

firm k , and i any output of firm k , $p_i = \beta_k f_i^k$, $p_{i^*} = \beta_k f_{i^*}^k$ [by (4^c)] so that eliminating β_k between these two equations we have at once [by the analog of (17)]

$$p_i = p_{i^*} f_i^k / f_{i^*}^k = p_{i^*} \partial \bar{y}_{i^*k} / \partial y_{ik} \quad \text{where } \bar{y}_{i^*k} = |y_{i^*k}|.$$

That is, the optimal price of i is the cost of the quantity of input i^* necessary to produce a unit addition in i , all other inputs and outputs held constant. In brief, it is the marginal *private* cost of i .³⁴ This means that if, for example, marginal smoke damage to the production of two firms of different types, k and k^* , is quite different (that is, if $\mu_k \partial f^k / \partial z \neq \mu_{k^*} \partial f^{k^*} / \partial z$), there will be no compensation to offset the differential effects on the optimal prices of the two outputs. Laundry and phonograph records will both sell at their private marginal costs though one cost is more heavily affected by smoke than the other. Thus, our results imply that, *in the presence of an externality, optimal resource allocation calls for pricing that involves zero taxation and zero compensation to those affected by the externalities* (but nonzero taxation of their generators). Laundry purchasers will then not be able to buy laundry at a price below the high marginal cost resulting from the presence of smoke, because there is no compensation.

d. Where the externality is shiftable, the result is a bit more complicated. It is still true that no compensation for the damage suffered should be paid to victims of the externality.³⁵ However, victims must now be subject to a tax on any shifting activities they choose to engage in. More specifically, they must pay a unit tax for shifting equal to the marginal social damage to the secondary victim. Such an extension of the Pigouvian tax will create the requisite incentives for an optimal allocation of the external product among victims.

³⁴ The reader may be disturbed at the notion of expressing marginal private cost in terms of the single input, i^* , rather than the full set of inputs used by the firm. However, in equilibrium the ratio of input price to marginal physical product will be the same for all of the firm's inputs. Hence, it will cost exactly the same whether the firm expands an output by a very small amount by using more of one input, i^* , or more of some other input, i^{**} , or any combination of these inputs.

³⁵ This result requires one important qualification. As Martin Bailey has shown recently, compensation need not result in any allocative distortions if such compensation is capitalized into property values. Such compensation payments become, in effect, an increment in rent (i.e., a lump-sum transfer). We explore Bailey's argument in Chapter 14 in conjunction with our treatment of subsidies for pollution abatement.

Uncertainty and the choice of policy instruments: price or quantity controls?

In the preceding chapter, we focused our attention on a particular policy measure: Pigouvian taxes (or effluent charges) as a means to regulate pollution. In this chapter, we will expand the set of available policy instruments to encompass marketable emission permits. Until fairly recently, most microeconomists would have argued that the two were virtually equivalent – nearly so in practice and surely so in theory. But a series of papers published in the 1970s have forced a major revision in this view.¹ These papers demonstrated that in the presence of uncertainty, the expected value of social welfare can differ markedly under a system of Pigouvian fees from that under a regime of marketable emission permits. These two policy instruments remain equivalent in a setting of perfect certainty. But as we shall see in this chapter, under particular forms of uncertainty, these two approaches to environmental management will yield very different outcomes. Depending on the shapes of the marginal damage and marginal abatement cost functions, the environmental authority will be better advised under some circumstances to use one of them, and under other circumstances to employ the other.

Our objective in this chapter is to describe the logic underlying the choice between these two policy instruments. For this purpose, we shall rely primarily on a series of simple diagrams that depict the basic propositions.

1 Effluent charges, marketable permits, and direct controls

It is important at the outset to be clear as to the precise character of the alternative policy instruments. As we have seen in Chapter 4, the

¹ See Martin L. Weitzman, "Prices vs. Quantities," *The Review of Economic Studies* XLI (October, 1974), 477–49; Zvi Adar and James M. Griffin, "Uncertainty and the Choice of Pollution Control Instruments," *Journal of Environmental Economics and Management* III (October, 1976), 178–88; Gideon Fishelson, "Emission Control Policies under Uncertainty," *Journal of Environmental Economics and Management* III (October, 1976), 189–97; and Marc J. Roberts and Michael Spence, "Effluent Charges and Licenses under Uncertainty," *Journal of Public Economics* V (April/May, 1976), 193–208.

Pigouvian fee is a tax (or effluent charge) per unit of emissions set equal to marginal social damage. In contrast, a system of marketable (that is, transferable) emission permits is one in which the regulatory authority effectively determines the *aggregate* quantity of waste emissions but leaves the allocation of these emissions among sources to market forces. To implement such a system, the environmental authority would issue permits for waste discharges such that, in the aggregate, total discharges would be at the level that equates marginal abatement cost and marginal social damage. Trading of these permits among sources would then establish the market-clearing price.

We emphasize that such a permit system is very different from the "direct controls" approach to permits or licenses. Under a system of direct controls, the environmental authority specifies for *each* source an allowable level of emissions. The emissions quota assigned to a particular source is not tradable so that there is no market in emission permits. We will have more to say about such systems of direct controls in Part II of this book. Here we simply note that in this chapter we are not concerned with such systems; our interest here is in marketable permits.

2 The equivalence of marketable emission permits and charges under certainty

It is clear that when the relevant functions are known with certainty by a welfare-maximizing regulator, exactly the same result will be achieved by a market in permits and by a system of effluent charges. If the optimal number of permits is issued by the environmental control agency, their price will be bid up on the free market to precisely the level of the Pigouvian tax. At that point, it will make no difference to the polluter whether he pays t dollars in effluent charges per unit of his emissions and pays it directly to the authorities, or whether, instead, he pays that same t dollars per unit of authorized emissions for the purchase of a permit on the unregulated permit market. In both cases, the polluter will restrict emissions by exactly the same amount, so each polluter who continues in operation will react in exactly the same way to the one incentive as to the other. The increased cost of doing business may also induce some exit from the field, but since the cost of any type of operation will not differ from one approach to the other, exit decisions will also be unaffected either in terms of the number of existing firms or the identity of the units that find it rational to cease operation.

Figure 5.1, the type of diagram we will be using throughout the chapter, depicts this outcome (it is based on the diagrams in Adar and Griffin). The horizontal axis of the diagram indicates the amount by which total

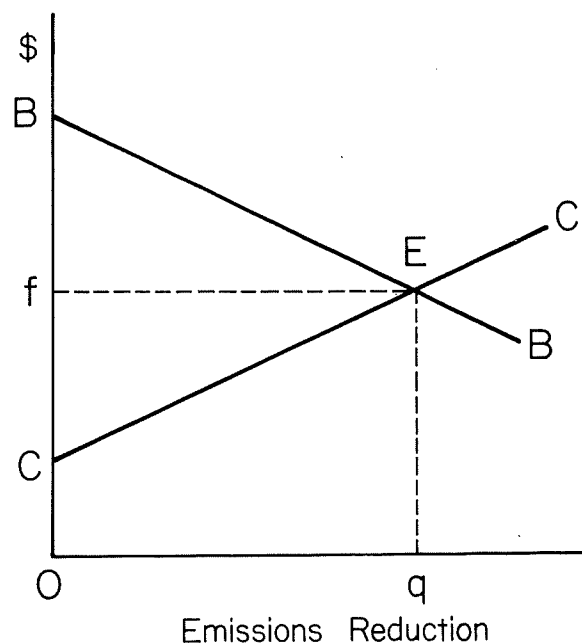


Figure 5.1

emissions are *reduced* below their uncontrolled level; the origin thus represents zero decrease in emissions below the level that would occur in the absence of an emissions-control policy. The curve BB represents the marginal social benefit of emissions reduction as a function of the quantity of emissions that has already been eliminated. In accord with the usual observations, it has a negative slope, indicating that the greater the degree of purity of air or water that has already been achieved, the less the marginal benefit of a further "unit" of purification. Similarly, CC , the curve of marginal control costs is increasing because of the rising cost of further abatement as the zero emissions point is approached. The optimal point is obviously E at which the marginal cost and benefit are equal. E can clearly be achieved by imposing a charge equal to f upon each unit of emissions; polluters would then find it more costly to pay the tax than to adopt measures that reduce their emissions up to the point where q units of emissions have been eliminated.

Similarly, the optimal solution can be attained if the environmental control agency issues a quantity of emissions permits just sufficient to lead to a q -unit reduction in discharges. If R is the amount that would be

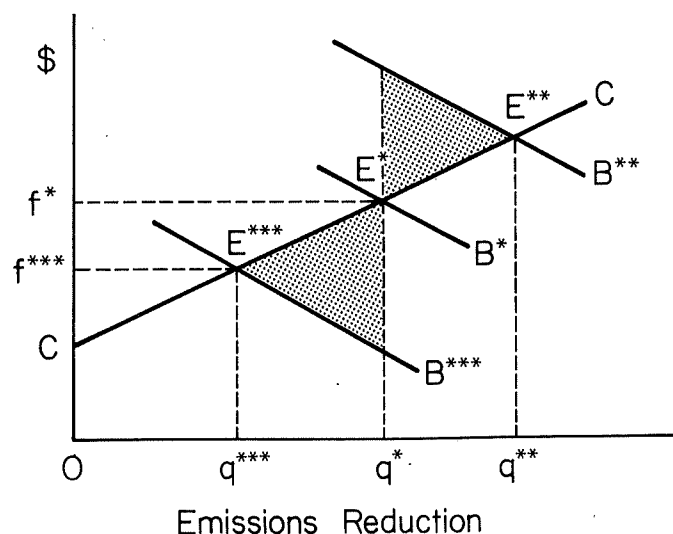


Figure 5.2

emitted in the absence of intervention by the public sector, the permits must allow, in total, $R - q$ units of emissions. Assuming that the market for permits is competitive, the price of a permit (allowing one unit of discharge) will be bid up exactly to f , that is, to the corresponding marginal cleanup cost.

Thus, both approaches will have the same result, reducing emissions to the optimal level and incurring the minimum cost for this level of control.

3 Regulatory uncertainty about the benefits function

Matters obviously become rather different when the regulator is unsure of the position of the pertinent curves and is therefore to be expected to make some error in calculating the optimal quantity of emissions reduction.

We will examine, in turn, the case of uncertainty about the benefits curve and uncertainty about the curve depicting marginal control costs. We will show in this section that when the regulator does not know the true position of the benefits curve, policy will, in general, not be optimal, which is to be expected. However, the resulting error and the corresponding social cost will be the same under effluent charges and marketable permits.

Figure 5.2 depicts this case. In the figure, the regulator has precise and correct information about the cost curve, CC . However, on the unfounded

belief that the benefit curve is B^* , the agency selects E^* as the optimal point, and so either introduces a fee, f^* , or issues the quantity of permits corresponding to q^* . Under either policy, the outcome will be the same if the market for permits is competitive. Emissions will be cut by q^* units and, if permits are issued, their price will rise to f^* .

If, however, the true benefit curve is B^{**} (rather than B^*) and the marginal benefits of abatement are thus greater than the regulator had thought, the q^* -unit reductions in emissions will be undesirably modest. In this instance, the optimal point is obviously E^{**} with a corresponding reduction in emissions of q^{**} . The social loss resulting from choosing q^* instead of q^{**} is equal to the shaded triangle to the right of and above point E^* (i.e., the excess of benefits over costs over the range from q^* to q^{**}). In contrast, were B^{***} the true benefits curve, then the choice of q^* would represent an excessive level of abatement activity. The social loss in this case would equal the shaded triangle to the left of and below E^* , which indicates the excess of control costs over benefits over the range q^{***} to q^* .

In short, an error in estimating the benefits curve necessarily has undesirable consequences, but those consequences and their undesirability will be exactly the same whether effluent charges or marketable permits are the regulator's chosen control instrument. It follows that uncertainty about the position of the benefits curve by itself offers no guidance on the choice between the two types of measures.

Why do the two approaches yield the same result when the cost curve is known? The answer is summed up in

Proposition One. Given any marginal control cost function $MC(q)$, then the regulator can be sure in a competitive market that emissions reduction q^* will emerge if price (the effluent fee) is set at $p^* = MC(q^*)$, and p^* will emerge as the equilibrium price of a unit emissions permit if a quantity of permits just sufficient to require emissions reduction of q^* is available. These prices and quantities depend exclusively on the cost function and are entirely independent of the shape or position of the benefit function.

This result follows because sources respond to the policy choice along the cost curve, CC . The authority can either set q^* directly through the issue of permits or, equivalently, can attain q^* indirectly by setting an effluent fee of f^* . The choice of a fee of f^* is identical in its effects to the issue of q^* of marketable permits.

4 Uncertainty about control costs

The key property of a system of marketable permits is that, if enforced, such a system guarantees a ceiling on emissions, *no matter how high or*

how low the cost of keeping them to that level. Similarly, the choice of an effluent fee guarantees that if the polluters are minimizers of the cost of achieving a given vector of outputs (level of utility), the marginal cost of emissions control will be equated to the level of the effluent charge that has been selected, *no matter how large or small the resulting quantity of emissions*. Thus, a regulator who adopts a system of marketable permits will be able to achieve the amount of reduction in emissions that he had decided upon beforehand, but he may be greatly surprised at the associated costs. In contrast, the regulator who employs an effluent fee can be confident about the resulting marginal control cost, no matter how uncertain he is about the cost function, but he cannot rely on this means as a device to achieve a target level of emissions reduction.

However, this should not be taken as a claim that the regulator who uses marketable permits will be satisfied *ex post* with the quantity of emissions that emerges or that he will be content *ex post* with the level of marginal control cost if he uses an effluent charge. On the contrary, if his estimate of the cost function turns out to have been imperfect, his original estimate of the optimal level of emissions reduction and the associated level of marginal control cost will both prove to be erroneous; he will wish, in retrospect, that he had been less successful in attaining his original target of whichever of the two he turns out to achieve. More specifically, we have

Proposition Two. When the position of the marginal cost curve is lower than expected, the emissions reduction will generally be inadequate under a system of permits and excessive under an effluent charge if both are set at what appear to be their optimal levels *ex ante*; the reverse will be true if the actual cost curve is higher than the expected one.

Figure 5.3 confirms these results. In the figure, the marginal benefits curve, BB , is, by assumption, known with certainty. The cost curve, in contrast, is now subject to uncertainty. We see part of the *anticipated* cost curve, C_a , and the associated "optimal" point, E_a . Assuming the regulator selects either the corresponding effluent fee, f , or the corresponding volume of permits, leading to an emissions reduction, q_p , we can now see the consequences if the true curve of marginal control costs, C_t , lies below C_a . First, we see that with the marginal benefit curve having a negative slope, the true optimum, point E_t , must lie below and to the right of the anticipated optimum, E_a . The optimal reduction in emissions, q_o , will thus always be greater than q_p , the quantity selected by the regulator under his misapprehension about costs. We also see that q_f , the emissions reduction achieved by effluent fee f , will be greater than either q_p or q_o , so that we must have $q_p < q_o < q_f$. This must be so if the BB curve has a negative slope and if the CC curve has a positive slope.

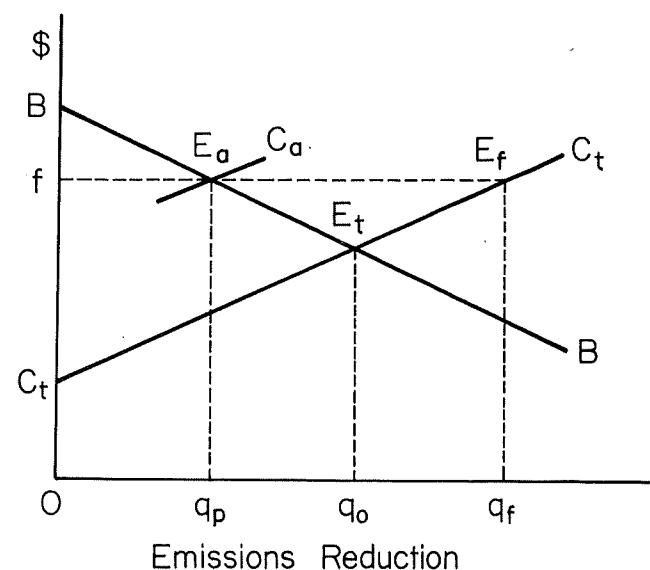


Figure 5.3

The intuitive explanation for the preceding result is straightforward. The emissions reduction q_p achieved under a system of marketable permits will be inadequate because it yields just the size of reduction the regulator initially thought to be optimal; the permit system offers no flexibility in adapting itself to the fact, which emerges subsequently, that additional emissions reductions are less costly than had been expected. The effluent fee, on the other hand, forces adoption of a level of q (denoted q_f) that incurs the same marginal cost as had initially been thought optimal; the fee approach does not adapt itself to the fact that at q_f the marginal benefit will have fallen below its level at q_o as a result of the diminishing marginal benefits to increased emissions reductions.

5 On the magnitudes of the relative distortions

Even though the true optimal value, q_o , lies between q_p and q_f , the emissions reductions achieved by marketable permits and effluent fees, respectively, do not lend any presumption that there will be anything close to equality either in the respective quantity distortions, $|q_p - q_o|$ and $|q_f - q_o|$ or in the resulting losses in consumers' and producers' surpluses. In both cases, the relative magnitudes or, rather, their expected values will depend on the shapes of the marginal cost and benefit curves and on the distribution of the random errors associated with the cost function.

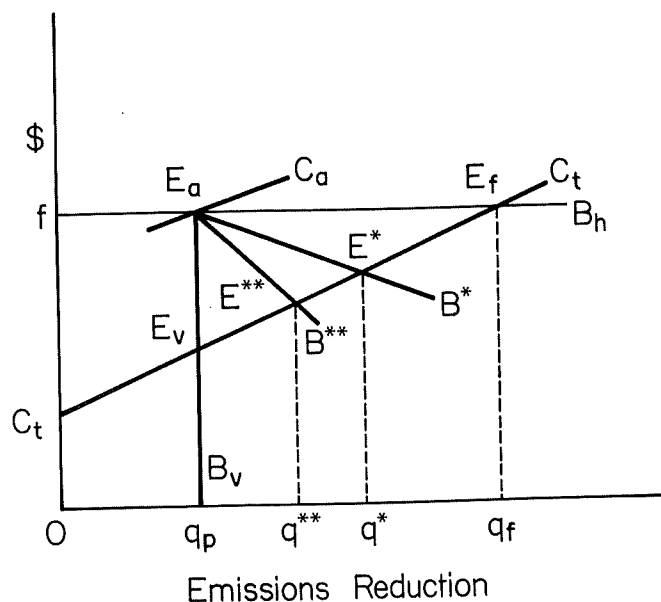


Figure 5.4

As an introduction to the relationship, we start with the quantity distortions and derive

Proposition Three. All other things being equal, the steeper the slope of the marginal benefits function, $MB(q)$ (i.e., the greater the absolute value of dMB/dq), the smaller will be the distortion $|q_p - q_o|$ resulting from regulatory error about the cost function under a system of marketable permits, and the greater will be the distortion $|q_f - q_o|$ yielded by an effluent fee.

This can once again be shown diagrammatically. Figure 5.4 depicts four marginal benefit functions, ranging from the horizontal benefits curve, B_h , to the vertical benefits curve, B_v , with B^* and B^{**} being of intermediate steepness. All four curves must go through E_a since, by hypothesis, they were known correctly by the regulator and so went through his estimated optimal point, E_a . Based upon the anticipated cost curve, C_a , the regulator would thus select a fee level, f , under a system of effluent fees or, alternatively, a quantity of permits, q_p , under a system of marketable permits. We see immediately that if C_t turns out to be the true cost curve, the fee approach will result in emissions reductions of q_f . We can now

compare the distortions under the two systems. For the extreme case of a perfectly horizontal marginal benefits curve, B_h , we see that the fee instrument achieves the true optimal outcome, q_f ; the distortion under the permit regime is, in contrast, relatively large, encompassing the entire range from q_p to q_f . In the other extreme case, a vertical marginal benefits curve, B_v , just the opposite is true; here, the permit approach produces the optimal outcome, and the fee system results in a large distortion ($q_f - q_p$).

The intermediate cases show that, starting from E_f , as the benefit curve grows steeper, the optimal point must move leftward along the true marginal cost curve, from E_f to E^* to E^{**} to E_v ; in short, it must move ever further from the effluent fee equilibrium, E_f , and ever closer to the permit equilibrium, E_v . Similarly, the optimal emissions reduction must move leftward, away from q_f and toward q_p .²

An identical argument shows that precisely the same relationships hold when the true marginal cost curve, C_t , lies above the estimated curve C_a . This completes the formal argument for Proposition Three.

Once again, an intuitive explanation is not difficult to provide. If the marginal benefits curve is declining very sharply, even a fairly severe fall in marginal costs will justify very little increase in the quantity of emissions eliminated since such an additional reduction will have little value to society. The same will obviously be true when the true marginal-cost curve turns out to be higher than the estimated cost curve (and the benefits curve is steep). In other words, in that case the quantity that was thought optimal in the erroneous *ex ante* calculation will still turn out to be very nearly correct, and so a system of marketable permits which enforces that quantity will turn out to produce a result very close to the true optimum.

On the other hand, when marginal benefits decline very little as q increases, then despite the error in the estimation of the cost function, the optimal value of the actual control cost will turn out to be very close to its estimated value *ex ante*, and the corresponding effluent fee will therefore come closer to yielding the desired results. Or, put slightly differently, if marginal benefits are relatively constant over the relevant range of waste emissions, then the fee, even though based on the anticipated function, will provide close to the right signal as a measure of external costs.

² A formal proof of the generality of the result is hardly necessary. C_t by assumption is a positively sloping curve that lies below and to the right of point E_a . Hence, if two negatively sloping curves emerge from E_a , and one uniformly has a greater absolute slope than the other, the former's intersection point with C_t must lie to the left of the latter's, and the result follows.

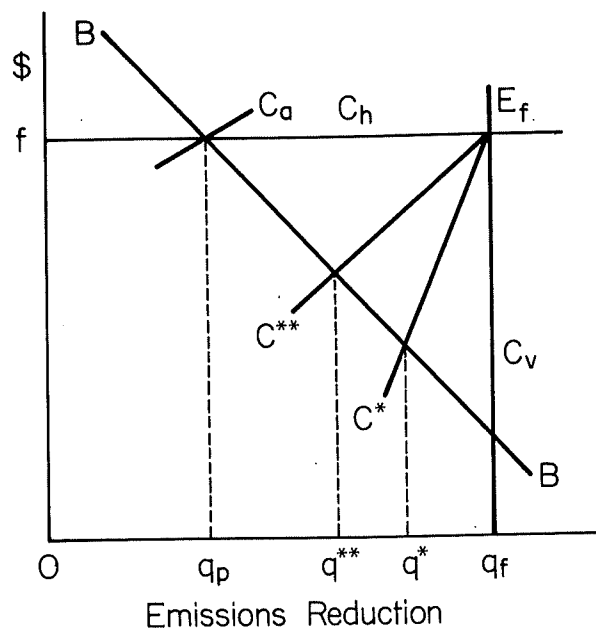


Figure 5.5

6 Effect of the slope of the cost function

In a similar manner we can derive

Proposition Four. All other things being equal, the steeper the curve of marginal control costs in the family of such curves meeting at q_f (the equilibrium value of the reduction in emissions under the effluent fee based on the erroneous cost estimate), the greater will be the distortion $|q_p - q_o|$ produced by a system of marketable permits and the smaller will be the distortion produced by the effluent fee.

The argument is indicated in Figure 5.5 and will only be sketched very briefly, since it is so similar to that of Proposition Three. In the figure, we see four cost curves through point E_f corresponding to effluent fee f and associated emissions reduction, q_f . As the true cost curve shifts from C_h to C^{**} to C^* to C_v (i.e., as it increases successively in steepness), the optimal value of q moves from q_p to q^{**} to q^* to q_f ; that is, it moves steadily further from the quantity that will be achieved under a system of permits and toward the quantity that will result under an effluent fee. We

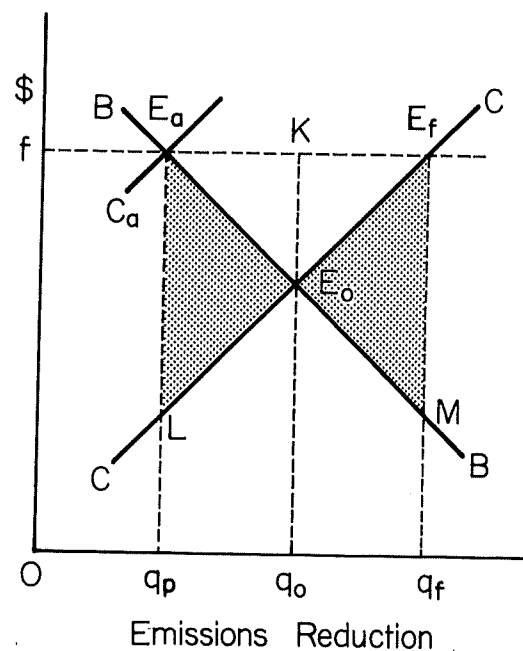


Figure 5.6

thus see that as the slope of the cost curve increases, the size of the distortion under the permit regime rises, whereas that under the fee system is diminished.

7 Relative slopes and the linear case

As a preliminary step toward a basic theorem first derived by Weitzman, we have

Proposition Five. When the marginal benefit and marginal cost curves are linear, marketable permits and effluent fees will produce the same absolute distortion when the regulator miscalculates marginal costs if the absolute values of the slopes of the two curves are equal. If the absolute value of the slope of the marginal cost curve is greater than that of the marginal benefit curve, effluent fees will lead to a smaller distortion, and *vice versa*.

The argument for Proposition Five is depicted in Figure 5.6, which shows the special case in which the slopes of CC and BB are equal in

absolute value (but no assumption is made about the magnitude of that slope). E_a is again the regulator's (mistaken) *ex ante* estimate of the optimum so that he either adopts effluent fee f or imposes emissions reduction q_p via a set of permits. The true optimum is q_o , corresponding to the intersection of the (true) cost and benefit curves, and the emissions reduction under the effluent fee is q_f . To prove Proposition Five, we extend the vertical line segment $q_o E_o$ to point K , where its height equals the value of the effluent fee. Then triangles $E_a K E_o$ and $E_f K E_o$ are congruent, since they are right triangles, share side $E_o K$ and have another angle in common because of the assumed equality of slopes of $E_a E_o$ and $E_o E_f$. Consequently, $E_a K$, the absolute distortion under permits, is equal to KE_f , the absolute distortion under the effluent fee.

This completes the proof for the case of equal slopes. The remainder of the result is an immediate corollary of what has just been shown and of Propositions Three and Four.

8 Relative slopes and consumers' and producers' surpluses

The true social damage from regulatory miscalculation is, of course, indicated by the consequent loss in consumers' and producers' surpluses, not by the distortion in emissions reductions with which we have been dealing so far. However, as we will see now, the analysis of the latter takes us a good part of the way toward analysis of the former. It is obvious that, in general, the permit approach will result in a loss of surpluses that differs, and can differ substantially, from the loss when the regulatory instrument is an effluent fee. This is illustrated in Figure 5.7. Here, q_p , q_o , and q_f indicate the reductions in emissions, respectively, under a system of permits, in an optimal solution, and under an effluent fee. The shaded areas represent the associated losses in consumers' and producers' surpluses. STE_a is the social loss under a system of permits and TUE_f is the loss under an effluent charge (both measured relative to the social optimum, q_o). In the diagram, the former loss is clearly greater than the latter, showing that the two need not be equal. Once again, their relative magnitude depends primarily on the slopes of the marginal benefit and marginal control cost curves. Indeed, here we have the fundamental theorem first derived by Weitzman:

Proposition Six. When the marginal benefit and marginal control cost curves are linear, where the former is known with certainty but an additive error term with zero expected value enters the equation for the latter (i.e., the cost curve), then a marketable permit system will produce the same expected welfare loss as an effluent fee regime if the slopes of the two marginal curves are equal. Otherwise, an effluent fee will be the pref-

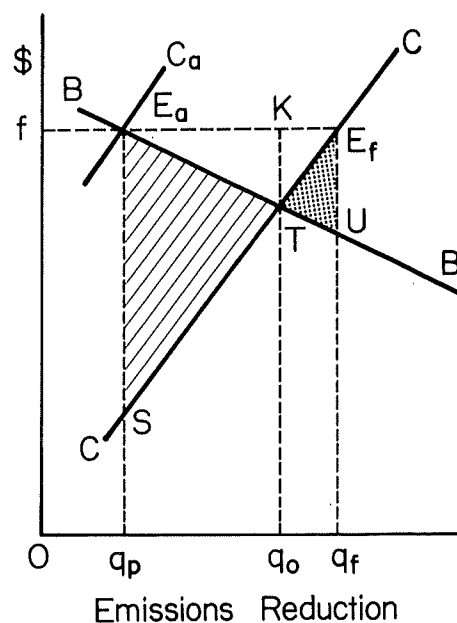


Figure 5.7

erable policy instrument to a risk-neutral regulator whose objective is welfare maximization if the marginal control cost curve is steeper than the marginal benefits curve, and vice versa. The choice is independent of the properties of the random element.

Note that this result is very similar to that in Proposition Five relating to the relative distortion of the value of q . As we will see in a moment, this is no accident. We shall provide a formal proof of this proposition shortly; however, a rough geometric argument leading to Proposition Six is now easily sketched.

We recall from Proposition Five that the absolute magnitude of the distortion under fees and under a permit system is the same if the slopes of the marginal benefit and cost curves are equal (in absolute value). In terms of Figure 5.6, we saw that this equality of slopes implied that $E_a K = KE_f$. But we can take this one step further. The equality in Figure 5.6 of $E_a K$ and KE_f implies that the shaded triangles, $E_a E_o L$ and $E_f E_o M$, have the same area.³ But these areas, of course, represent the welfare losses, respectively, under systems of tradable permits and effluent fees. More

³ This follows since the two triangles are necessarily similar and have the same height, $E_a K = KE_f$, measured from point K .

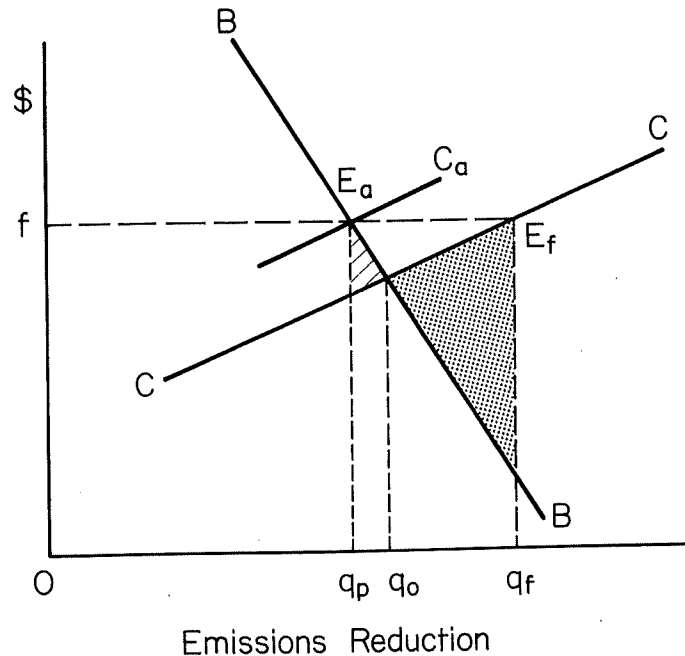


Figure 5.8

generally, as illustrated in Figure 5.7, the size of the distortion is related to the magnitude of the welfare loss as follows:

$$(E_a K)^2 / (KE_f)^2 = (\text{Area } E_a ST) / (\text{Area } TUE_f).$$

This is clear, because the two welfare loss triangles are necessarily similar and $E_a K$ and KE_f are their respective heights measured from point K .

We thus see in Figure 5.7 that when the marginal control cost curve is steeper than the marginal benefits curve, the expected welfare loss under the fee approach is less than the expected loss under a permit system. Similarly, as depicted in Figure 5.8, where the marginal benefit function is the steeper of the two curves, the permit instrument is the preferred policy because it promises a smaller welfare loss than a fee regime.

9 A more formal proof of Proposition Six

The geometric approach used so far in dealing with the propositions of this chapter suffers from some limitations, most notably from its inability

to take account of the influence of the stochastic properties of the regulator's uncertainty about the pertinent functions. We therefore summarize a formal derivation of Proposition Six, mostly to illustrate the method of approach generally adopted in the literature.

Weitzman's proof deals with the *total* cost and *total* benefit functions and assumes that the random error is sufficiently small to justify quadratic approximations to those functions, that is, linear approximations to the marginal cost and benefit functions. Adar and Griffin simply *assume* such linearity and their exposition is, consequently, a bit easier to follow. We will, therefore, employ the general procedure of the latter. For our notation, let

q = quantity of emissions reduction

u = a random error, $E(u) = 0$

E = the expected value operator

$$MB = a - bq = \text{the marginal benefit of } q \quad (1)$$

$$MCC = w + vq + u = \text{the marginal control cost} \quad (2)$$

and * denotes optimal value.

Then the objective of the regulator is to select either a reduction in emissions, q^* , achieved by a set of marketable permits, or an effluent fee, f^* , which yields the expected reduction in emissions that maximizes the expected value:

$$W = E \int_0^Q [MB(q) - MCC(q, u)] dq. \quad (3)$$

Our procedure will be to calculate optimal values of q^* and f^* and then, by substituting these successively into (3), to obtain expressions for the respective welfare gains under permits and effluent charges. The difference between these two expressions will be used to measure the expected net benefit of the one policy over the other; this will yield Proposition Six (and a bit more).

To find f^* , the optimal fee, we first solve for q as a function of f by noting that profit maximization requires $f = MCC = w + vq + u$, or

$$q(f, u) = (f - w - u)/v. \quad (4)$$

Next, we note that optimality requires the derivative of the welfare gain (i.e., the derivative of expression (3) with respect to f) to be equal to zero after substitution of the linear expressions (1) and (2) for MB and MCC , respectively. Integrating, we obtain

$$\frac{dW}{df} = \left(\frac{dq}{df} \right) \frac{dE}{dq} \left[(a - w)q(f^*, u) - \frac{(b + v)q(f^*, u)^2}{2} - uq(f^*, u) \right] = 0. \quad (5)$$

Because $E(u) = 0$, we have, after differentiating with respect to q ,

$$a - w - (b + v)E[q(f^*, u)] = 0$$

or

$$E[q(f^*, u)] = (a - w)/(b + v). \quad (6)$$

From (4), we have

$$E[q(f^*, u)] = (f^* - w)/v. \quad (7)$$

Equation (6), incidentally, indicates that f^* should be chosen so that the marginal benefits from emissions reductions equal the *expected* marginal control costs.

Similarly, we obtain q^* , the optimal quantity of permits, by differentiating (3) with respect to q , yielding

$$q^* = (a - w)/(b + v) = E[q(f^*, u)]. \quad (8)$$

From Equations (8) and (6), we see that both policies will yield the same expected value of q . However, they will not promise the same expected level of social welfare. To see this, we first calculate the welfare gain $W(q^*)$ offered by q^* , the optimal quantity of permits. Substituting q^* into (3), the linear expressions for benefits and costs, we obtain, by integration,

$$W(q^*) = (a - w)q^* - [(b + v)/2]q^{*2} - E(q^*u) \quad (9)$$

with the last term equal to zero because $E(q^*u) = q^*E(u) = 0$. To find $W(f^*)$, the welfare gain under an optimal effluent fee, we again integrate (3), obtaining the same expression as (9) but this time with (7) and (4) substituted for q^* , q^{*2} , and uq^* . We thus have

$$W(f^*) = (a - w)E[q(f^*, u)] - [(b + v)/2]E[q(f^*, u)^2] - E[uq(f^*, u)]. \quad (10)$$

Now, (4), (6), and (8) give us

$$q(f^*, u) = q^* - u/v$$

so with $E(u) = 0$,

$$W(f^*) = (a - w)q^* - [(b + v)/2][q^{*2} + E(u/v)^2] + E(u)^2/v$$

or, by (9),

$$W(f^*) = W(q^*) - E[(b + v)u^2/2v^2] + E[u^2v/v^2]$$

or

$$W(f^*) - W(q^*) = E(u^2)(v - b)/2v^2. \quad (11)$$

Equation (11) is the basic result given in Proposition Six. Since b and v are, respectively, the absolute slopes of the marginal benefit and the cost curves, (11) shows that where these slopes have the same absolute value, the expected welfare gains under the two policy regimes are equal. Where the slope of the marginal control cost curve, v , exceeds the absolute value, b , of the marginal benefits curve, the fee approach is to be preferred, and vice versa. Equation (11) also shows that $E(u^2)$, the variance of u , affects the magnitude of the difference between the welfare yields of the two policies, but it does not affect the choice of policy instrument.⁴

Of course, in the more general case, where the marginal cost and benefit functions are nonlinear, where the disturbance terms do not necessarily enter randomly, and where the regulator is not risk neutral, the results are less straightforward. The choice of instrument may now depend on the parameter values in the cost and benefit functions, the way in which the random variables enter the functions, and the frequency distributions of those variables. In such circumstances, qualitative generalizations are no longer easy to obtain.

10 On the choice between marketable permits and fees in practice

Proposition Six does not in itself establish any real presumption in favor of one of our policy instruments over the other. In particular cases, of course, where one has some empirical information about the shapes of the marginal cost and marginal benefit functions or even some *a priori* grounds on which to base an informed guess about their magnitudes, Proposition Six does provide some guidance to the regulator in choosing between them. But for the economy as a whole, we should not be surprised to find that each instrument will prove the better choice in some considerable number of cases.

Not everyone has drawn so noncommittal a conclusion. Weitzman, for example, expects that quantitative control instruments (such as marketable permits) will most frequently prove preferable. We quote his reasoning (which is obviously generalized well beyond the environmental issues that concern us here), leaving an evaluation to the judgment of the reader.

There is, it seems to me, a rather fundamental reason to believe that quantities are better signals for situations demanding a high degree of coordination. A

⁴ In Figure 5.7, $E(u^2)$ reflects the expected distance between C_a , the cost curve anticipated by the regulator, and CC , the true cost curve. The additivity of the error term implies that the two curves must be parallel, meaning that their distance from one another can be measured unambiguously.

classical example would be the short run production planning of intermediate industrial materials. Within a large production organization, be it the General Motors Corporation or the Soviet industrial sector as a whole, the need for balancing the output of any intermediate commodity whose production is relatively specialized to this organization and which cannot be effortlessly and instantaneously imported from or exported to a perfectly competitive outside world puts a kink in the benefit function. If it turns out that production of ball bearings of a certain specialized kind (plus reserves) falls short of anticipated internal consumption, far more than the value of the unproduced bearings can be lost. Factors of production and materials that were destined to be combined with the ball bearings and with commodities containing them in higher stages of production must stand idle and are prevented from adding value all along the line. If on the other hand more bearings are produced than were contemplated being consumed, the excess cannot be used immediately and will only go into storage to lose implicit interest over time. Such short run rigidity is essentially due to the limited substitutability, fixed coefficients nature of a technology based on machinery. Other things being equal, the asymmetry between the effects of overproducing and underproducing are more pronounced the further removed from final use is the commodity and the more difficult it is to substitute alternative slack resources or to quickly replenish supplies by emergency imports. The resulting strong curvature in benefits around the planned consumption levels of intermediate materials tends to create a very high comparative advantage for quantity instruments. If this is combined with a cost function that is nearly linear in the relevant range, the advantage of the quantity mode is doubly compounded (p. 487).

The Weitzman argument, however, does not seem applicable to many issues in the design of environmental policy. The case for the use of fees, in some instances, seems compelling. David Harrison, for example, in a study of the control of airport noise, finds that the marginal benefits from noise control are fairly constant over the relevant range.⁵ On the basis of the results embodied in Proposition Six in this chapter, Harrison recommends the use of fees, rather than marketable permits, to regulate noise levels at local airports. Similarly, it has been argued that the use of quantity instruments (in this instance, direct controls) for the regulation of automotive exhausts in the face of rapidly rising marginal control costs has resulted in large welfare losses through excessive severity of controls and the associated high costs. A fee approach might, in this instance, have reduced social costs significantly.

On the other hand, where the marginal benefits function is quite steep, close control over quantity becomes important. For various hazardous wastes, for example, a permit system may well be preferable since it provides greater assurance against excessive, and possibly highly destructive, emissions of such pollutants. We shall return to these issues in a somewhat

⁵ "The Regulation of Aircraft Noise," in Thomas C. Schelling, Ed., *Incentives for Environmental Protection* (Cambridge, Mass.: M.I.T. Press, 1983), pp. 41-144.

broader context in Part II when we consider the roles of various policy instruments in the arena of environmental decision-making.

11 Mixtures of instruments for pollution control

Up to this point, we have treated quantity and price instruments as alternative policy tools. However, Roberts and Spence have constructed an ingenious hybrid control instrument that employs marketable permits supplemented by an effluent fee and a subsidy. The fee serves as an escape mechanism to limit the detrimental consequences of extreme misjudgment of the optimal value of q and in the corresponding quantity of emissions permits issued. The system works as follows: the regulator issues a number of marketable emission permits, and on the market for those permits there emerges an equilibrium price per permit. Let us now use p to represent that price. At the same time, the regulator allows polluters to generate emissions without permits (or in excess of the quantity authorized by their permit holdings) but charges an effluent fee, f , per unit of such emissions. Finally, the regulator offers the polluter a subsidy, s , per unit for any unused permits, where $s \leq f$.

It is easy to show that in equilibrium we must have

$$s \leq p \leq f. \quad (12)$$

This is so because if p were greater than f , no one would purchase a permit but would pay the effluent charge instead, so p would have to fall. On the other hand, if s exceeded p , it would pay to purchase as many permits as were available and hold them unused at a profit of $s - p$ per unit; but obviously no one would be willing to sell a permit at that price.

It is also easy to see that by an appropriate choice of values of s and f , the mixed system can be transformed either into a pure permit scheme or a fee regime. It becomes a pure permit system if one sets $s = 0$, $f = \infty$ so that both the subsidy and effluent charge elements are effectively eliminated. It becomes a pure effluent charge of any desired magnitude, k , if one sets $s = f = k$ so that by (12) the price of a permit is driven automatically to that level.

It follows at once that if the three regulator-controlled parameters in the system, s , f and the number of permits issued, l , are selected so as to maximize expected welfare, the result must be at least as desirable as either a pure permit regime or a pure effluent fee. For if it should transpire that either of the latter is optimal, then the maximization calculation will automatically select the corresponding parameter values that effectively eliminate the mixed system.

There are at least two other ways to describe the reason for the superiority of the mixed system. As we have seen, the use of a permit system limiting effluent quantity to q^* gives rise to two possible dangers: First, the regulator's estimate of cost on which his choice of q^* is based may turn out to be far too high, in which case the selected quantity of effluent reduction, q^* , could be far below the optimum. Second, the cost estimate might be far too low, in which case q^* will be correspondingly excessive. The mixed system induces polluters to avoid both errors when they are extreme. For example, if actual marginal cleanup cost turns out to be higher than f , it will pay polluters operating under the mixed arrangement to emit more than the permits allow and to pay the fee, f , on all emissions that exceed those justified by their permit holdings. Similarly, if cleanup costs turn out to be lower than s , it will pay polluters to continue to reduce their emissions and hold their excess licenses unused in return for the subsidy payment. Automatically, then, whenever the permit system would have performed very badly, it pays the polluters under the mixed scheme to act in a way that transforms it into a fee regime.

The reason a pure effluent fee is apt to perform badly when the regulator is uncertain about the cost function is that the fee is a fixed number, f , which corresponds to the marginal benefits of emissions reduction *only* at one value of q – the one the regulator considers, *ex ante*, to be optimal. Suppose, instead, it were somehow possible and feasible to institute a variable total effluent fee $F(q)$ that varied with q and whose marginal payment always equalled the marginal social benefit corresponding to that value of q , so that $dF(q)/dq = MB(q)$. Then in our diagrams, instead of corresponding to a horizontal line, the marginal effluent payment curve would coincide with the marginal benefit curve. In that case, at any value of q not equal to the optimal value, a polluter would find that the marginal cleanup cost was not equal to the marginal effluent charge; it would then pay the polluter to alter his emissions in the direction of the optimum. Note that to operate such a scheme, the regulator needs absolutely no information about marginal control costs, so his uncertainty on that subject becomes irrelevant. A penalty function that coincides with the damage function will yield the optimal outcome irrespective of the level of control costs.⁶

Now, the mixed system of Roberts and Spence represents a compromise between the horizontal effluent curve, f , and the pure variable payment $F(q)$. Instead, it is a step function which constitutes an approximation to the marginal benefit curve, as shown in Figure 5.9. We see there

⁶ For another such scheme, see Robert A. Collinge and Wallace E. Oates, "Efficiency in Pollution Control in the Short and Long Runs: A System of Rental Emission Permits," *Canadian Journal of Economics* XV (May, 1982), 346–54.

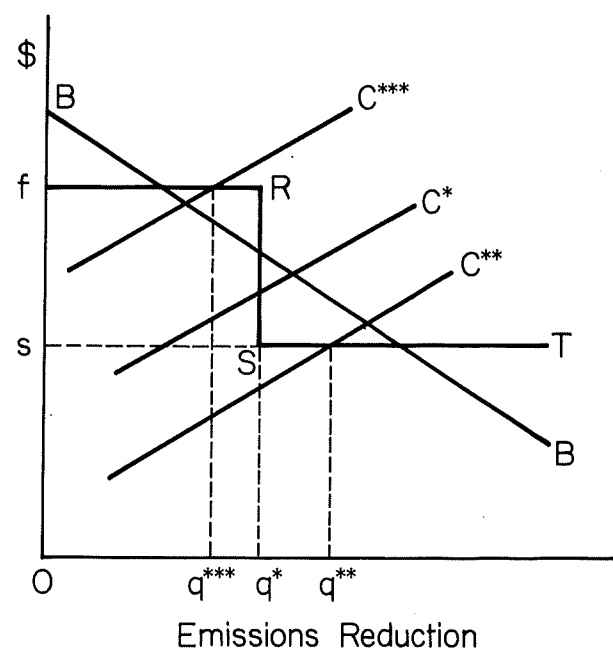


Figure 5.9

the three regulatory decision variables, q^* , f , and s . We also see that for an emissions reduction less than the prescribed quantity, q^* , there is an effluent fee, f , whereas for emissions reductions greater than q^* , incremental emissions have a low opportunity cost (equivalent to an effluent charge), s . Along the vertical segment SR , the effective fee is some value p , where $f \geq p \geq s$. It is clear that the implicit effluent locus $fRST$ is a better approximation to the marginal benefit curve, BB , than is any horizontal line. We also see how extreme errors in the regulator's estimate of marginal control costs (like curves C^{**} and C^{***}) can lead to adaptations in the value of q , unlike a pure permit system.

Roberts and Spence go on to show how one can design an effluent fee system whose graph constitutes a number of steps (unlike the one-step curve $fRST$ in Figure 5.9). With such a procedure, effluent payments can be made to approximate the marginal benefits curve, BB , even more closely, leading, in general, to an expected welfare level that is higher still. In the limit where the number of steps goes to infinity, we would, of course, have the variable fee system that we described earlier and which produces the optimal outcome.

12 A final note

In this chapter, we have found that in the presence of particular forms of uncertainty, fee and permit systems are unlikely to produce optimal outcomes and, moreover, may produce quite different results. We note in concluding this discussion that the analysis in this chapter has been wholly static in character. Moreover, we have considered only "once-and-for-all" choices of the policy variables. The results would be "softened" somewhat if we were to allow the environmental authority to amend policies that turn out, *ex post*, to involve significant welfare losses. Environmental regulators have, for example, revised environmental targets and rules for compliance. There are, of course, costs associated with changes in policies, and, in certain instances, policy decisions may be virtually irreversible in terms of their effects. However, for a wide range of environmental policy choices, there exists some opportunity for later corrections.

We raise this matter in anticipation of the analysis in Part II, where we will explore the design and implementation of environmental policy in a somewhat more realistic setting that incorporates a more diverse set of criteria for the selection of policy instruments. Expected welfare gains and losses in the standard static sense employed in this chapter will figure as an important consideration in the design of policy measures – but certainly not the only consideration.

CHAPTER 6

Market imperfections and the number of participants

In Chapter 4 we derived our results for optimal taxes and payments from a competitive model in which individuals and firms both behave as price-takers. In this framework, prices *and* our prescribed fees are parameters for individual decision-makers; they take them as given and simply respond so as to maximize utility or profit.

In this chapter, we consider some of the complications that market imperfections introduce into the analysis. More specifically, we will examine the implications of two sources of such imperfections. First, a firm that generates externalities (smoke emissions) may not sell its output in a competitive market. For example, we will consider how a profit-maximizing monopolist will respond to the Pigouvian taxes prescribed in Chapter 4. We will show that an emissions tax rate that is appropriate for the pure competitor will not, in general, induce behavior that is consistent with optimality in the second-best world inhabited by a monopolist.

We then consider a second source of imperfection: the presence of polluters who are not "fee-takers." There can be situations involving few polluters, the manipulation of whose activity levels can influence the *unit* tax paid on waste emissions. In such cases, we will see that producers (and perhaps also consumers) of externalities will have an incentive to adjust their behavior so as to influence not only their tax bills, but also the tax *rate* they pay per unit of pollution. As for the monopolist, this necessitates some modifications in the prescription for an optimal fee. In fact, we will find one case to which the Coase result calling for a tax on victims is applicable.

Finally, we offer a proposition that relates, not only to the number of producers of an externality, but also to the number of consumers or victims. This establishes a strong presumption that increasing marginal costs will characterize many sorts of externality-generating activities (particularly those involving congestion).

1 Externalities produced by a monopolist

We found in Chapter 4 that a unit fee on the polluting activities of competitive firms equal to the costs at the margin that they impose on other