Environmental Taxes in the Long Run

Henrik Vetter

Abstract

The efficiency of the Pigouvian tax suggests that price-based regulation is the proper benchmark for efficient regulation. However, results due to Carlton and Loury (1980, 1986) question this; when harm depends on scale effects a pure Pigou tax is inefficient regulation in the long run. In this note we make precise that there is an efficient tax scheme for controlling harm as long as social optimum exists. In particular, the efficient tax scheme is based on a tax rate equal to marginal harm. Hence, price regulation is the right benchmark for regulation even in the presence of scale effects in the harm function.

Authors

Henrik Vetter, ✉ State and University Library, Aarhus, hv@statsbiblioteket.dk

This work is not the result of a for-pay consulting relationship; neither does any party have a financial interest in the results.

Citation Henrik Vetter (2013). Environmental Taxes in the Long Run. Economics Discussion Papers, No 2013-29, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2013-29

Introduction

Pigouvian taxes are central to understanding environmental policy because they set a standard for efficient regulation. They are the basis for evaluation of environmental taxes under monopoly, (Barnett 1980) and other forms of imperfect competition (Xepapadeas, 1997), as well as the measuring point for second best taxes (Pang and Shaw, 2011). Also, the efficiency of Pigouvian taxes is the starting point of comparisons of quotas and price regulation (Weitzman, 1974, and recently Kato, 2011). In the short run a tax with a tax rate equal to marginal harm is efficient regulation. However, there are dissenting views on the efficiency of the Pigouvian tax with respect to the long run effects of regulation (Carlton and Loury, 1980 and 1986, Pezzey, 2003).

With respect to regulation in the long run efficient regulation might imply a policy that changes output per firm in addition to achieving optimal exit or entry. The conventional view is that the number of polluting firms is optimal under the Pigouvian tax (Pezzey, 2003, page 329). Of course, under a charge per unit, each firm produces until the point where economies of scale are fully utilised. Thus, the efficiency of a flat rate tax presupposes that the socially optimal production per firm minimises the firm's average cost. Clearly, the upshot of this is that the tax is optimal only when harm is independent of scale. In general, when scale effects matter for environmental harm a pure Pigou tax cannot lead to long run social optimum (Carlton and Loury, 1980, 1986).

The purpose of this paper is to characterize the relationship between the externality function and the optimum scale of production and discuss the optimal long run tax. We show, when each firm ideally produce under increasing returns to scale, that tax threshold and a tax equal

to marginal harm at the optimum is efficient regulation. Oppositely, when firms ideally produce under decreasing returns a combination of a flat-rate tax and a fee is efficient regulation. Moreover, the efficient scheme is in fact no more impractical compared to the conventional flat rate Pigou-tax. In this way a marginal tax equal to marginal harm is efficient regulation in a qualified way also in the long run.[1] Hence, a tax rate equal to marginal harm continues as the proper benchmark for environmental regulation in the presence of scale dependent harm. For example, it is straightforward that a combination of an entry fee and a tax compares to a combination of an entry fee and a quota. Parallel, the combination of a tax threshold and a tax compares to a combination of tradeable permits where emissions are required only for production in excess of some production threshold.[2]

Finally, we discuss how existence of social optimum is a prerequisite for a tax scheme when marginal harm is decreasing. Most likely, more serious limitations to tax schedules based on charges equal to marginal harm (either on their own or combined with some tax threshold or a fee for being in the market) arises when marginal harm is decreasing in output per firm and the number of firms. As a general rule, the second-order conditions for a social-optimum allocation fail when the marginal damage of per-firm output is a decreasing function and prices cannot guide the economy to the proper maximum, even if a complete set of markets for externalities should happen to exist. This obviously hinders any kind of tax scheme to bring about an efficient allocation. However, a unique social optimum still exists if demand is sufficiently strong and negatively sloped. In this situation, as shown by Baumol and Bradford (1972), it is possible to regulate successfully whenever the polluter's profit function is well-behaved in spite of problems relating to non-convexities of the damage function. It turns out

---

[1] Carlton and Loury (1980) notice that a tax threshold or a fee in combination with a tax is efficient regulation whenever harm is a non-decreasing function of output. They do not relate the result to the exact characteristics of the harm function.

[2] Unless permits are priced the long run effects of a permit threshold and permits would of course differ from the effects of a combination of a tax threshold and a tax (see Pezzey, 2003).

in this case, that each active firm should ideally produce at a point of decreasing returns. We explain why this calls for a combination of an entry fee and a unit tax equal to marginal harm.

1. Results

We consider a competitive industry where a number of firms produce a homogenous product. The size of the industry is regulated by entry and exit until the profit of each active firm is zero. Each firm's private production cost is $C(q)$ when q units are produced. Average production cost per firm, $C(q)/q$, is assumed to be U-shaped, and average costs are minimized when the firm produces q_s. Production within the industry generates an unpriced harm on society. We assume that damage is a function of firm output and the number of firms, called n, according to $D(n,q)$. Conventionally, damage increases at an increasing rate as per-firm output goes up. Likewise, damage increases at an increasing rate when the number of firms goes up. Also, it is usual to assume that the increase in damage that follows an increase in per-firm output depends positively on the number of firms and vice versa. We state this as Assumption 1.

A1: The harm function satisfies $D_q(n,q) > 0$, $D_{qq}(n,q) > 0$, $D_n(n,q) > 0$, $D_{nn}(n,q) > 0$, and $D_{qn}(n,q) > 0$.

Clearly, one can think of externalities where it is more appropriate to assume that damages increase with higher production when production is low, but after some threshold level, the marginal damage is falling or goes to zero. This is described by Assumption 2.

A2: The harm function satisfies $D_q(n,q) > 0$, $D_{qq}(n,q) < 0$, $D_n(n,q) > 0$, and $D_{nn}(n,q) < 0$ without restrictions on the sign of $D_{qn}(n,q)$.

The inverse demand function is $P(nq)$, where, $P'(nq) < 0$. Hence, the long-run social optimum is characterized by (see Carlton and Loury, 1980, for details):

(1) $$P(nq) = C_q(q) + \frac{1}{n} D_q(n,q)$$

and

(2) $$qP(nq) = C(q) + D_n(n,q).$$

When Assumption 1 is satisfied and private production costs and demand functions satisfy standard assumptions, the optimum values of output per firm and the number of firms are unique. The unique optimum allocation that satisfies (1) and (2) is called $\{q^*, n^*\}$.

With scale effects in the damage function, it is clear that the socially optimum per-firm output is different from the level of output that minimizes the firm's private production cost. To find out whether an active firm ideally produces with increasing or decreasing returns to scale, notice that the social average cost, called SAC, is $(C(q) + D(n,q))/q$, and, moreover:

(3) $$\frac{dSAC}{dq} = \frac{1}{q}\left(C_q(q) + D_q(n,q) - \frac{C(q) + D(n,q)}{q}\right)$$

If each firm fully utilizes all economies of scale, per-firm output satisfies $C_q(q_s) = C(q_s)/q_s$, and evaluating how social average cost changes with production scale at q_s we have:

(4) $$\left.\frac{dSAC}{dq}\right|_{q_s} = \frac{1}{q_s}\left(D_q(n,q_s) - \frac{D(n,q_s)}{q_s}\right).$$

Because $D_q(q) > D(n,q)/q$ under Assumption 1, the socially optimal production scale satisfies $q* < q_s$. If we denote by $\hat{q}$ and $\hat{n}$ output per firm and the number of firms, respectively, in long run competitive equilibrium without regulation we have Proposition 1.

Proposition 1. Suppose that the harm function satisfies Assumption 1. Output per firm in the long-run competitive equilibrium, $\hat{q}$, exceeds the socially optimal output per firm, $q*$. The number of firms in the long-run competitive equilibrium, $\hat{n}$, can exceed or fall short of the optimal number of firms, $n*$.

The first claim follows immediately by noticing that $\hat{q} = q_s$ because each firm fully exhausts economies of scale in long run equilibrium. To see the second claim notice that— when the socially optimal output per firm is different from the output that exactly exhausts economies of scale the average cost—it follows immediately form equation (2) that the socially optimal price exceeds the long run price in the absence of regulation. Since demand is inversely related to total output we have $\hat{q}\hat{n} > q*n*$. The only restriction that derives from this is $n* < \hat{n}\hat{q}/q*$. Thus, efficient regulation can actually involve entry of firms relative to the unregulated long run equilibrium. If firms ideally produce so that they exactly take advantage of economies of scale it follows that efficient regulation involves exit of firms.

Proposition 1 shows that an active firm should ideally produce at a point of increasing scale when Assumption 1 is met. In turn, the price falls short of the firm's average cost at the social optimum. Thus, for the firm to break even under a socially optimal allocation, a corrective unit tax equal to marginal harm must be combined with some other instrument that makes up for the difference between price and average cost. The efficient tax scheme is a combination

of a charge per unit of output, called t, and a tax threshold, which means that only output in excess of S/t is taxed. We show Proposition 1 in the appendix.

Proposition 2. Suppose that the harm function satisfies Assumption 1. The long-run competitive equilibrium is coincident with the full social optimum under a simple unit tax equal to marginal harm when firms are exempted from the tax for output below a certain threshold. The tax threshold is S/t, the tax rate is $t = D_q\left(n^*, q^*\right)/n^*$, and $S = \left(tq^* - D_n\left(n^*, q^*\right)\right)$.

Of course, there are types of externalities that are not described by Assumptions 1 and 2, for example externalities where the harm function satisfies Assumption 1 except for the sign of $D_{nn}\left(n, q\right)$ and $D_{qn}\left(n, q\right)$. Going through the proof of Proposition 2, it can be seen that the signs of these derivatives play no role, so that the result still applies. We state this as Proposition 3.

Proposition 3. The tax scheme in Proposition 2 is efficient when Assumption 1 is relaxed with respect to the signs of $D_{nn}\left(n, q\right)$ and $D_{qn}\left(n, q\right)$ as long as a unique social optimum exists.

Together propositions 2 and 3 show that a tax equal to marginal harm per firm measured at the social optimum in combination with a tax threshold provides the right incentives for competitive firms in the short as well as in the long run as long as the marginal harm of per firm output is increasing in firm output. Because the efficient tax rate equals marginal harm of increased output per firm the firm's net of tax profit function is not dependent on the

relationship between harm and the number of firms. This explains why the tax scheme described in Proposition 2 can be used under the conditions listed in Proposition 3. However, the way the number of firms changes harm might imply that we cannot find a unique social optimum. Hence it is necessary to assume that a social optimum exists. In Proposition 2 this assumption is unnecessary since the harm function ensures that second order conditions are satisfied.

The explanation for the need for a subsidy is straightforward. Regulation should be aimed at reducing per-firm output to a point of increasing returns whenever marginal harm of per firm output exceeds average harm, cf. equation (4). This means that the firm's profit is negative unless there is some subsidy. The subsidy can be implemented by exempting from the tax some part of the firm's output.

Of course, it can be argued that it is, in practice, difficult to calculate the correct tax rate as well as the tax threshold. But notice that the information needed to construct the efficient tax schedule in the presence of a scale effect in the damage function is the same as the information needed for finding the efficient flat rate Pigou-tax: One must know consumers' willingness to pay for a marginal increase in output over the relevant range. That is, the demand function $P(nq)$ must be known in the range that includes the optimum. Next, calculation of the correct tax value calls for knowledge about firms' cost function across the relevant range of output. Finally one needs to know the value of marginal harm around the optimum which is secured, if there is knowledge about the harm function over the relevant range. But these are also the informational requirements that are supposed to be met for

Proposition 2 to apply. [3] Knowing optimum per-firm output, and knowing the willingness to pay for additional output at the optimum, the optimum number of firms are known through the demand function. This allows calculation of the optimum tax rate according to $t = D_q\left(n^*, q^*\right)/n^*$. The optimum threshold follows from $S = \left(tq^* - D_n\left(n^*, q^*\right)\right)$ which requires knowledge of the harm generated by one additional firm measured at the optimum in addition to the information that allow calculation of the optimum tax rate.

Moreover, with respect to practical policy, notice that the exemption level is always positive; this follows because each firm produces at a point of increasing scale. We have $S = q^*\left(C(q^*)/q^* - C_q\left(q^*\right)\right) > 0$ from equations (5) and (6), and overall revenue is positive as $t\left(q^* - S/t\right) = D_n\left(n^*, q^*\right)$ from equations (2) and (6). In this way, there will be no problem with raising revenue by the lump-sum taxation of other sectors. Thus, in the standard case, the efficiency of a tax that closes the gap between the private and the social marginal cost is not seriously harmed.

We are left with the case covered by Assumption 2. It is well known that, in these circumstances, the solution to the problem of maximizing welfare can fail to have a unique solution. As shown by Baumol and Bradford (1972) the consequence can be the existence of multiple equilibria or, that the polluting firm should either produce the profit-maximizing quantity or not produce at all. Assuming, however, that demand is sufficiently strongly negatively sloped, the conditions will be satisfied and a unique social optimum will exist. We state this as Assumption 3.

---

[3] It can be argued there is a difference since one needs to know $D_n\left(.\right)$ to find the policy described in Proposition 1, and this term does not enter into the traditional Pigou-tax. However, to be sure about the efficiency properties of the flat-rate Pigou-tax, one needs to know the damage function to rule that scale effects matter.

A3. Demand is sufficiently strongly negatively sloped to ensure existence of a social

optimum under Assumption 2.

When the harm function satisfies Assumption 2 the socially optimal output per firm occurs

where the firm produces under decreasing returns to scale. To see this notice that

$dSAC/dq|_{q_s} = \left(D_q(n,q_s) - D(n,q_s)/q_s\right)/q_s$ when each firm produces at the point where all

economies of scale are exhausted, that. The slope of the social average-cost curve at this

point is negative when $D_q(n,q) > 0$ and $D_{qq}(n,q) < 0$. It follows that $q^* > q_s$. In comparison

to the long run equilibrium in the absence of regulation we have Proposition 4.

Proposition 4. Suppose that the harm function satisfies Assumption 2. Output per firm in the

long-run competitive equilibrium, $\hat{q}$, falls short the socially optimal output per firm, q^*. The

number of firms in the long-run competitive equilibrium, $\hat{n}$, exceeds the optimal number of

firms, n^*.

The first claim follows as before by noticing that $\hat{q} = q_s$. Because $q^* > q_s$ the marginal cost

increases and since marginal harm of increased output per firm is positive it follows from

equation (1) that the price in the optimal allocation exceeds the price that obtains in long run

unregulated equilibrium. Thus, aggregate output goes down under efficient regulation. This

gives $q^*n^* < \hat{q}\hat{n}$. It then follows immediately that $n^* < \hat{n}$. Consider the workings of a unit tax

on firm output, called s, combined with the payment of a fee of the order of E for a permit to

be in the market at all. Profit maximization results in $P(nq) = C_q(q) + s$, and long-run

equilibrium occurs when $qP(nq) - C(q) - sq - E = 0$. Now, combining these conditions with

the first-order condition for a social optimum (equations (1) and (2)), we have $s = D_q(n*, q*)/n*$ and $E = D_n(n*, q*) - D_q(n*, q*)q*/n*$. It is obvious that the tax rate is positive. Another way to look at the entry fee is to use the zero-profit condition that allows us to write $E = q*(C_q(q*) - C(q*)/q*)$, which is positive because marginal costs exceed average cost, as production is at a point of decreasing returns to scale. We state this as Proposition 5.

Proposition 5. If a unique social optimum exists and Assumptions 2 and 3 apply a combination of an entry fee and an output tax equal to marginal harm at the social optimum is efficient.

Using equation (4) it is easy to see why a fee is necessary when $D_q(q) < D(n, q)/q$. In this case firms ideally produce under decreasing returns to scale meaning that profit per firm is strictly positive. This clearly is inconsistent with long-run equilibrium and the solution is to introduce a fee. The working of the regulatory scheme described in the proposition is that the entry fee restricts the number of firms. If the entry fee is the only instrument per firm output would be inefficient. In order to harness each firm's production to the socially desirable level, the entry fee, which drives up the price, must be combined with a tax. The overall revenue is clearly positive showing that the problems relating to the overall economy budget restriction, as pointed out by Baumol and Bradford (1970) and Baumol (1979), are absent.

2. Conclusion

The efficiency of the Pigouvian tax strongly suggests that price-based regulation is the proper benchmark for efficient regulation. However, at first glance the results of the papers by Carlton and Loury (1980 and 1986) question the long run efficiency of a Pigouvian tax. In

this note we have made precise how a slight modification of the Pigouvian tax is in fact efficient regulation.[4]

When the socially optimal scale of production differs from the scale minimizing the private production cost regulation cannot take advantage of private firms' attempts to minimize production costs. Of course, this restricts the efficiency of a flat-rate tax. But following Kaplow and Shavell (2002) and, initially Roberts and Spence (1976), one can expect the efficiency of the competitive process to be secured by a flat-rate tax schedule only in the case where the regulation problem is linear. When, in real situations, a regulation problem is nonlinear, for example when harm is affected by changes in the number of firms, it is of interest to ask whether corrective taxes are reasonably practical and simultaneously an efficient way to deal with external damages. In this note, allowing for scale effects in the harm function, we make precise that 1) there is an efficient tax scheme whenever a social optimum exits, 2) the tax scheme is a flat rate tax in combination with either a fixed fee or a tax exemption, 3) the regulator is not constrained by informational limitations, at least compared to using a simple flat rate Pigou tax, and 4) net revenue is positive making the scheme is feasible in the meaning that direct payments to firms are unnecessary. Moreover, the suggested regulation is no more impractical than a simple flat-rate Pigou tax. In this way our conclusion is that price regulation is not limited efficiency regulation even when damages are sensitive to scale effects.

---

[4] As noted, Carlton and Loury (1980 page 565) speaks about a scheme that meets a budget restriction. However, they do not discuss taxes in relation to problems of existence of a socially optimal level of production. Neither is it clear how the second order derivates of the harm function determines whether the flat-rate tax is to be supplemented with either tax exemption rules or a fee. Proposition 3 is not covered by exiting analysis.

References

Baumol, W.J. 1979. Quasi Optimality: The Price We Must Pay for a Price System. *The Journal of Political Economy* 87: 578–599.
http://ideas.repec.org/a/ucp/jpolec/v87y1979i3p578-99.html

Baumol, W.J. and D.F. Bradford. 1970. Optimal Departures from Marginal Cost Pricing. *American Economic Review* 60: 265–283.
http://ideas.repec.org/a/aea/aecrev/v60y1970i3p265-83.html

———. 1972. Detrimental Externalities and Non-Convexity of the Production Set. *Economica*, N.S., 39 (154): 160–176.
http://ideas.repec.org/a/bla/econom/v39y1972i154p160-76.html

Barnett, A.H. 1980. The Pigouvian Tax Rule under Monopoly. *American Economic Review*, 70: 1037–1041.
http://ideas.repec.org/a/aea/aecrev/v70y1980i5p1037-41.html

Carlton, D.W. and G.C. Loury. 1980. The Limitation of Pigouvian Taxes as a Long-Run Remedy for Externalities. *The Quarterly Journal of Economics* 95:559–566.
http://www.econ.brown.edu/fac/glenn_loury/louryhomepage/papers/carleton-loury.pdf

———. 1986. The Limitation of Pigouvian Taxes as a Long-Run Remedy for Externalities: An Extension of Results. *The Quarterly Journal of Economics* 101:631–634.
http://ideas.repec.org/a/tpr/qjecon/v95y1980i3p559-66.html

Kaplow, L. and S. Shavell. 2002. On the Superiority of Corrective Taxes to Quantity Regulation. *American Law and Economics Review* 4:1–17.
http://www.law.harvard.edu/faculty/shavell/pdf/4_Amer_Law_Econ_Rev_1.pdf

Kato, K. 2011. Emission Quota versus Emission Tax in Mixed Duopoly. *Environmental Economics and Policy Studies* 13:43–63.
http://ideas.repec.org/a/spr/envpol/v13y2011i1p43-63.html

Pang, A. and D. Shaw. 2011. Optimal Emission Tax with Pre-existing Distortions, *Environmental Economics and Policy Studies* 13:79–88.
http://ideas.repec.org/a/spr/envpol/v13y2011i2p79-88.html

Pezzey, J.V.C. 2003. Emission Taxes and Tradeable Permits, *Environmental and Resource Economics* 26:329–342.

Pigou, A.C. [1920] 1932. *The Economics of Welfare*, 4th ed. London: Macmillan.

Roberts, M.J. and M. Spence. 1976. Effluent Charges and Licenses under Uncertainty. *Journal of Public Economics* 5:193–208.
http://ideas.repec.org/a/eee/pubeco/v5y1976i3-4p193-208.html

Weitzman, M. 1974. Prices versus Quantities. *Review if Economic St*udies 41:477-491.
http://ideas.repec.org/a/bla/restud/v41y1974i4p477-91.html

Xepapadeas, A. 1997. *Advanced Principles in Environmental Policy*. Edward Elgar, Cheltenham.

Proof of proposition 1.

The result follows straightforward from the analysis in Carlton and Loury (1980), but we provide the proof for convenience. Once the firm has to pay a charge per unit of output, called t, in combination with a tax threshold, which means that only output in excess of S/t is taxed, profit maximization implies:

(A.1) $\qquad P(nq) = C_q(q) + t$,

and entry and exit are brought to a stop when:

(A.2) $\qquad qP(nq) - C(q) - t(q - S/t) = 0$.

If we use $t = D_q(n^*, q^*)/n^*$ in equation (A.1), we have

(A.3) $\qquad P(nq) = C_q(q) + D_q(n^*, q^*)/n^*$.

Plainly, when $n = n^*$ and $q = q^*$, equations (A.3) and (1) are coincident. Using the definition of S in equation (A.2) we have:

(A.4) $\qquad qP(nq) - C(q) - t\left(q - \left(q^* - D_n(n^*, q^*)/t\right)\right) = 0$,

or, when $n = n^*$ and $q = q^*$,

(A.5) $\qquad q^* P(n^* q^*) - C(q^*) - D_n(n^*, q^*) = 0$.

This shows that equations (A.2) and (2) are coincident when $n = n^*$ and $q = q^*$.

End of proof.

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

http://www.economics-ejournal.org/economics/discussionpapers/2013-29

The Editor