one, multiplied by $PG(v)$ to reflect the expected fine size. The second component will be the tax that will be collected on unreported but discovered wastes. Thus, in addition to any expected fine and the tax on reported wastes, there is an expected tax on unreported wastes of $Ppvt$. To reduce the number of terms in later expressions we will define $B = pv$. We note that

$$B' = p + p'v \tag{6a}$$

$$B'' = 2p' + p''v. \tag{6b}$$

If we again use $R(x)$ to denote total revenue as a function of output, we may write expected profits as

$$\Pi_t = R(x) - C(x, w) - tw_r - FPG(v) - PB(v)t. \tag{7}$$

The firm now has three control variables, x, w, and w_r, with which it attempts to maximize expected profits. Accordingly, we will have three first order necessary conditions for a maximum, which are:

$$R_x - C_x = 0, \tag{8a}$$

$$-C_w - FPG' - PB't = -C_w - G' - B't = 0, \tag{8b}$$

$$-t + FPG' + PB't = -t + G' + B't = 0. \tag{8c}$$

Condition (8a) informs us that marginal revenue should equal marginal cost, the same type of result one would get if there were no pollution tax (although the level of marginal costs will be affected by the actual level of pollution control). Condition (8b) indicates that the marginal cost of reducing actual waste should be equal to the marginal increase in the expected penalty (expected fine plus additional tax) with a unit increase in violation size. Condition (8c) tells us that the increase in expected penalty from a unit decrease in reported wastes should equal the unit tax on reported wastes.

If these conditions hold, it is obvious but interesting the $-C_w = t$. In other words, the marginal cost of actual pollution reduction by the firm will equal the unit tax on *reported* pollution. This implies that the actual waste level of the firm does not directly depend upon the size of our shift parameters for the fine or the probability of discovery of the violation. Furthermore, if the expected punishment levels are generally so high that the firm maximizes expected profits by having actual and reported wastes the same [$(G' + B't) > t$ at all positive v would cause this situation], then it would still clearly be true that the marginal costs of waste reduction would equal the unit tax on reported wastes, since reported and actual wastes are the same.

At the other extreme of corner solutions, however, this result would not hold. If expected punishment costs are sufficiently low and increase so slowly that the unit tax on reported wastes is everywhere greater than the increase in expected penalty with respect to violation size, one would get a case where reported wastes are equal to zero, and the marginal costs of actual waste reduction would be equal to the smaller magnitude of the increase in expected penalty costs.

Practically speaking, corner solutions where reported wastes are zero are unlikely. It would be irrational to set penalties so low that no pollution tax at all was collected. Moreover, if it is obvious that every firm generates some wastes, reporting zero wastes would be a clear signal of a violation.

Deriving the Hessian matrix of second order derivatives we can characterize the second order conditions for a maximum in terms of restrictions on the signs of its determinant and principal minors. We have the restrictions that the determinant

$$H_t = \begin{vmatrix} (R_{xx} - C_{xx}) & -C_{xw} & 0 \\ -C_{xw} & (-C_{ww} - A) & A \\ 0 & A & -A \end{vmatrix} < 0, \tag{9}$$

where $A = G'' + B''t$, and that all the diagonal elements must be negative, while all second order minors should be positive. These restrictions imply that the expression represented by A must be positive for an interior maximum to be unique, and a global maximum if there is no discontinuity in the expected penalty function at the zero violation level. Thus, the expected penalty function must have an increasing slope to satisfy the second order conditions and provide the possibility of an interior solution in the pollution tax case, whereas the expected fine function did not have to have a positive curvature in order for an interior solution to exist. It is still true, however, that the expected fine function, by itself, does not have to have a positive curvature if the function B has a sufficiently strong positive curvature.

Although we have a larger matrix, the comparative statics analysis turns out to be simpler in some ways than in the standards case. Equating the shift parameters to one we have the following results:

$$\frac{\partial w}{\partial P} = \frac{\partial w}{\partial F} = 0 \tag{10a}$$

$$\frac{\partial x}{\partial P} = \frac{\partial x}{\partial F} = 0 \tag{10b}$$

$$\frac{\partial w}{\partial t} = \frac{-(R_{xx} - C_{xx})A}{H_t} < 0 \tag{10c}$$

$$\frac{\partial x}{\partial t} = \frac{-C_{xw}A}{H_t} < 0 \tag{10d}$$

$$\frac{\partial w_r}{\partial F} = \frac{G'[(R_{xx} - C_{xx})C_{ww} + C_{xw}^2]}{H_t} > 0 \tag{10e}$$

$$\frac{\partial w_r}{\partial P} = \frac{(G' + B't)}{G'}\left(\frac{\partial w_r}{\partial F}\right) > 0 \tag{10f}$$

$$\frac{\partial w_r}{\partial t} = \frac{-[(R_{xx} - C_{xx})C_{ww} + C_{xw}^2](1 - B')}{H_t} + \frac{\partial w}{\partial t} < 0 \tag{10g}$$

$$\frac{\partial v}{\partial t} = \frac{\partial w}{\partial t} - \frac{\partial w_r}{\partial t} > 0 \tag{10h}$$

First of all, quite consistently with previous statements, we find that the actual level of wastes released is independent of the level of fines or probability of detection and punishment. Small changes in P and F have no effect on w. Furthermore, small changes in P and F have no effect on the firm's level of output. This follows quite intuitively; if the optimal level of actual waste does not change, then costs as a function of output would not change, and, therefore, marginal revenue would equal marginal costs at the same output level both before and after any shift in fines or the probability of punishment.

We find that actual wastes decline as the tax on reported wastes is increased. This occurs because of the connection of both actual and reported wastes to the rate of change in the expected penalty with respect to violation size. If costs are not separable in output and wastes, we find that output is a negative function of the tax on wastes. If costs are separable, then the firm's output will not be directly affected by the tax on reported wastes.

Result (10f) indicates that increases in the probability of capture at all sizes of violation will increase reported wastes. Since actual wastes are not affected by such a parameter shift, this implies a reduction in violation size at the same absolute rate. Using this result and explicitly writing out and simplifying (10f), one can derive the following relationship between the elasticity of violation size with respect to the general probability of detection, and the elasticity of the slope of the expected penalty function.

$$\left(\frac{\partial v}{\partial P}\right)\frac{P}{v} \equiv e_{P_v} = \frac{-1}{e_{vz}} \equiv -\left(\frac{Z}{Z'}\right)\frac{P}{v} \tag{11}$$

where $Z = G' + B't$, and use is made of the normalization $P = 1$.

This relationship says that the elasticity of violation size (in absolute value) with respect to a proportional shift in the probability of punishment is inversely related to the relative positive curvature of expected punishment as a function of violation size. To say it another way, if the rate of increase in the increase in expected punishment with violation size is relatively small, then a proportional increase in the probability of being fined at all levels of punishment will elicit a relatively large decrease in violation size.

The rate of change in reported wastes with respect to a change in the tax on reported wastes may be broken up into two parts: A term exactly equal to the effect on actual wastes, and a term which reflects the direct tax and penalty effects of a tax shift. It has already been determined that an increase in the pollution tax reduces actual wastes, and accordingly this component reduces the reported wastes to the same extent. The other term in (10g) will be negative if (1-B') is positive. An examination of the first order condition (8c) indicates that this must be true if the slope of the expected fine function is positive, which we assume to be the case. Since both terms are negative, reported wastes will decline to a greater extent than actual wastes when the pollution tax is increased. This, as (10h) indicates, means that an increase in the pollution tax increases the size of the violation that the firm chooses. As one might expect, raising the tax rate causes more evasion and presumably a more difficult enforcement problem.

Given the insensitivity of actual wastes to the general level of fines and the probability of punishment, one might be tempted to conclude that their exact magnitude is of little or no consequence over a considerable range. This idea has a rough validity for all firms in the short run, and firms in noncompetitive industries in the long run. It is definitely not true in the long run for those firms in industries where expected long run economic profits tend toward zero. In these industries it is the long run average costs which determine the industry size in terms of output. A lower level of enforcement of payment of pollution taxes may not change the actual level of wastes emitted by any firm, but the implied lower level of some combination of pollution taxes, and penalty payments will imply lower average costs and more firms in the industry. In this regard it should be noted that, ceterus paribus, firms with the same expected fine function will have lower average costs under the imposition of standards than under the imposition of pollution taxes. It is roughly on this point, as Buchanan and Tullock [7] explain, that standards fail to produce a Pareto optimal (or pollution tax equivalent) resolution to a pollution externality under perfect competition.

This model may not be irrelevant to the current situation in the area of water pollution control. The recent federal legislation states that municipalities must charge fees to firms for accepting their wastes, fees which must be designed to cover the costs of treating those wastes. If the firm were not legally allowed to channel any wastes directly into public waters, then, with proper interpretation, this model might almost directly apply. The fee for municipal acceptance of firms' effluent would correspond to the tax on reported, in this case delivered, wastes. Violations might take the form of illegal dumping of wastes in the river, or channeling more wastes to the municipal treatment plant than the firm pays for in effluent fees. (More complicated and subtle versions of the same problem might involve the firm reneging on pretreatment requirements or negotiating reductions in other types of taxes (such as the local property tax) to compensate for the newly instituted effluent fees.) Even if the firm is allowed to release a fixed level of wastes, w_s, directly into public waters, we may simply define the size of violation as $v_o = [w - (w_s + w_r)]$, and the previous results are not qualitatively altered, although chosen values of the control variables would differ.

If one modifies the definition of the violation size as just suggested, one gets the additional results that $\partial w_r/\partial w_s = -1$, and $\partial w/\partial w_s = 0$. The former result says that there would be a one to one reduction in reported waste (subject to effluent fees by the municipality) with any increase in the allowable standard of wastes disposable directly into public waters. Put another way this result is quite striking. It tells us that making the standard stricter would not increase the violation size, but simply cause an increase in the amount of wastes going to the treatment plant and *paying* the effluent fee. The latter result indicates that changing the standard would have no effect on the actual amount of wastes created, which is in contrast to the pure standards case, but quite in spirit with our results in the pollution tax case.

Inevitably one runs into difficulty in trying to fit reality to a model. Relevant here is the fact that legislatively described fines are often in terms of days of violation rather than being geared to the quantity of pollutant. Even so, or perhaps because of this, the actual level of fines is left to the discretion of the courts within the constraint of maximum and minimum fines.[3] This methodology could produce its own peculiarities. Another complicating factor is the actual variety and interactability of real world wastes. And, with at least two, and perhaps three, levels of government (with varying views of benefits and costs) participating in pollution control with regard to the same set of potentially geographically mobile firms, the actual situation becomes extremely complicated.

IV. NORMATIVE IMPLICATIONS

The analysis heretofore has been predictive in the sense of simply drawing out the implications for a firm's behavior under expected profit maximization. We will now attempt a discussion of the normative implications of an analysis that includes the possibility of evading the pollution taxes, which are imposed pre-

[3] For example, within Title III, Section 309 of the Federal Water Pollution Control Act Amendments of 1972 it is stated that any firm or municipality (or the persons responsible) violating an applicable pollution standard ". . . shall be punished by a fine of not less than \$2250 nor more than \$25,000 per day of violation, or by imprisonment for not more than one year, or by both." Except for the expectation that the courts will use their discretion wisely in gearing fines to the actual damages caused by the pollution, this would appear to be an unsatisfactory way to relate fines to the severity of the violation.

sumably to reduce an externality to an efficient level. The discussion will speak in terms of a pollution tax rather than a standard, since it is the former which has been most favored by economists for its potential to improve efficiency. We will use the intuitive, but not completely rigorous, concepts of damages and costs measured in dollar values. Hopefully this problem may be treated within a rigorous general equilibrium framework at some future point. At this point, however, the technical problems with more precise approach appear too complex to explore.

The results derived indicate that the larger the pollution tax, the greater the evasion of that tax. This points to increasing enforcement costs as attempts are made to reduce actual pollution. We therefore have increasing marginal costs of reducing pollution which include additional costs of enforcing pollution taxes (or standards). This would argue for a reinterpretation of the rule that pollution should be eliminated to the point where marginal damages are equal to the marginal costs of treatment or prevention. Under present assumptions we should have marginal damages equal to the marginal costs of treatment or prevention plus the marginal costs of enforcing the treatment or prevention.

This point can be argued somewhat more rigorously by a modification of Becker's [5] approach to the optimal level of crime and law enforcement, which is to minimize the sum of the damage caused by crime plus the costs of capturing and punishing criminals. In this context we wish to minimize the sum of the damages caused by pollution plus the costs of treating or preventing pollution plus the costs of enforcing the taxes or other instruments by which pollution control is encouraged.

Let us define the value of the damage done to society by an aggregate level of waste W as $D(W)$. On the basis of previous analysis we would expect aggregate wastes to be a decreasing function of the unit pollution tax t. It is also consistent with previous results to assert that the aggregate level of waste should be a decreasing function of the general level of fines F and the general probability of detection and punishment for an evasion of pollution taxes, P. (P and F are now viewed as control variables of the government rather than as parameters.) Previous results indicate that the later two factors may ordinarily have no effect on the actual level of the firm, but, as we have argued before, expected punishment levels will effect the number of firms (and thus aggregate wastes) in a competitive industry via the effects on expected average costs. The costs of removing or preventing wastes are an increasing function of how much lower wastes are than than they would be if no efforts were made to incur or prevent wastes. This base level of wastes will be labeled W_o, while the aggregate costs of neutralizing wastes will be $S = S(Y)$, where $Y = (W_o - W)$.

Costs of enforcement of the pollution tax, E, will be an increasing function of the probability of capturing violators, P, and an index of the number of violators, V. No analysis of the shape of the function relating the probability of detection and the size of the violation will be attempted. However, the variable V is taken to reflect some index of the number and size of violations, which would in general depend upon the ease of detecting violations of different sizes. The level of the index of violations will itself be a negative function of the probability of detection and punishment, and the general level of fines F, but it should be a positive function of the height of the unit tax t on reported wastes. Thus, we have $E = E[P, V(P, F, t)]$.

Lastly, consistent with Becker, it will be assumed that there are social costs of punishment. These social costs are assumed to be a linear function of the general probability of punishment times the general magnitude of punishment. Specifically, total punishment costs are hPF, where h is a parameter which reflects general population size and other factors. Again, for simplicity, we avoid consideration of the structure of fines over the violation size. Given the fact that punishment in this context is likely to be a fine, the assumption that there are non-negligible social costs to punishment may not be very appealing to some. Harris [9] suggests another rationale for social costs of punishment: that of wrongful punishment. Occasionally, the innocent may be fined and this can be looked upon as causing social costs. On a more general note Stigler [12] points to the idea of marginal deterrence rather than social costs of punishment as the critical factor in the determination the appropriate level of punishment for any crime. This may be interpreted as saying that the higher the punishment for a given type of crime the greater the occurrence of other types of crimes with their attendant social costs. Further discussion of how, and how well, these rationales serve in arguing for social costs of punishment is not warranted here, except to say that Stigler's line of thought quickly leads us back to the tricky problem of determining an optimal structure of punishment.[4] For the moment we will simply work with our simple assumption.

We may now write the sum of pollution damage costs plus the costs of treatment/prevention plus the costs of enforcement of pollution taxes as

$$L = D(W) + S(Y) + E(P, V) + hPF. \tag{12}$$

The social goal is to minimize total social costs through the choice of F, P, and t. The three first order necessary conditions for minimum social costs are: where the subscripts indicate partial derivatives.

$$(D_W - S_Y)W_t + E_V V_t = 0 \tag{13a}$$

$$(D_W - S_Y)W_P + E_P + E_V V_P + hF = 0 \tag{13b}$$

$$(D_W - S_Y)W_F + E_V V_F + hP = 0, \tag{13c}$$

For present purposes it is condition (13a) that is of main interest, and so we will forego any comment on conditions (13b) and (13c), except to say that they indicate the usual balancing of marginal damages and marginal costs as they relate to the variables P and F, respectively. Condition (13a) can be rearranged to read

$$D_W = S_Y - (E_V V_t)/W_t. \tag{14}$$

[4] If one examines the concept of marginal deterrence closely it becomes evident that, unless there is some maximum feasible costless punishment, the need for marginal deterrence alone will not always lead one to a finite punishment for any crime. The basic reason for this is that one can mathematically conceive of infinite *rates* of punishment. In the context of this paper, one might have an infinite amount of fine per unit of violation and thus preserve marginal deterrence without having finite punishments. Clearly, in any practical sense, there is a maximum feasible costless punishment. Assuming for the moment that fines have zero social costs, it is clear that any fine cannot exceed the economic worth of the firm or individual fined. Given that there is a maximum feasible costless punishment, the concept of marginal deterrence should enable one to reasonably consider what an optimal structure of punishments should be.

Given our assumptions about the various functional relationships involved, the expression to the right of the minus sign is itself negative. This implies that the damages caused by one more unit of waste should be greater than the costs of physically eliminating the unit by the amount of the extra expenditure on enforcement induced by the increase in tax evasion caused by the additional pollution tax required to reduce pollution by the last unit.

Of course, a full assessment of the normative implications of pollution controls must recognize that the imperfections may be just as great in enforcing non-pollution types of taxes and controls. Imperfections in controls are simply another set of factors to consider in forming policy toward the existence of pollution types of externalities, along with market structure, various uncertainties, and dynamic considerations.

V. CONCLUSION

This paper has presented models of an expected profit maximizing firm under imperfectly enforceable pollution standards and under imperfectly enforceable pollution taxes. We have found that under standards increasing the expected level of the penalty will reduce the level of wastes released by the firm, but that increasing the strictness of the standard will only reduce the firm's wastes if the slope of the expected fine function with respect to the size of the standards violation is increasing. We have also found that the use of cost subsidies for pollution control expenses can serve the useful function of reducing the level of the firm's violation of the standard and thus its actual level of wastes.

Under imperfectly enforceable pollution taxes the analysis has established the neat result that the marginal costs of pollution reduction by the firm will be equated to the constant rate of the pollution tax as long as the slope of the expected penalty function is increasing. This implies that on the firm level the amount of pollution tax evasion is independent of the actual wastes, which are determined by the pollution tax rate. The level of tax evasion is determined by equating the increase in expected penalty with respect to a unit reduction in reported wastes to the decrease in tax paid. Therefore, the general level of fines and the probability of punishment affect reported wastes but not actual wastes. Further results indicate that increasing the tax rate on reported wastes will reduce actual wastes released by the firm, but it will reduce reported wastes even more, implying an increase in pollution tax evasion.

We have also suggested that, with proper reinterpretation of the violation size as the difference between actual wastes and an allowable standard of wastes plus an amount of (reported) wastes going to a fee charging treatment plant, one can apply the pollution tax model to current situation in the area of water pollution. Under this reinterpretation, changes in the allowable standard of wastes has no effect on wastes leaving the firm, or violation size, but are offset exactly by the amount of wastes going to the treatment plant.

In considering the implications of these results for policy it has been mentioned that effects of various policies on the expected average costs of firms are important and may not always work in the same direction as the effect on the individual firm's marginal decision. Furthermore, in Section IV, we suggested a reinterpretation of the intuitive rule of efficiency in pollution control that the marginal damages of pollution should equal the marginal costs of eliminating pollution. The version which this paper suggests is that the marginal damages should be equal to the marginal costs of physically eliminating the pollution plus the additional costs of enforcing any pollution control instrument to the extra degree required to induce the unit reduction in pollution.

Throughout this paper we have assumed that the pollution control agency faces firms which believe that they cannot effect the penalty structure or subsidy rate of the agency. If the number of different polluters is small, this assumption may not be valid. Firms may adopt various kinds of strategic behavior and threats in order to affect the penalty structure. They may threaten to go out of business, or to move to a region where the agency does not have authority. Firms might collude to violate pollution control laws simultaneously, thereby overloading the agency's ability to enforce its laws with any effectiveness. The agency may be able to adopt various counterstrategies in this regulatory duopoly game. The outcome of such a situation is not easily determined, but there are counterparts to it in many types of regulation. Assuming that it is desirable to minimize the possibilities of such strategic behavior on the part of firms, it appears preferable to formulate and administer pollution controls at the most inclusive level of government possible.

This paper raises a number of (other) unsolved problems. One of the more interesting ones is the development of the concept of an optimal structure of penalties. This will likely involve an examination of both the structure of fines and penalties per se, and the technical and resource allocational nature of the probability of detecting and punishing violations of different sizes. Another problem would be to develop a more explicit and rigorous model to analyze the optimal level of pollution, pollution taxes, and enforcement of pollution taxes than that developed in Section IV of this paper. That analysis might be interpreted as suggesting that one should have a lower pollution tax rate in a case where there are enforcement problems than when there are not, but such a conclusion is not explicit, and we suspect that it may not always be true. A more explicit analysis would clarify this and other significant ambiguities.

REFERENCES

1. Robert J. Anderson, Jr., Environmental management economics, *in* "Regional Environmental Management: Selected Proceedings of the National Conference," (Coate and Bonner, Eds.), pp. 201–208, John Wiley and Sons, New York (1975).
2. J. S. Bain (Ed.), "Environmental Decay," Little Brown and Company, New York, (1973).
3. W. J. Baumol, On taxation and the control of externalities, *Amer. Econ. Rev.* **62**, 307–322 (1972).
4. W. J. Baumol and W. E. Oates, "The Theory of Environmental Policy," Prentice-Hall, Englewood Cliffs (1975).
5. G. Becker, Crime and punishment: An economic approach, *J. Polit. Econ.* **6**, 169–217 (1968).
6. J. M. Buchanan, External diseconomies, corrective taxes, and market structure, *Amer. Econ. Rev.* **59**, 174–177 (1969).
7. J. H. Buchanan and G. Tullock, Polluters profits and political response: Direct controls versus taxes, *Amer. Econ. Rev.* **65**, 139–147 (1975).
8. M. E. Darby and E. Karni, Free competition and the optimal amount of fraud, *J. Law Econ.* 67–88 (1973).
9. J. R. Harris, On the economies of law and order, *J. Polit. Econ.* **78**, 165–174 (1970).

10. S. Rose-Ackerman, The economics of corruption, *J. Pub. Econ.* **4**, 187–203 (1975).
11. W. Schulze and R. C. d'Arge, The Coase proposition, information constraints, and long-run equilibrium, *Amer. Econ. Rev.* **64**, 763–772 (1974).
12. G. J. Stigler, The optimum enforcement of the laws, *J. Polit. Econ.* **78**, 526–536 (1970).
13. U. S. Congress, "Clean Air Amendments of 1970," (PL91-604) USGPO, Washington, D. C. 1970.
14. U. S. Congress, "The Federal Water Pollution Control Act Amendments of 1972," (PL92-500) USGPO, Washington, D. C. (1972).
15. D. Zwick, "Water Wasteland," Bantam, New York (1972).