Pigouvian Taxes as a Long-run Remedy for Externalities

Henrik Vetter

Abstract: It has been suggested that price taking firms can not be regulated efficiently using Pigouvian taxes when damages are sensitive to scale effects. This conclusion follows because the Pigouvian tax cannot simultaneously get right the number of firms and the output per firm. In this paper, we argue that a simple nonlinear unit tax is efficient under standard assumption about demand and damages.

Jel No.: D61, D62.

Keywords: Externalities, Taxes.
1 Introduction

A way to deal with externalities is to impose a per-unit tax on the externality producing action. On this issue Carlton and Loury (1980, 1986) argue that a Pigouvian tax alone is inefficient in the long run when there are scale effects in the damage function. This follows because each firm, under a tax, produces until all economies of scale is exhausted, and this production volume is only coincidently optimal when scale matters.

Of course, when the Pigouvian tax is combined with a lump-sum tax or a lump-sum subsidy, it is possible to realize an efficient outcome. But lump-sum instruments are not practical because they are based on the presumption that the regulator is essentially not constrained by informational limitations. Even if these concerns were somehow easily dismissed, Baumol and Bradford (1970), and later Baumol (1979) alone, argue that lump-sum transfers can be ruled out by the economy-wide budget. We rule out lump-sum transfers as a general rule and ask if it is the upshot of the results reported by Carlton and Loury (1980, 1986) that one should dismiss unit taxes on grounds of inefficiency. More precisely, as mentioned by Carlton and Loury (1980), a unit tax with some exempting level can to a certain degree replace a unit tax without exempting and a subsidy. We ask in this paper if such a nonlinear unit tax—the nonlinearity being the exemption level—will correct pollution problems in common circumstances, i.e., when demand is negatively sloped and when marginal pollution is positive and escalates at an increasing rate.

In perfectly competitive markets, entry and exit of firms regulate production per firm to the point where each firm fully utilizes all economies of scale. This implies that average and marginal costs are equal and, in turn, that profit is zilch when price equals marginal cost. It is straightforward that lump-sum transfers are needed when concern about damage makes it optimal to have price taking firms produce at a point of either increasing or decreasing scale. If firms ideally should operate at a point of decreasing returns a unit subsidy
and a lump sum tax is called for. Oppositely, if firms should operate at a point of increasing returns a unit tax and a lump sum subsidy is called for. What we argue in this paper is that the most widespread assumptions about damage functions imply that firms should ideally operate at a point of increasing scale. In turn, this means that firms must be given a subsidy. We show that this subsidy, as noted by Carlton and Loury (1980), for practical purposes can take the form of a tax exemption threshold. That is, the proper Pigouvian tax will most often provide the correct incentive in the short run as well as in the long run.\footnote{When taxes are constructed this way it is obvious that the problems listed by Baumol and Bradford (1970), and Baumol (1979) do not arise.} It must be stressed that our results can be found in passing in Carlton and Loury (1980), but the practical implications—that a Pigouvian tax is a proper instrument in most circumstances—are not noticed.

2 Results

We consider a competitive industry where the number of firms, called $n$, is regulated by the entry and exit of firms until the profit of each active firm is zero. Private production costs are given as a function of firm output according to $C(q)$, and average production costs per firm, $C(q)/q$, are assumed to be U-shaped. It is supposed that each firm utilizes all economies of scale when it produces $q_s$. Clearly, the postulation that the market is perfectly competitive requires us to assume that actual production per firm is negligible relative to aggregate output. Since $q_s$ is what the untaxed firm produces, the assumption of perfectly competitive products markets is equivalent to the assumption that $q_s$ is small relative to aggregate production. With respect to regulation the assumption of perfectly competitive markets extends straightforwardly to regulated markets when output per firm goes down and the number of active firms goes up, while it is not straightforward that regulated markets
continue to be perfectly competitive if output per firm is regulated in the upwards direction while there are fewer firms.

Production has external effects, and we assume that damage is a function of firm output and the number of firms according to \( D(n, q) \). Under standard assumptions damage increases at an escalating rate when production increases, that is, \( D_q(n, q) > 0 \) and \( D_{qq}(n, q) > 0 \). Likewise, damage increases at an increasing rate when the number of firms goes up, that is, \( D_n(n, q) > 0 \) and \( D_{nn}(n, q) > 0 \). Finally, it is usual to assume that the increase in damage that follows an increase in firm output depends positively on the number of firms, \( D_{qn}(n, q) > 0 \). These kinds of assumptions make sure that once an optimum level of damage exists, it will be unique. Clearly, one can think of externalities where it is proper to assume that damage increases with higher production when production is low, but after some threshold level the marginal damage falls or goes to zero. For example, imagine a firm that emits chemical substances into a river and suppose that damage is measured by the number of dead fish. When the river is sufficiently affected, all of the fish are dead and further emissions are costless. Such kinds of assumptions result in nonconvexities, and a consequence of this can be the existence of multiple equilibria, or, for example, each firm should either produce the profit-maximizing quantity or not produce at all (Baumol and Bradford, 1992). Our main concern is with the standard case of increasing marginal damages.

If the inverse demand function is \( P(nq) \), the long-run social optimum is given by (see Carlton and Loury 1980 for details):

\[
(1) \quad P(nq) = C_q(q) + D_q(n, q)/n
\]

and

\[
(2) \quad qP(nq) = C(q) + D_n(n, q)
\]
With the proper second-order conditions being satisfied, the unique optimum allocation satisfies (1) and (2), and we denote this allocation by \( \{q^*, n^*\} \). Now, the social average cost, called \( \text{SAC} \), is \( (C(q) + D(n,q))/q \), and we have:

\[
\frac{d\text{SAC}}{dq} = \frac{1}{q} \left( C_q(q) + D_q(n,q) - \frac{C(q) + D(n,q)}{q} \right)
\]

Output per firm satisfies \( C_q(q_s) = C(q_s)/q \), when each firm fully utilizes all economies of scale, and evaluating how social average cost changes with production scale at \( q_s \) we have

\[
\frac{d\text{SAC}}{dq} \bigg|_{q_s} = \frac{1}{q_s} \left( D_q(n,q_s) - \frac{D(n,q_s)}{q_s} \right)
\]

When, for any number of firms, \( D_q(n,q) > 0 \) and \( D_{nq}(n,q) > 0 \), it follows that \( D_q(q) > D(n,q)/q \), and the social average cost is thus increasing in output per firm. In turn, this shows that the social optimum production scale satisfies \( q^* < q_s \). That is, each active firm should ideally produce at a point of increasing scale. Of course, the conclusion that optimum production scale satisfies \( q^* < q_s \) makes it safe to assume that firms are price takers under regulation if it is also the case that the number of firms goes up. To see that it actually does so notice that \( C_q(q^*) < C_q(q_s) \). Also, at the unregulated equilibrium, called \( \{q_s, n_s\} \) we have \( P(n,q_s) = C_q(q_s) \) while the optimum allocation is characterized by \( P(n^* q^*) = C_q(q^*) \).

It follows that \( P(n,q_s) > P(n^* q^*) \) or \( n_s q_s < n^* q^* \) when demand is downward sloping. From \( n_s q_s < n^* q^* \) it is clear that \( n^* > n_s \) since \( q^* < q_s \). That is, when each firm is small relative to aggregate production without regulation it will also be small under regulation that changes the output-per-firm-number-of-firms combination in direction of the socially optimal combination. Oppositely, if output per firm were to be increased by regulation the supposition of markets being competitive could be out of line with the aims of regulation.
since in this case regulation should decrease the number of firms and let remaining firms
increase production.

Of course, when each firm should ideally produce at a scale where the average cost
exceeds the marginal cost, the competitive firm is unable to break-even unless the unit tax
that regulates firm output to the social optimum is somehow modified. Once firms are subject
to a charge, called \( t \), on marginal output profit maximization calls for \( P(nq) = C_q(q) + t \). Let
us suppose that the tax system is nonlinear with a charge of \( t \) per unit of output that exceeds
\( S/t \). In this case, entry and exit is brought to a stop when \( qP(nq) - C(q) - t(q - S/t) = 0 \). We
can now show proposition 1.

**Proposition 1.** The long-run competitive equilibrium is coincident with the full social
optimum under a simple unit tax when firm output beyond \( S/t \) is taxed at the rate of
\( t = D_q(n^*, q^*)/n^* \) where \( S = (tq^* - D_n(n^*, q^*)) \) and is otherwise not taxed.

**Proof.**
The result follows straightforward from the analysis in Carlton and Loury (1980), but we
shall give the proof for convenience. If we use \( t = D_q(n^*, q^*)/n^* \) in equation
\[ P(nq) = C_q(q) + t \] we have \( P(nq) = C_q(q) + D_q(n^*, q^*)/n^* \). Plainly, when \( n = n^* \) and
\( q = q^* \) equations (1) is satisfied. Using the definition of \( S \) in equation
\[ qP(nq) - C(q) - t(q - S/t) = 0 \] we have \( qP(nq) - C(q) - t(q - (q^* - D_n(n^*, q^*)/t)) = 0 \), or,
when \( n = n^* \) and \( q = q^* \), \( q^* P(n^* q^*) - C(q^*) + D_n(n^*, q^*) = 0 \) showing that equation (2)
is satisfied. Finally, we have implicitly assumed that each firm produces a quantity that is
larger than \( S/t \), that is, \( S/t < q^* \). To verify that this is actually satisfied, notice that
\( S/t = q^* - D_n(n^*, q^*)/D_q(n^*, q^*) \) when \( D_q(n, q) > 0 \) and \( D_n(n, q) > 0 \) as is assumed here.

*End of proof.*
Proposition 1 demonstrates that a nonlinear Pigouvian tax provides the right incentives in the short as well as the long run when the damage function fulfills standard assumptions, when demand is downward sloping, and when firms are price takers in the unregulated market. Thus, there are no serious limitations on Pigouvian taxes as a cure against externalities as it seems to be the conclusion set forth by Carlton and Loury (1980, 1986). With respect to the practical implications of proposition 1, notice that the exemption level is always positive; this follows because each firm produces at a point of increasing scale. In turn, the tax system specifies a zero unit tax for some output and a positive unit tax for the rest of each firm’s output so that revenue must be positive. In this way, the problems of raising revenue by lump-sum taxation of other sectors that are discussed by Baumol and Bradford (1970), and later Baumol (1979) does not arise. Also, since the number of firms increases under regulation and each firm produces less the assumption of perfect competition continues to apply.

As noted, there are certain kinds of damage that are characterized by decreasing marginal damage, that is, \( D_{qq}(n,q) \) and \( D_{nn}(n,q) \) or both are negative. This implies that the solution to the problem of maximizing welfare does not in general have a unique solution. This follows immediately from the second-order conditions

\[
np'(nqu) - C_{qq}(q) - n^{-1}D_{qq}(n,q) < 0 \quad \text{and} \quad q^2 P'(nqu) - D_{nn}(n,q) < 0.
\]

Since \( P'(nqu) \) is negative, these conditions are met for any quantity-number-of-firms pair when \( D_{qq}(n,q) > 0 \) and \( D_{nn}(n,q) > 0 \). Conversely, if either \( D_{qq}(n,q) \), or \( D_{nn}(n,q) \), or both are negative, the second-order conditions might not be satisfied. Assuming however, that demand is sufficiently sloped in a strong, negative manner a unique social optimum still exist. Consider the case where marginal damage of increased output is positive and increasing while the marginal damage of adding firms is positive but at a decreasing rate. Going through the proof of
proposition 1, it can be seen that the signs of $D_m(n,q)$ and $D_{qn}(n,q)$ play no role, so the result still applies, and it is still the case that the socially optimal output per firm is less than the output that utilizes all economies of scale and that the optimum number of firms is higher than the number surviving in the unregulated market, that is, $q^* < q_s$ and $n^* < n_s$. We state this as proposition 2.

**Proposition 2.** As long as $D_q(n,q) > 0$, $D_n(n,q) > 0$, and $D_{qq}(n,q) > 0$ the nonlinear unit tax defined in proposition 1 is efficient when the problem of maximizing welfare has a unique solution.

The explanation for proposition 2 is straightforward. Whenever $D_q(n,q) > 0$ and $D_{qq}(n,q) > 0$ each firm’s production technology is convex and it is therefore possible to sustain any desirable allocation as has been shown for example by Baumol and Oates (1975, chapter 8). Also it follows from $D_{qq}(n,q) > 0$ that regulation should be aimed at reducing firm output to a point of increasing returns and increase the number of active firms. In this sense it is safe to postulate that regulated markets are competitive when markets are competitive without regulation. That is, all of the standard assumptions for price instruments to be efficient are met. The effect of assuming that $D_{nn}(n,q)$ is negative and $D_{qq}(n,q)$ is positive is that the social production-possibility set might be non-convex. This means that it is uncertain whether or not a unique solution exists, but proposition 2 makes it clear that once such a solution exists a tax is an efficient instrument.

We are left with the case where marginal damage of increased output is positive and decreasing. Again, this produces a non-convexity of the social production possibility frontier but firms can be regulated by a tax as long as private costs satisfy the standard conditions. However, as noted, social average cost changes according to $dSAC/dq|_{q_s} = (D_q(n,q_s) - D(n,q_s)/q_s)/q_s$ when each firm produces at the point where all
economies of scale are exhausted. In turn, the slope of the social average cost curve at this point is negatively sloped when $D_q(n,q) > 0$ and $D_{qq}(n,q) < 0$. This means that output per firm should be pushed into a region of decreasing returns to scale and that the number of firms should ideally decrease, that is, $q^* > q_s$ and $n^* < n_s$. In relation to assuming that markets are perfectly competitive, this is clearly a problem. In models of perfect competition, profits are nullified by the entry and exit of firms until each firm produces at the point where there are constant returns to scale. Assuming that firms are price takers is to assume that each firm produces a negligible amount of total production when the firm utilizes all economies of scale. In the case of $D_{qq}(n,q) > 0$, firm production is adjusted downwards, and we are therefore certain that firm production under regulation is small relative to the aggregate output. This is not so when $D_{qq}(n,q) < 0$; here each firm should grow and there are ideally to be fewer firms meaning that each firm produces at a level that might be non-negligible so that the assumption of price-taking behavior is questioned. It is easy to see this point. When the marginal damage of increasing output from a firm is falling strongly, it is (relatively speaking) beneficial to have few large firms and, in such cases, the scope of unit taxes as devices against externalities can be limited because the number of firms makes it implausible to assume that each firm is a price taker.

3 Conclusion

In this paper, we have revisited the question of how efficient Pigouvian taxes are in the long run. We emphasize, once more, that the results including the existence of a unit tax with an exemption level are mentioned in passing by Carlton and Loury (1980, 564). However, the point that a simple nonlinear unit tax is efficient and does not require the regulator to raise funds from other sectors when we make standard assumptions about damage is missing. Thus, for many practical policy issues, the analysis suggests that Pigouvian taxes correct
externalities efficiently. Clearly, the analysis does not suggest that unit taxes are efficient in the presence of severe non-convexities and non-uniqueness problems. Also, there is one further case in which the working of Pigouvian taxes is subtle. This is when the marginal damage of increases in production is positive but falling. In this case, a unit tax with an exemption level would not work. However, we suggest that one of the main problems in such circumstances is whether one can think of a perfectly competitive industry at all—the reason being that each firm should ideally produce at a point of decreasing returns. The natural or necessary assumption to make in models that are based on competitive product markets is that each firm is small when the firm produces at the point of constant returns to scale. But, once it is ideal to have fewer firms produce more, the idea of each firm producing a negligible amount of aggregate output is not naturally satisfied, and this of course questions the general applicability of a unit tax which relies on firms being price takers.

References


---

2 It is easy to demonstrate, in this case, that each firm should be given a subsidy according to $s(S/s-q)$ where $s = D_s(n^*, q^*)/n^*$ and $S = (sq^*-D_s(n^*, q^*))$. But, since entry and exit nullifies profit, it is better for firms not to produce and collect the subsidy $S$. 

10