5 Regulatory Strategies for Pollution Control

A. Michael Spence and
Martin L. Weitzman

Introduction

The debate over regulatory policies to control pollution has been dominated by advocates of both effluent standards and effluent charges. This paper argues for a combination of the two that achieves the objectives of each but is more finely attuned than either alone to the practical problems that are often encountered in regulation. The conventional case for effluent charges makes some good sense, but it is based on an unrealistic assessment of the amount of information actually available to regulators, information about the sources, extent, and costs of pollution. It therefore misrepresents the practical options actually open to government regulatory bodies. On the face of it, the case for effluent standards might appear to be stronger. Yet, as we shall argue, the strategy for setting standards can also be improved.

The regulatory problem, in general terms, is to constrain sectors of the economy to achieve the "right" reductions in pollution levels. At the same time, however, the regulatory authority must hedge against the major sources of uncertainty that arise when limited information about polluters and effluents is available. Dealing effectively with this uncertainty involves building into the regulatory process sufficient flexibility to avoid the worst consequences of mistakes. This paper argues that a combination of effluent charges and standards will be the most effective policy.

The problem of uncertainty is not inconsequential. Pollution entails substantial adverse effects for society. But the costs of cleanup or the costs to the industrial sector of equipment to reduce effluents at their source are also substantial. Any mistake about these relative costs and the implied appropriate regulatory policy can cost the economy as a whole many millions of dollars. Because of the magnitude of resources involved, it is important to devote a sufficient amount of effort to ensure that the incentives faced by the regulated sector are consistent with social objectives.

Economists tend to think of pollution as a problem that results from a market failure. That failure is the complete absence of a market for the effects that polluters have on those adversely affected by pollution. (A market, in this context, should be thought of as an arrangement in which people pay for the things they do that affect others.) The fact that markets in the effects of effluents do not and are not likely to exist implies that the costs of pollution are *not* necessarily a factor in the decisions of the polluters. That implication, in turn, tends to result in excessive effluents.

Economists do not just identify the problem as one of a missing market. With a certain rationality, we also incline toward fixing the problem with something that looks like and acts like a market, that is, a system in which prices are attached to effluents. It is usually called an effluent charge or an effluent fee system. The effluent charges are imposed. The regulated sector is then allowed to adjust to them by some combination of paying the charges and reducing effluents.

To most people, the identification of the pollution problem as one of a missing market or a market failure seems odd, to say the least. There never was a market in pollution, and there is not likely to be one in the near future. Pollution, for many people, is a problem like that of preserving wilderness areas or keeping the streets clean. It has little to do with markets, potential or actual. Opponents of prostitution do not react favorably to proposals to control that industry by imposing (perhaps sizeable) excise taxes that approximate the social and moral costs of the activity. There has been a similar feeling about effluent charges as a response to the degradation of the atmosphere. Recently, opposition to effluent fees by regulators and environmental groups seems to have declined somewhat.

To the extent that the presumption in favor of effluent charges is based upon the view that the underlying problem is a missing market, the case has not been convincing to the majority of the public or to noneconomist professionals such as engineers and lawyers. Moreover, public acceptance is not entirely an academic matter. Many of the major regulatory efforts in the United States have been fashioned and modified in the Congress. The general form of the regulations does not usually emanate from a technically oriented agency like the EPA.

Those who mistrust the effluent charge approach typically favor effluent standards. These are maximum levels of effluents that are deemed acceptable and consistent with the maintenance of the quality of the environment in which people live. They are set by the political process and met by the regulated sector. This approach to the regulation of pollution corresponds much more closely to most people's perception of the problem. There is a collective decision about what is and is not acceptable conduct, and the government's task is simply to enforce the laws. The setting of standards is

the major perceived alternative to effluent charges. Thus far, standards have been winning the battle in the political arena.

The economists' case for effluent charges is not entirely based upon the missing market hypothesis. To suggest that it is would be to do an injustice to its proponents. The case is often amplified and buttressed by other arguments. Standards are rigid, at least in principle, and therefore are insensitive to costs. Of course, if costs turn out to be much higher than expected, the standards can be relaxed. But the effluent charge advocate would argue that this type of relaxation occurs automatically and in a controlled way with an effluent charge. Moreover, the knowledge that a standard may be relaxed can create an incentive for the regulated to create the impression that costs are or will be high.

It is argued further that standards distribute the cleanup among polluters in ways that are potentially inefficient. What one would like is a system that causes the sources with the lowest cleanup costs to do most of the cleaning up. Effluent charges do distribute cleanup activity among sources efficiently.[1] Whatever the level of effluents actually achieved, it is achieved at least cost.

There is a final argument. It is that the government should take the position of standing in for the public, whose interest is not represented in the absence of regulation. The public interest is properly represented when the additional benefits of reduced effluents are commensurate with the additional costs that result from effluent reduction. The public, after all, pays these costs in the end, in the form of higher prices, displacement of jobs, and so on. Effluent fees, it is sometimes argued, are a reasonable way of putting the public benefit into the equation. The fees "represent" the benefits to the public of reduced effluents or, equivalently, the costs of pollution to the public.[2] By contrast, it is said, rigid standards seem to imply that the social costs of exceeding the standard are high enough to make it unreasonable to contemplate emissions in excess of the standards. They implicitly misrepresent the damages from pollution. No one really believes that the social cost of exceeding the standard is infinite. Then, the argument goes, one is left wondering about the rationale for adopting what must inevitably be seen as a somewhat arbitrary standard in the first place.

There are two basic problems in regulating pollution, setting aside enforcement problems for the moment. One is distributing cleanup among effluent sources in an efficient manner. The second is trading off costs and benefits and adjusting effluent levels until costs and benefits are commensurate. Effluent charges accomplish the first objective. If the initial effluent charge is not the correct one, it can be adjusted and, if necessary, readjusted until the appropriate trade-off between costs and benefits is achieved. Thus, the second objective is achieved by a process of trial and error. In the course of this process, the regulators need only adjust one number, the effluent charge. It is adjusted when the incremental benefits of effluent reduction are perceived to differ significantly from the effluent charge, for it is the effluent charge that is supposed to approximate these benefits. The idea, then, is that the regulators go through a cycle: setting charges, obtaining a response from the regulated sector, and resetting the charges. A similar argument could be made for effluent standards, but the economist would argue that it is more difficult to iterate in this case because there are more variables to control.

The underlying presumption is that the regulated sector can and will effortlessly, costlessly, and frictionlessly adjust to changes in the regulations until the hypothetical optimum is reached. This is not a very accurate description of our world. The image of a frictionless and responsive regulated sector is misleading for several reasons. First, the limited information about costs and cleanup technologies that the government possesses is also a problem for effluent sources. The real expenditures required to meet particular standards or to respond to particular regulations cannot be taken back. They are sunk costs. Each time the regulations are changed there are additional costs.

Second, the cleanup technology is usually highly capital intensive. The investments required to respond to a particular set of standards cannot easily be reversed. In fact, they may not be reversible at all.

Third, it takes time to mobilize any organization to engage in a new activity. Once in motion, most business organizations do not easily change direction. Frequent changes in regulations may create serious implementation problems for a well-intentioned business management.

Fourth, the organizational inertia just described applies also to the regulatory organization involved in implementation and enforcement. Learning on the job is rendered significantly complicated by frequent changes in the rules that are being enforced.

Fifth, if the regulated sector anticipates regulatory changes that, in turn, are responses to costs in the regulated sector, then the simple model of the regulated sector responding myopically to each new regulation, be it a standard or a charge, is not realistic.[3] If the regulated sector's behavior affects the rules and the polluters know it, they are unlikely to take each new set of rules at face value.

These factors conspire to make any regulatory process that involves repeated adjustment to new information costly, time-consuming, and perhaps infeasible. These remarks apply to any fine-tuning regulatory process and not just to the effluent charge approach. The same comments would apply to a system of continuously adjusted standards.

If it is costly or impossible to deal with uncertainty by adjusting until the relevant aspects of costs and benefits are known, then it is important to think of the regulatory problem as one of imposing rules based on the best available information, however limited. The combination of the rules and the regulated sector's responses will produce results (effluent levels, cleanup costs, and price changes for many products) that must be endured for some extended period of time. This is not to say that that the rules are immutable but rather that the initial rules are important because their effects will last for an extended period.

One cannot help feeling that the occasional hostility of the debate between proponents of standards and effluent charges is in part the result of a failure of communication among the interested parties about what the practical constraints on regulatory activity may be. Perhaps the two sides have been operating with essentially different models. Economists, using the previously described frictionless model, have labored with some success to explain the merits of the price system and appear to have difficulty understanding why there are so many recalcitrants among the policy makers. The proponents of standards have felt that their approach is safer and more practical, although the case for standards as a control strategy has been less effectively defended on formal grounds.

In the literature and in discussions of control problems with those who have had to frame, implement, and respond to regulation, certain facts emerge upon which most would agree. They are facts that can act as guides in developing regulatory strategies.

First and most important, both the benefits and costs of effluent reduction are uncertain at the time the regulations are imposed. The uncertainty can be reduced through the expenditure of time and resources, but it cannot be eliminated.[4]

Second, adjustments in levels of effluents cannot be made easily or costlessly, for the reasons cited earlier.

Third, any control system requires some form of monitoring. The costs of and available technologies for monitoring vary from one pollution problem to the next.[5]

Fourth, the link between ambient air quality standards and emission levels at effluent sources is imperfectly understood. The diffusion models required to predict ambient air quality from emissions are complicated. They have not been available long, and those that are available are not necessarily understandable to state enforcement agencies.

Fifth, regulation not only affects the regulated sector's incentives to reduce emissions, but it also affects the incentives for research and development in the area of new technologies for cleanup, a subject to which we shall return.

Regulation places restrictions and imposes costs on the regulated sector. There is a large variety of different kinds of restrictions that can be imposed. Perhaps unfortunately, two have attracted almost all of the attention. They are standards and effluent charges. After looking at the important characteristics of costs and benefits in the second section, in the third we examine standards and charges. The reason is essentially two-fold. First is to argue that, between the two, the preferred option depends on some important features of pollution damages and cleanup costs, especially the uncertainty about costs. Second, an examination of the relative merits of each alternative as a response to limited information is essential to understand what we think may be a better practical alternative: combining standards with effluent penalties for emissions in excess of the standards.

Therefore, in the fourth section we outline a regulatory strategy based on standards supplemented by penalties that look and act like effluent charges. This strategy combines the better features of both of the currently debated alternatives. We believe it will appeal to the practically oriented as a reasonable and useful modification of current control systems based on standards.

The Nature of Benefits, Costs, and Uncertainty

Although we can and must quantify the benefits of clean air to make intelligent decisions about desirable levels of air quality, it would be a mistake to think that it is easy to determine the damages from pollution. Ideally, we would like what the economist calls a damages function—a dollar measure of the harm or disutility caused by various levels of pollution. But such numbers are hard to come by. For one thing, there is a large psychic component—polluted air is undesirable, in part, because it is unpleasant. Attaching numbers to people's preferences is always difficult, especially when they disagree. And even when less subjective elements are involved, many of the more tangible damages from pollution, including health effects, are hard to measure.

So a crucial aspect of pollution damages is uncertainty. This acknowledgment does not mean we know nothing about the benefits of clean air. We may know upper and lower limits but be somewhat fuzzy about the area in between.

The uncertainty in benefits does not really affect pollution control strategy per se (the choice between standards and fees, for example). Whatever course is actually tried out may not make us less uncertain about benefits. There is a slight qualification because changes in ambient air quality attributable to regulation provide new data that can be used to improve estimates of the health effects and other impacts of pollution. But the "value" of cleaner air would still remain largely uncertain. In contrast, the way that the regulated sector responds to a particular set of regulations will, over time, tend to reduce our uncertainty about cleanup costs. Uncertainty in benefits can be narrowed only by more research, carried out presumably by some branch of the public sector. Under the circumstances, the economist tends to suppress the uncertainty in benefits by working with the reasonable compromise of an expected damages function. The expected damages function may be higher or lower depending on whether damages are anticipated to be higher or lower. It represents our best single estimate of damages at the time when a regulatory decision must be reached. Henceforth when we speak of damages, we will implicitly be speaking of expected damages.

There remains a problem of translating the various consequences of pollution into an index that is commensurable with cleanup or abatement costs. In discussing this problem with a variety of people who have been involved in the framing and implementation of the clean air act, we have discovered a way of phrasing the issue that seems to have some appeal. Imagine that pollutants are currently at some fixed levels. One can ask what maximum additional abatement costs would be tolerated to achieve a further 10 percent reduction in pollutants. The answer to this question translates into a statement about marginal damages measured in dollars. But it seems easier to confront the issue by comparing the consequences of effluent reduction and abatement costs directly in this way than to attempt to attach a dollar value to effluent reduction abstracted from abatement costs.

Perhaps the most important single property of pollution damage is that the extra damages of an additional unit of effluent often increase (or at least do not decrease) with the overall level of pollution. This is sometimes called the "principle of increasing marginal damages." Although it is not universal, the principle seems to have general validity. When the air is fairly clean, an extra unit of effluent does less damage than when the air is already heavily polluted.

As we shall see presently, the form of an optimal pollution strategy very much depends on the shape or curvature of the damages function. Two extreme cases merit special attention.

A relatively straight damages function means that marginal damages do not increase very much with pollution levels. This function would be characteristic of a situation in which increased pollution leads to a steady, even deterioration without any dramatic changes.

When the damages function is highly curved at some level of pollution, marginal damages are increasing rapidly around that point. This might be a fair description of a "threshold effect" in pollution. In such situations, marginal pollution damages rise precipitously as pollution starts to become dangerous or uncomfortable.

The principle of nonconstant marginal damages means that it is difficult to place a single, unambiguous price tag on pollution. Unfortunately, marginal damages depend on the level of pollution. To price effluents correctly, we would need to know what the level of pollution is or will be. The fact that we do not know the pollution level in advance makes it difficult for a fee system to function well. Naturally this problem is going to be more acute when marginal damages are changing rapidly than when they are relatively constant.

Turning now to costs, our starting point is the cleanup cost function. This is simply a schedule giving the dollar outlays necessary to obtain a certain reduction in emissions.

Most analysts believe that the incremental cost required to eliminate an extra unit of pollution goes up as the effluent level declines.[6] That is, it is less costly to eliminate the first 5 percent of effluents than it is to eliminate the second 5 percent, and so forth. Economists call this phenomenon the "principle of increasing marginal costs."

The shape of the cleanup cost function will have some bearing on the form of an optimal pollution strategy, just as does the shape of the damages function. Sharply increasing marginal costs give rise to more highly curved cost functions, whereas slightly increasing marginal costs are associated with relatively straighter cost functions. There are no general principles for determining the curvature of cost functions. It depends on the situation and varies from case to case. In short, it is an empirical matter.

It seems to be a fact of life that the regulators don't know cleanup costs to a high degree of accuracy at the time the regulations are imposed. It is especially true when a new or unproven technology is involved, as with auto emission controls. There is no way of knowing beforehand exactly what it will cost to achieve a certain cleanup level. Estimates can be made, but the final costs will not be known until mass-produced equipment is in place, if then.

The uncertainty in cleanup costs is essentially due to lack of experience. It can be reduced by research but not altogether eliminated. No one knows precisely what it will cost to achieve some cleanup level because it has never before been tried. Once a full-scale effort has been launched, the relevant costs will eventually become known but not before.

There is another important feature of cleanup costs that goes along with the uncertainty. Not only are costs unknown, but it is also difficult and expensive to find out what they are. Sometimes economists and others share a tendency to conceptualize regulation as a process of continual fine-tuning. A certain strategy is adopted, and marginal costs and marginal benefits are observed; if they are not equal, the fees, standards, or other parameters are smoothly adjusted until an optimum is obtained.

As argued earlier, this may be an inappropriate way of viewing the problem. In order to have a chance to work, a regulatory strategy must be left in place for an extended period after it has been adopted. As we see it, analysis of regulatory strategy should start from the following point of departure: *the regulators are forced to make decisions in an uncertain environment and they must live with the consequences for some time.* Among these consequences is the possibility that costs will turn out to be higher or lower than was expected. The above principle will provide a framework for analyzing certain important issues that are outside the scope of the fine-tuning model. With it in mind, we turn to the relative merits and demerits of effluent fees and standards.

Fees vs. Standards

The two best-known regulatory strategies for controlling pollution are effluent *standards* and effluent *fees*. They are easily comprehended and are frequently contrasted. In this section we propose to analyze carefully the comparative advantage of each of these control strategies.[7]

It is useful to analyze standards and fees for at least two reasons. For one thing, this issue is of interest in itself because there is a long-standing policy debate about the comparative merits of these two control modes. For this reason alone it is important to understand how fees and standards work and to be able to identify situations in which each one is likely to outperform the other. A second motivating factor is our own interest in promoting a mixed standard-fee system, which we, and others, feel may be superior to either standards or fees alone. To understand how the proposed mixed system works and just exactly why it is better requires a thorough acquaintance with the basic subcomponents out of which it is constructed.

By far the easiest pollution strategy for the public to comprehend is the one based on standards. Some branch of the government acting on behalf of society's interests establishes upper limits on emissions for each polluting firm. It might appear at first glance that with standards cleanup costs are borne by the polluter, but in fact they will eventually be passed on to the consumer in the form of increased prices, reduced employment, and so on. The standards approach is popular in large part because it represents a direct assault on the problem that fits comfortably with legal, moral, and historical traditions. If the problem is that some identifiable group is overpolluting, the obvious remedy is to force it to clean up to a level more in keeping with society's needs as a whole. What could be simpler than decreeing that pollution be cut back to some level that approximates the public's interest?

An alternative strategy, one frequently favored by economists, is the effluent fee system. Effluent fees are prices that attach to effluents. A polluter pays for his effluent an amount proportional to his discharge volume. In controlling the fee, a regulatory agency indirectly controls effluent discharge by manipulating the incentive to engage in cleanup. A higher fee encourages polluters to clean up more whereas a lower fee elicits more pollution.

To many people a fee system is an unfamiliar and peculiar method of controlling pollution. To make a fair evaluation, it is important to understand exactly how it works. When a fee is imposed on emissions, it indirectly controls pollution in the following manner. A polluter, in order to maximize profits or minimize costs, will fix emissions at that level at which the incremental cost of cleaning up an extra unit of pollution equals the fee. If the fee exceeds the marginal cleanup cost, money could be saved by making the extra investment needed to cut back pollution slightly, and vice versa when the marginal cleanup cost is greater than the fee.

Effluent fees and standards differ in how the burden of cleanup costs is shared. A full analysis of the distributional implications of either system would constitute an excessively lengthy aside. However, it is worth noting one point: an effluent fee system generates government revenues, which can be used for public expenditures or to reduce the burden of taxes collected in other ways. How they are used will in large part determine the distributional impact of the control strategy. Economists regard these distributional issues as impossible to decide on purely economic grounds, but this difficulty does not mean they are unimportant. In our experience, many practitioners react negatively to the potentially large payments that must be made under a fee system.

In an uncertain world in which cleanup costs are not precisely known to the regulators at the time a decision must be made, the comparative advantages of standards or fees derive from the following basic observation: *standards fix pollution levels but leave cleanup costs uncertain;* in contrast, *fees fix (incremental) cleanup costs but leave pollution levels uncertain*. Which of these features is more desirable depends on the underlying economic situation. As we shall see, sometimes one feature is more important, sometimes the other is. We propose to examine a few extreme cases to illustrate the general principles that are involved.

The fact that effluents are fixed under a standards system tends to make that approach relatively more desirable as the damages function is more highly curved. When marginal damages rise rapidly around some threshold level, it would probably be foolish to use effluent fees to control pollution because the pollution level remains uncertain. If the marginal social benefit of clean air is low in some range but increases precipitously as pollution starts to become dangerous or uncomfortable, then for a wide range of costs the effluents should be at or near the threshold level. And under standards they will be. But if effluent fees are used, pollution levels will vary with costs. Should cleanup costs turn out to be higher than anticipated, profit-maximizing polluters could elevate pollution levels into the danger zone. If costs are lower than expected, polluters may be motivated to clean up well beyond the threshold level, to an extent that is not socially justified because marginal damages at that point are insignificant. In a world of cost uncertainty in which regulators must work with fixed fees or standards, a threshold effect in pollution damages makes a strong case for standards.

The opposite kind of conclusion holds with respect to the curvature of the cost function. The straighter the cleanup cost function, the stronger the case for standards. If incremental costs increase only slightly with cleanup levels, it is very difficult to control pollution by fees. Suppose a fee is named. A polluter will set emissions at the level at which marginal cleanup costs equal the fee. Now, when marginal costs vary little as pollution changes, it means that pollution varies greatly as marginal costs change. Because polluters set marginal cleanup costs equal to the effluent fee, even slight fee changes will be translated into large swings in the pollution level. If the fee is set correctly in the first place, everything will be fine. But, as we have tried to emphasize, there is a large amount of uncertainty in any real-world regulatory environment. With relatively straight cleanup costs, the slightest miscalculation of the fee will result in either much more or much less than the desired pollution level. In such a situation, standards tend to be better because a high premium is placed on the rigid output controllability that only they can provide under uncertainty.

Just as a more highly curved damages function favors standards, so a relatively straight damages function is more conducive to fees. If the damages function is close to being linear, it would be foolish to name standards. If the marginal social damage is approximately constant in the relevant range, a superior policy is to confront the effluent sources with an effluent charge equal to the marginal damages. Then the polluters will automatically bring themselves close to a social optimum by picking the emission level that equates marginal cleanup costs to the fee.[8] With a straight damages function it is much better to have the polluters find their own desired emission level on the basis of a fee than to have the regulators determine it for them by setting a rigid emission standard at a time when costs are uncertain. In this case, the fee system is more attractive because it gives the ability to fix marginal costs in an uncertain world.

Perhaps the main reason that some economists traditionally favor fees over standards is that fees automatically induce an efficient distribution of cleanup effort among different sources. The economist tends to view standards as piecemeal regulation that offers no guarantee that the overall level of pollution will be attained at least cost.

Here is an example. Suppose it has been decreed on the basis of crude cost calculations that mobile sources should cut back sulfur dioxide emissions by 30 percent and stationary sources by 50 percent. Because the regulators don't really know what the actual costs of pollution abatement will turn out to be (and maybe even if they do know), there is no guarantee that the incremental costs of a small further reduction will be the same for both sources. It would be better to require the source with the smaller incremental cleanup costs to pollute a little less and to permit the other source to pollute a little more. The same overall sulfur dioxide level would be attained but at less total cost.

It is important to understand that imposing a uniform fee on all emitters of a specific pollutant automatically guarantees that the overall cleanup level will be obtained at least total cost. Each polluter, in order to maximize profits or minimize costs, will set emissions at that level at which the marginal cost of cleaning up an extra unit equals the fee. Because the marginal cleanup costs of each polluter are equal to the same fee, they are equal to each other. This is the hallmark of a least-cost allocation of cleanup activity. Only when marginal cleanup costs differ would it be possible to obtain the same overall pollution level at less cost. As in the previous example, this

effect would be accomplished by allowing less pollution from the source with low marginal cleanup costs and more from the high-cost polluter.

The automatic efficiency of a fee system is a definite point in its favor because it results in cost savings. If we lived in an infinitely flexible control environment where the regulators could continuously and costlessly adjust the fee, the efficiency argument might be overwhelming. But, in practice, adjustments are usually very costly. The consequences of any regulatory action are going to be with us for a while. And then the uncertainty about pollution levels that is inherent in a fee system can become troublesome for the reasons just discussed. Standards may be preferable to fees in a multiple-source setting even though standards are inefficient. It all depends. If it is important to hold overall pollution to some prescribed level, that need may take precedence over having a cost-minimizing way of achieving an uncertain level of pollution.

The cost-saving or efficiency argument for fees becomes more significant as the variety of polluters increases. A fee system enables the regulators automatically to screen out the low-cost polluters by encouraging them to clean up more relative to the high-cost polluters. This screening effect is more significant if there are many different types of polluters because the possible cost savings are greater. If there are three distinctly different types of sulfur dioxide emitters that have independent cleanup technologies instead of one large pollution source that yields the same aggregate effect, the case for fees is strengthened, other things being equal.

Note that the desirable cost-screening effect of fees doesn't work unless the different pollution sources really are different. A fee system for automobiles is not likely to permit much cost screening because cleanup costs are not likely to differ much from one auto to another. If costs of several polluters are highly correlated, as with automobiles, it is best to lump the units together and view them as one *type* of pollution source. The more different types of polluters there are, the greater the potential cost savings of a fee system due to the screening effect.

It is perhaps useful to summarize at this stage. The comparative advantage of fees and standards for controlling pollution depends on the shapes of the damages and cost functions, on the magnitude of the uncertainty, and on the number of effluent sources with relatively independent cleanup costs. Standards are favored as damages are curved or costs are straight. Fees are favored as damages are straight or there is a larger number of independent polluters.

The purpose of this section has been to give the reader a feeling for the way fees and standards work and when each one is likely to work better than the other. In the next section we are going to propose a mixed standards-fee system that outperforms either pure system.

The Pressure-Valve Approach to Regulation: Standards with Effluent Penalties

In the presence of limited information, standards are often preferable to effluent charges because they prevent the levels of effluents from running up when costs turn out to be higher than the initial estimates. This feature of standards is particularly important when the incremental damages increase rapidly with the level of effluents. On the other hand, standards are rigid and therefore unresponsive to situations in which cleanup costs turn out to be high. Under these conditions, of course, it is possible to relax the standards. Relaxation, in the form of a delay, has occurred in the case of automobile emissions. And it is done elsewhere when the need arises. But the conditions under which the standard is relaxed and the way it is relaxed are of extreme importance in determining the effectiveness of the control program.

The problem, in a nutshell, is to devise a mechanism that responds to high cleanup costs for individual sources and at the same time does not create an incentive for noncompliance with the standard. What is needed is a penalty, specified in advance and paid by the source in case its emissions exceed the standard set for that source. Moreover, to maintain the incentive to clean up, the penalty should increase with the amount by which emissions exceed the standard. A practical way to achieve this goal is to establish a penalty *per unit* of emissions in excess of the standard.[9] That penalty will act like a high effluent charge for the sources that turn out to have high costs. Those sources that have costs close to prior expectations will find it desirable to meet the standards. For them, and they will be in the majority, the system will function as if there were simple standards. For the very high cost polluters, the system will function as if there were fees.

The proposal is to add effluent penalties to the standards that take effect only when the standards have been exceeded. It is a system in which a set of standards is supplemented by *pressure valves,* which release only when the costs for an individual source exceed the estimated costs by a significant amount. Each individual polluter decides for himself whether his cleanup costs are high enough to justify paying the fee for exceeding the standard. Moreover, when the escape valve releases, it is not a complete release. The standard for a high-cost polluter is *not* reset at no cost to the source. Rather, the polluter pays a penalty, which can be avoided at a future date, in the

event of a reduction in the costs of cleanup. Therefore the incentive to maintain the effort to reduce the cleanup costs is retained.

This approach has several attractive features. Most sources will have costs in the neighborhood of the average of prior expectations. They will therefore meet the standards. Only those sources whose costs are significantly higher than estimated will choose to exceed the standards, and they will pay the penalties. High-cost sources will clean up less than those with lower costs. This is one of the more important attractive properties of the effluent charge approach, but it is one that can be overridden by other considerations. The protection of standards against high pollution levels is substantially maintained. So long as the penalty is set above expected marginal damages, the escape valve will operate only when costs and benefits differ by a sizeable amount.

The penalty that the individual polluter faces more closely approximates the social costs of his contribution to pollution than under pure standards or effluent charges. The penalty is neither prohibitive, as in effect would be the case under rigid standards, nor is it too lenient, as when it is equal to the incremental damages at the anticipated outcome. In the latter case, the system would function like an effluent charge system and would therefore have the problem of suboptimally high levels of pollution that high costs would cause.

The setting of the optimal penalty is a matter of some complexity.[10] For practical purposes, however, there is only one important principle. The penalty should be related to the damages or social costs of pollution. As a rough approximation, the penalty per unit of effluent should be somewhat higher than the marginal damages at the levels of effluents that would obtain if all the standards were met.[11] It is important to remember that the incremental damages increase with the level of effluents. Therefore, as a first approximation, one could do worse than setting the penalty equal to the marginal damages at the level of effluents prior to the imposition of regulation. Or to put it in the form of a decision rule, assess the marginal damages of pollution at existing (precontrol) levels of effluents and set the unit effluent penalty equal to that number. If that is done, then the escape valve will function only for those sources whose costs are so high as to exceed the benefits of cleanup at relatively high precontrol levels of pollution. If marginal damages increase dramatically with effluents, then that penalty will be high, as it should be. On the other hand, if marginal damages do not increase rapidly, then the penalty will be lower, and that relationship also is desirable. The setting of the escape valve can and should be tied to the curvature of the damage functions. Setting the penalty at the precontrol marginal damages is one relatively simple way of accomplishing this end, although there are others.

Analysts have worried that the effluent charge system, even when it is desirable (for the reasons discussed earlier), possesses some embarrassing features on the financial side. It generates a lot of revenue that then has to be disbursed. And effluent sources pay double: once for cleanup and once for the effluent fees on the emissions after cleaning up. That system may impose a rather heavy financial burden on some sources. The standards and penalties approach does not have this pair of problems. Most sources do not pay anything to the government. They meet the standards. Some sources pay some penalties. But the penalties are only on the emissions in excess of the standards. The payments are therefore smaller by orders of magnitude than the fees that would be paid under a simple effluent charge system. Indeed, another way of describing this control strategy is as an effluent charge that doesn't take effect until certain targets or standards (one for each source) have been exceeded.

In most of the pollution problems with shich we are familiar, damages increase at an increasing rate with the levels of pollution. These are the circumstances under which the use of standards is preferred, when the choice is between standards and simple effluent charges. We are suggesting that one can improve upon standards by instituting an automatic escape valve that applies to individual sources and prevents the worst dislocations in the event of unexpectedly difficult cleanup problems. The escape valve takes the form of a high (but not prohibitively high) penalty that is proportional to the amount by which effluents exceed the standard.

The basic argument for this modification is that it makes the penalties for the individual polluter more closely approximate the actual damages he is causing than is the case under either rigid standards or effluent fees. The modified system is a form of flexible standard approach, but it is preferable to systems in which standards are simply relaxed without penalty (then no one would comply). The costs of exceeding the standards are nontrivial and nonzero. These costs are also specified in advance. The system therefore maintains the incentives to attempt to meet the standard and to reduce levels of effluents. It also removes the incentive to try to appear to have high cleanup costs and thereby impress the regulators to relax standards. In addition, the penalty-modified standards will not result in the collection of enormous volumes of revenue from the private sector. And implementation of the system does not require a major change in the direction of policy. All that is required is a redefinition of how to assess penalties for non-

compliance. It is a practical and, we think, useful addition to the current system of control via standards.

There is another rather important set of issues that have emerged in discussions with regulators and interested observers. They concern the incentives for research and development in pollution control technology that the regulations create. It is clear that the regulations not only affect cleanup activity in the private sector, but they also determine the way in which technological development proceeds in this relatively new industry. The effects of regulation on the patterns of technological development are potentially among the most important long-run effects that regulations can have.

This subject can be rather difficult and complicated, but there are a few observations that can be made in support of the penalty-supplemented standards that are being proposed here. Both standards and effluent charges have potentially distortionary effects on the pattern of research and development. From a social standpoint, the R and D problem is one of investing in technologies that are good "gambles." Good gambles are technologies that have a significant chance of reducing the costs of effluent reduction and at the same time do not run significant risks of being ineffective in reducing emissions. Society wants resources and effort expended in attractive but nonspeculative ventures.

We can think of an R and D investment program as having a cost (the initial investment) and a distribution of outcomes defined in terms of the levels of effluent reduction that can be achieved at the conclusion of the program. It is useful to think of that distribution of outcomes as having a mean and a variance. What we want is a program with relatively low costs that has a high mean and a low variance. The low variance is desirable because it means that the chances of a disaster (no significant reductions in pollution) are reduced.

With these objectives in mind, what can be said about effluent charges and standards? Consider effluent charges first. Because the penalties facing the regulated sector are proportional to emissions under effluent charges, that sector will respond by trading off the costs of the R and D program on the one hand and the mean of the distribution of effluent reductions on the other. The variance is likely to be ignored as a factor in the choice of an R and D program, but the variance is irrelevant only when the damages function is linear. Therefore, the effluent charge approach in the context of the R and D problem has the potentially fatal flaw that it fails to provide the regulated sector with an incentive to respond negatively to the variance of the outcomes.[12] It therefore fails to provide protection against the risk of potentially high pollution levels. This problem is the analogue of the tendency of effluent charges to generate excessive levels of effluents when costs turn out to be high. Both problems result from the fact that a curved damage function is approximated by a straight line.

Standards pose a different problem. Under standards, provided they are not expected to be relaxed, the private sector will respond by selecting a program that minimizes the probability of failing to meet the standards. That may be a very costly program. And thus, in an important sense, standards may cause the R and D program to be excessively costly and conservative. A substantial reduction in costs with a small increase in the probability of noncompliance will be rejected even if it might appear to be a rather good decision from a social point of view. Costs figure in the R and D program under standards only if there are programs that assure the meeting of the standards. The private sector will then select the least-cost program that meets the standard.

As a rough approximation, standards respond first to the mean and variance, especially the variance, and then to costs if the standards can be met with certainty. This pattern also produces distortions. Costs and benefits may not be properly traded off.

Because the penalty-modified standards more closely approximate the damages, these sorts of problems will not arise in as severe a form. The private sector will hedge against uncertainty but not to the exclusion of cost considerations. Thus the case for the mixed approach is based in part on the need to structure the incentives for research and development in an appropriate way. The automobile air-pollution case is an example of the desirability of having a somewhat flexible standard in a situation in which the development of the control technology is a central feature of the problem. It is a complicated case, which Mills and White treat in detail elsewhere in this volume.[13]

The Number of Effluent Sources

The preceding analysis has assumed that firms are the source of pollution and that their number is fixed or varies only slowly over time. It further assumes that the number of sources is not particularly sensitive to the regulations, whatever form they take. There are, however, important problems in which this assumption may not be true. The best-known and perhaps most important is the automobile case. When the number of sources is fixed, the issue is confined to determining the optimal effluent level per source. When the number of sources varies there are two issues. One concerns the effluents per source; the other is the number of sources. In the automobile case, effective pollution policy involves controlling both the effluents per vehicle and the number of vehicles.

In the short space available to us, we cannot deal completely with this problem. However, some comments are in order. First, many of the considerations discussed previously are applicable to this situation and point to the merits of a mixed approach to controlling effluents per vehicle. One difference is that a strong case can be made that the control program directed at the effluents per vehicle should be supplemented with a tax on the vehicles themselves. Without such a tax, the pollution cost of an additional vehicle is not borne by the purchaser; therefore, unless the demand is inelastic, there will be too many vehicles on the road.

Even in the case of stationary sources, there is a long-run decision as to their number and locations. For these decisions to be made correctly, something more than the correct marginal incentives is required. The absolute magnitude of the penalties paid by an effluent source should equal the estimated incremental social cost of that source. Otherwise the long-run entry and exit decisions may not be the correct ones. This possibility suggests that the mixed system proposed earlier should have appended to it an effluent charge equal to or slightly below the marginal damages at the pollution levels implied by the standards. This charge has two properties. It is paid on all units of effluent, so that not only the marginal social cost but also the total cost of the source is internalized. And the effluent fee is below the supplementary penalty that takes effect only when the standard has been exceeded.

Thus, the appropriate regulatory response to cleanup cost uncertainty would include an effluent charge of the conventional kind and a standard accompanied by a penalty for emissions in excess of the standard. This system would act like a nonrigid standard system. It would also provide an incentive to clean up beyond the standard for those sources that turn out to have low abatement costs. And it would approximate the appropriate long-run incentives in the private sector for the correct entry and exit of effluent sources.

Notes

This paper is based on some recent work by the authors and others on the subject of controlling via quantities and prices. Weitzman was the first to study the advantages of quantity controls as a response to uncertainty. Spence and Roberts used the device of transferable licenses in the pollution context to examine the merits of effluent standards and of mixtures of fees and standards. The purpose of transferable licenses is to maintain efficiency in the distribution of effluent reductions among sources and to employ an overall quantity control. The authors would like to thank the members of an early workshop on the Clean Air Act for their encouragement and their aid in helping us to understand various facets of the problems. We want particularly to thank Douglas Allen for his help. Robert Dorfman, Peter Diamond, and Robert Solow pointed to some problems with which we have tried to deal. There remain some issues of a theoretical kind that are best confronted in the more technical papers listed in the references.

1. The reason is that the sources of pollution clean up to the point at which the marginal costs of cleanup equals the effluent fee. But then the marginal costs of cleanup are the same for all sources, the implication being that cleanup cannot be shifted among sources in a way that reduces costs.

2. The benefits, of course, are extremely difficult to measure and to agree on. And no regulatory approach is excused from undertaking this task.

3. It is not easy to predict how the regulated sector will respond to a situation in which regulations are anticipated to change in response to costs. It seems reasonable to surmise that the regulated sector's efforts might be sluggish under these conditions.

4. The decision about how far to go in collecting information is one of the more important policy decisions that arises. It is rarely made with a view to what difference further information will actually make.

5. Moreover, monitoring technology is capable of changing over time. It appears, for example, that the monitoring technology for particulates from stationary sources is insufficiently advanced to permit the use of effluent charges in that context. On the other hand, it is also true that this situation may change if the government devotes more resources to the advancement of monitoring technology. At present, this essentially public R and D investment seems to be deficient.

6. See, for example, the costs derived by Donald N. DeWees in Chapter 7 of this volume.

7. In the interest of focusing sharply on the essential economic differences between fees and standards, we are abstracting away a host of "noneconomic" factors. This treatment does not mean that we feel they are unimportant. On the contrary, there may be significant legal, administrative, or historical reasons for favoring standards or fees in one situation or another. Even such factors as the capability of the monitoring system play no small role. At the present time we do not seem to have an in-place technology capable of continually monitoring stationary particulate sources to the degree of accuracy required for making an effluent fee system work effectively. The desirability of having the government encourage research in the monitoring area is something that cannot be stressed enough. Certainly the regulated sector has little direct incentive to develop or improve the monitoring system.

8. Of course, effluent sources may first have to expend resources to learn about the costs and control technologies.

9. There are other possibilities. The penalty per *unit* of emissions in excess of the standard could increase with the amount of the excess.

10. For a technical discussion of related issues, see Martin L. Weitzman, "Prices vs. Quantities," *Review of Economic Studies,* October 1974; and Marc J. Roberts and Michael Spence, "Effluent Charges and Licenses Under Uncertainty," *Journal of Public Economics,* 5 (1976):193–208.

11. There is currently a movement to employ penalties as part of the enforcement programs in some states. Connecticut is one of them. These penalties are set to equal the cleanup costs avoided by those not in compliance. It is important to stress that this is a technique *for enforcing standards*. It should not be confused with the proposal put forward here and elsewhere. The penalties in the escape-valve approach are related to damages and may be translated into maximum tolerable abatement costs, as discussed earlier. But maximum tolerable costs and actual costs are very different.

12. It may be that the regulated sector behaves as if it were risk averse. If it does, then it will respond negatively to the variance even though the penalties are linear in effluents. This behavior will mitigate, to some extent, the tendency of effluent fees to cause risk to be ignored in R and D programs. A full analysis of this complex set of problems is unfortunately beyond the scope of this paper.

13. See Edwin Mills and Lawrence J. White, Chapter 8 of this volume.