10-6-1997

On the Superiority of Corrective Taxes to Quantity Regulation

Louis Kaplow
Harvard Law School

Steven Shavell
Harvard Law School

Follow this and additional works at: http://lsr.nellco.org/harvard_olin
Part of the Law and Economics Commons

Recommended Citation
http://lsr.nellco.org/harvard_olin/215

This Article is brought to you for free and open access by the Harvard Law School at NELLCO Legal Scholarship Repository. It has been accepted for inclusion in Harvard Law School John M. Olin Center for Law, Economics and Business Discussion Paper Series by an authorized administrator of NELLCO Legal Scholarship Repository. For more information, please contact tracy.thompson@nellco.org.
On the Superiority of Corrective Taxes to Quantity Regulation

Louis Kaplow and Steven Shavell

Abstract

The traditional view of economists has been that corrective taxes are superior to direct regulation of harmful externalities when the state’s information about control costs is incomplete. In recent years, however, many economists seem to have adopted the view that either corrective taxes or quantity regulation could be superior to the other. One argument for this view, identified with Weitzman (1974), holds only if the state is constrained to use a fixed tax rate (a linear tax schedule) even when harm is nonlinear. Corrective taxes are indeed superior to quantity regulation if -- as seems more plausible -- the state can impose a nonlinear tax equal to the schedule of harm or can adjust the tax rate upon learning that it diverges from marginal harm. Another argument, associated with Baumol and Oates (1988), is that quantity regulation gains appeal when the state is uncertain about the harm caused by an externality. In this case, however, a corrective tax schedule (equal to the expected harm schedule) is superior to quantity regulation.

JEL Classes H23, D62, K32, L51

---

*Harvard University and National Bureau of Economic Research. We thank William Baumol, Lucian Bebchuk, David Bradford, Don Fullerton, Richard Revesz, Robert Stavins, Kip Viscusi, Martin Weitzman, and workshop participants for comments, George Wang for research assistance, and the John M. Olin Center for Law, Economics, and Business at Harvard Law School for financial support.
In the past, the consensus view of economists was that corrective taxes are superior to quantity regulation as a means of controlling harmful externalities. This belief, however, has changed in recent years. Due in large part to the influence of Weitzman (1974) and Baumol and Oates (1971, 1988), economists now commonly state that either corrective taxes or quantity regulation could be superior to the other. This newer view is presented in textbooks and surveys on environmental economics (Bohm and Russel, 1985, Cropper and Oates, 1992, Pearce and Turner, 1990, Tietenberg, 1996) as well as in general undergraduate textbooks (Pindyck and Rubinfeld, 1995, Rosen, 1992).

Our purpose here is to explain why, in fact, the traditional notion of the superiority of corrective taxes should continue to be a benchmark for economists’ thinking about the control of externalities. We first consider the standard context of a single externality-producing firm and review the conventional argument that corrective taxes are superior to quantity regulation. We observe that the contrary claim in Weitzman (1974) and related literature depends entirely upon the assumption that the corrective tax is constrained to be fixed and linear even when harm is nonlinear. The rationale for this restriction on taxes, though, is not evident, and we suggest that it is generally inappropriate. We next consider the idea, associated with Baumol and Oates (1971, 1988), that quantity regulation may be desirable when the state is uncertain about the magnitude of harm. As we explain, however, corrective taxes (equal to expected harm) remain superior to quantity regulation even when harm is uncertain.

We also examine the choice between corrective taxes and quantity regulation (and permit schemes) in situations in which multiple firms jointly create an externality. Although our discussion of this case relies in part on informal judgments, it suggests that corrective taxes (and
modified permit schemes) possess the same basic advantage over quantity regulation as in the single-firm case: they harness firms’ information about control costs, making possible a result in which the level of the externality is optimal (or more nearly so).

**Single Firm: Control Costs Unknown to the State**

*The traditional view of the superiority of corrective taxes.* We begin with the classic problem of a harmful externality generated by a single firm. In this case, we make the usual assumption that the schedule of harm as a function of the level of the externality is known to the state, whereas the firm’s control cost schedule is known only to the firm.¹ We also presume, as is conventional, that the marginal harm schedule rises with the level of the externality. This is illustrated in our figure, where e is the level of the externality, MH(e) is the marginal harm schedule, and MC_H(e) and MC_L(e) are two possible marginal control cost schedules.

---

¹Control costs should be interpreted as including not only direct expenditures made to abate the level of the externality but also forgone profits from reduction of output. (In some literature on externalities, the control cost schedule is referred to as the benefit schedule because firms benefit from not having to control pollution.)
The traditional view of externalities and corrective taxes, due to Pigou (1932), is that imposing a tax equal to the magnitude of the harm caused by the externality will induce efficient behavior because under such a scheme the firm will bear the full social cost of its behavior. In the present context, this means that the state should set the corrective tax schedule equal to the harm schedule. In other words, the marginal tax rate should equal MH(e). Then, the first-best outcome will always be achieved. The reason is that the firm will choose the level of externality where its marginal control cost -- whether MC_H(e) or MC_L(e) -- equals the marginal tax rate, which by construction equals the marginal harm.

However, the first-best outcome generally will not result under quantity regulation. Under quantity regulation, the state announces a limit on the level of the externality, say, e'. But because the control cost schedule is not known to the state, it cannot calculate the optimal level of the externality for the particular firm, and it typically will make a suboptimal choice. For example, if the marginal control cost schedule is MC_H(e), e' will be suboptimal because the marginal control cost will exceed the marginal harm.
The reason that the corrective tax is superior to quantity regulation is evident. The corrective tax effectively harnesses the firm’s information about its control costs. Although the firm does not directly reveal its information under the corrective tax, it does choose the level of the externality at which its actual marginal control cost equals the marginal tax rate, which in turn equals the marginal harm, and thus it behaves optimally.

The Weitzman argument that quantity regulation may be superior to corrective taxes. In Weitzman (1974) and related literature, it is assumed that, when the state employs a corrective tax, it must use a linear tax schedule (a fixed tax rate) even though harm may be nonlinear. In general, this produces a suboptimal result. For example, suppose that the state chooses the tax rate t (see our figure) and the control costs are $MC_H(e)$. Then, the marginal harm exceeds the marginal control cost at the level of the externality chosen by the firm.

Because the use of linear corrective taxes, like the use of quantity regulation, involves error, either corrective taxes or quantity regulation may be superior. As shown by Weitzman (1974) and others, under certain conditions, corrective taxes are superior when marginal control cost schedules are steeper than the marginal harm schedule, and quantity regulation is superior in the converse case.

Somewhat surprisingly, the assumption that taxes must be linear either is only casually motivated or is not even mentioned in the relevant literature. Moreover, imposing nonlinear corrective taxes seems straightforward. All that the state has to do is announce its nonlinear tax schedule, and all that the firm has to do is compute its taxes from the schedule (or the firm could

---

2 Results similar to Weitzman’s were developed by Adar and Griffin (1976), Fishelson (1976), Roberts and Spence (1976), and Rose-Ackerman (1973).

3 For example, Weitzman (1974, p. 481) briefly justifies limiting his analysis to linear prices on the ground that they are “simple messages, easily comprehended, traditionally employed, and frequently contrasted.” Myles’s (1995) graduate text refers to the “considerable administrative difficulties” of a nonlinear scheme but does not identify them.
read its taxes from a table, as taxpayers actually do with the nonlinear personal income tax). Thus, the state should be able to achieve the optimal outcome with corrective taxes.

We also note that the state needs less information to impose the optimal nonlinear corrective tax than it needs to implement either a linear tax or quantity regulation. To determine the nonlinear tax schedule, the state only needs to know the harm schedule. (We discuss below the situation where the state does not know the harm schedule, in which case the state only needs to know the expected harm schedule.) In contrast, to determine the optimal linear tax or the optimal quantity -- even though each is just a single number -- the state must know both the harm schedule and the distribution of possible marginal control cost schedules, and it must make calculations to determine the (second-best) optimal linear tax rate or quantity level.

---

Cropper and Oates’s (1992) survey presents Weitzman’s argument without mentioning the assumed restrictions on the state.
Finally, suppose that for some reason the state is constrained to choose between using a linear tax and quantity regulation. Then we wish to observe that a linear tax is more attractive than is suggested by our discussion thus far. Realistically, firms usually cause externalities over time, not at one moment. When the state collects taxes from a firm, the state learns the level of the externality and thus can make inferences about the firm’s marginal cost schedule. Using this information, the state can adjust the tax rate, allowing a higher level of welfare to be achieved.4

*Other criticisms of corrective taxes.* It has been suggested -- again assuming that corrective taxes are linear and fixed -- that changes in control costs or in harm pose problems for corrective taxes. See, for example, Butler and Maher (1982) and Rose-Ackerman (1973). In the case of changes in control costs, however, a preexisting optimal nonlinear tax would automatically produce an optimum: the firm would choose a different, but efficient level of the externality because the firm would continue to equate its (now different) marginal control cost to the marginal harm. We note, by contrast, that under quantity regulation the permitted level of the externality would have to be altered when control costs change.

In the case of changes in the harm schedule, a new tax schedule equal to the new harm schedule would produce the optimum. Note that the state would also have to change the allowed level of the externality under quantity regulation.

In the latter case, the state would wish to adjust taxes as soon as the harm schedule changes. This may not always be possible, and some take this to be an argument for quantity regulation. Weitzman (1974) and Baumol and Oates (1988), for example, cite emergencies as prototypical situations in which quantity regulation may be superior to corrective taxes. But rapid adjustment of quantities also will be difficult in emergencies. If there is, say, a sudden

---

4The firm will behave strategically, anticipating the effect of its choice of e on the level of the tax, and the state’s optimal adjustment strategy should take this into account. Accordingly, it will not generally be possible to achieve the first-best outcome, as can be done with a nonlinear tax.
atmospheric inversion that makes emissions unusually harmful on a particular day, we do not see why instructing firms to reduce their emissions by a given amount could be done more quickly than informing firms of a higher tax on emissions.

**Single Firm: Harm and Control Costs Unknown to the State**

The superiority of corrective taxes to quantity regulation has also been questioned on account of the state having only imperfect information about the harm caused by an externality. If the harm schedule associated with an externality is uncertain, the state cannot set the tax schedule equal to the schedule of actual harm. Baumol and Oates (1971, 1988) are most associated with the view, which seems to be part of our folklore, that quantity regulation therefore gains appeal. (We surmise that a motivation for this view is that quantity regulation somehow provides better protection against a large harm.)

However, the basic logic favoring corrective taxes -- that they harness a firm’s information about its control costs -- applies in the case in which the schedule of harm as well as the control cost schedule is unknown to the state. In this case, the corrective tax schedule should equal the *expected* harm schedule. Of course, the outcome under corrective taxes can no longer be first best: if marginal harm is in fact greater than its expected value, the level of the externality will be too high, and conversely if marginal harm is lower than its expected value. A corrective tax equal to expected harm is, however, second best (that is, best given that harm is uncertain), and it is superior to quantity regulation.

To demonstrate that a corrective tax equal to expected harm is second best, consider the following standard formulation of the state’s planning problem. The state knows that the harm caused by the externality is given by the function \( h(e, \eta) \), which depends upon the level of the externality, \( e \), and an unobservable parameter \( \eta \), which has density \( f(\eta) \). The level of the
externality, e(x,θ), depends upon a firm’s expenditures to control the externality, x, and a parameter θ, which has density g(θ); θ might correspond to the efficiency of expenditures x in reducing the level of the externality. The firm observes θ before it chooses x, but the state does not observe θ.

The state chooses a tax schedule t(e) that may depend only on the level of the externality, because the state cannot observe η or θ. Given the tax schedule, the firm selects x to minimize its total costs (control cost plus tax obligation),

\[ x + t(e(x,\theta)). \]

Thus, the firm’s choice, denoted x(θ,t), is a function of θ and the tax schedule t. The state will select the tax schedule t to minimize expected social costs (the sum of control costs and harm),

\[
\int [x(\theta,t) + \int \eta f(\eta) d\eta] \int g(\theta) d\theta,
\]

where the state takes into account that e depends on θ and on x (which the firm is known to choose after it observes θ).

Clearly, the state can do no better than could a benevolent dictator who observes θ and then commands the firm to choose an x. Since such a dictator knows θ, he would choose x to minimize

\[ x + \int h(e(x,\theta),\eta) f(\eta) d\eta, \]

the sum of the control cost and the expected harm. But the state (without knowing θ) can induce the firm to choose this same x. The state can do so by setting the corrective tax schedule t(e) equal to the expected harm schedule:

\[ t(e) = \int h(e,\eta) f(\eta) d\eta. \]

Then the firm, whose problem as expressed above is to choose x to minimize x + t(e(x,θ)), will
be led to minimize

\[ x + \int h(e(x, \theta), \eta) f(\eta) d\eta, \]

which is exactly the dictator’s problem. Thus, setting the tax schedule equal to the expected harm schedule achieves the second-best optimum.

This analysis reveals that the intuition from the certainty case -- that a corrective tax allows the state to harness the firm’s information about control costs -- is applicable to the case in which harm is uncertain. The only differences here are that the optimal corrective tax equals expected harm and that the outcome is second best rather than first best.

The result that the optimal corrective tax equals expected harm implies that quantity regulation is inferior to this corrective tax. The reason is that quantity regulation is equivalent to a particular nonlinear tax, one that equals zero for all levels of the externality below some stipulated quantity and infinity (i.e., some prohibitive amount) for levels above that quantity. Given the extent to which quantity regulation diverges from a tax equal to expected harm, one would expect quantity regulation to be substantially worse than the optimal corrective tax. Thus, there is no basis for the view that quantity regulation may be desirable when uncertainty makes it impossible to set a tax schedule equal to actual harm.

The result that the optimal tax schedule equals the expected harm schedule depends on the assumption that harm and control costs are uncorrelated. If they were correlated, the optimal corrective tax schedule would be somewhat different, but would again be superior to quantity regulation. Regardless, the assumption of no correlation usually seems appropriate: the factors

\[ \text{Roberts and Spence (1976) observe that quantity regulation is an instance of a nonlinear tax.} \]
\[ \text{The optimal scheme would take into account that the firm’s choice of e conveys information about its control costs, which in turn are now assumed to be correlated with harm and thus bear on the optimal marginal tax rate, and it would also reflect that the firm may behave strategically in choosing e.} \]
\[ \text{Stavins (1996) and Weitzman (1974), among others, show that positive correlation would tend to favor quantity} \]
that determine harm, such as how a pollutant affects the human body, would seem to have little connection with the factors that determine control costs, such as the feasibility of various abatement technologies.\textsuperscript{7}

**Multiple Firms**

Here we discuss situations involving multiple firms that jointly create an externality, such as firms in a city that discharge sulfur dioxide, the total quantity of which determines pollution harm. In this setting, we consider corrective taxes versus quantity regulation as well as corrective taxes versus permit schemes, which constitute a form of quantity regulation. We assume that control costs are unknown to the state but that harm is known. (Our conclusions would be similar were we also to allow for uncertainty about harm, for essentially the reasons given in the single-firm case.)
Corrective taxes versus quantity regulation. When multiple firms together contribute to an externality, then in strict logic corrective taxes cannot implement the first-best outcome. To do so, the state would have to confront each firm with a tax schedule under which the marginal tax rate equals the marginal harm, which now depends on the total level of the externality. But, at the moment when any particular firm makes its decision about the level of the externality to generate, neither it nor the state knows this total, which is a function of the decisions of all the firms.  

Nevertheless, we suggest that, as a practical matter, nonlinear corrective taxes retain their general advantage over quantity regulation. It is true that because firms’ taxes depend on the total level of the externality, firms would not learn their tax obligation until the state received reports of all firms’ quantities and summed them; thus, firms might not know their tax bill until the end of, say, a one-week period. But it seems plausible that, after a modest number of periods, firms would be able to make reasonably accurate estimates of their tax obligations (just as firms generally are able to make reasonably good estimates of next week’s prices for many inputs even though these prices will depend on the demands of other firms).

The case of multiple firms has received some attention in the literature comparing corrective taxes and quantity regulation. Mendelsohn (1984) suggests that the presence of many firms might complicate the use of nonlinear taxes, but he does not explore the problem. Karp and Yohe’s (1979) model of the multiple-firm case assumes linear, fixed taxes. Weitzman (1978) examines fixed nonlinear taxes, but he requires that each firm’s tax be independent of other firms’ actions. We will explain that, in practice, these assumptions limit the benefits from corrective taxes.

Further, we remark that there is a mechanism that can be used to solve the problem under discussion without the need for iteration. Specifically, each firm initially submits a demand schedule indicating its level of externality as a function of the tax rate. The regulator aggregates these schedules, determines the proper marginal tax rate, and thereby tells each firm its level of externality and its tax bill -- all in advance of firms’ actions. This mechanism, as well as that in the text, works nearly perfectly as the number of firms is large. With small numbers, there will be strategic behavior, as we note in the single-firm case in note 4. For a mechanism that induces truthful revelation even with a small number of firms, see Dasgupta, Hammond, and Maskin (1980).
Instead of imposing a nonlinear corrective tax, the state might initially impose a linear tax. In this case, because the tax rate is constant, firms would immediately know their tax obligations, but the induced level of the externality would typically be inefficient. Accordingly, after each period the state could adjust the tax rate in the light of the reported total level of the externality, as we discussed in the single-firm case. One supposes that the tax rate would soon be close to the marginal harm at the prevailing total quantity, so optimality would be approximately achieved.

Now, contrast the case of quantity regulation. Not only will the level of the externality generally be suboptimal when it is chosen by the state, but also the state will not be able to correct its mistakes over time. This is because -- unlike with corrective taxes -- the state cannot learn anything about firms’ control costs from the level of externality that they generate: firms simply do as they are told. Therefore, if one were to consider the extended periods over which most pollution control schemes operate -- such as the decades since the passage of Clean Air Act and the indefinite future -- it seems likely that the advantage of corrective taxes over quantity regulation would be substantial.

Corrective taxes versus permit schemes. Under conventional tradeable permit schemes, the state issues a fixed quantity of permits, each one of which authorizes the holder to generate a stated amount of the externality; these permits may be traded among firms, giving rise to a market in the permits. Permit schemes are a form of quantity regulation because the total level of
the externality is determined by the quantity of permits issued by the state. As a result, permit schemes are likely to be inferior to corrective taxes for the reason we have emphasized throughout: under permit schemes, the state chooses the level of the externality without knowing firms’ control costs, whereas under corrective taxes firms’ decisions determine the level of the externality.

As an aside, we note that, when control costs differ among firms, permit schemes do have a well-appreciated advantage over regulation of the externality at the level of the individual firm. Namely, the costs of achieving a given total level of the externality will be minimized because all firms face the same permit price and thus firms’ marginal control costs are equated. Yet it remains true that this total level is stipulated by the state.

The state need not, however, be confined to using conventional permit schemes. In particular, the state can employ modified schemes that harness the information contained in market prices for permits (the market price will equal each firm’s marginal control costs). Under such modified schemes, the quantity of permits in use would be adjusted in the light of the price at which permits trade. We now describe two types of modified permit schemes that possess the advantages of nonlinear corrective taxes. (It should not be surprising that modified permit schemes may have the advantages of nonlinear corrective taxes, for permit schemes in which the total level of the externality is made a function of the permit price are a kind of dual to nonlinear corrective taxes, namely, prices that are a function of the total level of the externality.12)

---

12For a general discussion of quantity-dependent pricing, see Spence (1977).
Under a straightforward modification of a conventional permit scheme, the state would alter the quantity of permits over time. Suppose that, given the currently issued quantity of permits, the equilibrium permit price exceeds marginal harm. Then, since the permit price equals the marginal control costs of firms, the state knows that marginal control costs exceed marginal harm. Therefore, it is desirable for the state to allow firms to increase the level of the externality. This can be done by issuing more permits. Conversely, if the equilibrium permit price is less than marginal harm, the state should reduce the number of permits. This process of adjustment in the number of permits will end when the permit price equals marginal harm, and optimality is thus achieved.

Under a rather different type of permit scheme, the state would issue tradeable permits with different “exercise” prices. In particular, permit 1 would have a usage fee -- paid to the state -- equal to the marginal harm caused by the first unit of the externality; permit i would have a usage fee equal to the marginal harm caused by the i’th unit of the externality. (Note, therefore, that the permit scheme embodies a nonlinear corrective tax equal to the marginal harm schedule.)

In the equilibrium of the market for the permits, some quantity, q, of permits will be exercised. Each of these permits obviously must have the same effective price -- market price plus usage fee -- and permit q must have a market price of zero. This implies that all firms will equate their marginal control costs to the usage fee of permit q, which by design equals the marginal harm at the equilibrium level q of the externality. The permit scheme thus results in an equilibrium that

---

13 Firms’ anticipation of this possibility would affect the scheme. We commented on this issue as it arises with respect to tax rate adjustments in note 10.
14 A familiar complication, which we do not consider here, is that, if the harm schedule is not convex, a local optimum (reached through iteration) may not be the global optimum.
15 Such mechanisms are analyzed by Collinge and Oates (1982), Laffont and Tirole (1996), and Roberts and Spence (1976, appendix).
16 Permits i > q will be unexercised and worthless, and each permit i with i < q will have a positive market price that equals the usage fee of permit q minus its own usage fee. Note that for this scheme to work, the state should issue a
is optimal.\textsuperscript{17}

**Conclusion**

Economists traditionally have favored the use of corrective taxes to reduce harmful externalities because taxes leave control decisions in the hands of individual firms, which have better knowledge of their own control costs than does the state. But economists have recently come to believe that either corrective taxes or quantity regulation could be superior to the other. One argument for this view rests on the assumption that corrective taxes must be linear even when harm and thus the optimal corrective tax is nonlinear. However, as we discuss, nonlinear corrective taxes are often easy to impose, particularly in the case of externalities generated by a single firm. Further, even if linear taxes are employed, it would seem that tax rates could be adjusted to improve welfare. Moreover, we explain that conventional permit schemes may be modified to simulate nonlinear corrective taxes.

We have also discussed the notion that quantity regulation might be superior to corrective taxes on account of uncertainty about the magnitude of harm. Yet, as we indicated, in this case a nonlinear tax schedule equal to the expected harm schedule is (second-best) optimal, not quantity regulation.
Of course, there are a number of factors bearing on the problem of externalities that we have not considered, such as administrative costs and political considerations. Nevertheless, we believe that the core intuition indicating the superiority of corrective taxes to quantity regulation is robust and provides the appropriate starting point for thinking about policy.

\footnote{With regard to the latter, see Keohane, Revesz, and Stavins (1997).}
References


Laffont, Jean-Jacques and Tirole, Jean. “Pollution Permits and Environmental Innovation.”


