

EFFICIENT DETERRENCE AND THE TAX TREATMENT OF MONETARY SANCTIONS

I. P. L. PNG* AND ERIC M. ZOLT

University of California, Los Angeles, CA 90024, USA

I. INTRODUCTION

Many business activities generate external harm. Oil refineries generate air pollution, overloaded trucks damage highways, and messenger services obstruct traffic as they double park. To control these harms, the government subjects offenders to monetary as well as nonmonetary sanctions. In this article, we examine how the government should adjust the amount or the tax treatment of monetary sanctions to achieve efficient deterrence of external harm by firms that are subject to income taxation.

We believe that the Becker (1968) formula—that the amount of sanctions be based on the external harm caused and the probability of enforcement—must be adjusted in a world in which firms pay income taxes. Most firms that engage in activities generating external harm are subject to federal and state income taxation. We contend that failure to consider tax consequences in the design of monetary sanctions distorts firms' choice of inputs, firms' scale of production, and total industry output. The government can achieve efficient deterrence only by either changing the tax rule to allow deductibility of monetary sanctions or adjusting the amount of such sanctions to reflect the offender's marginal tax rate.

The tax law does not treat monetary sanctions in a coherent fashion. The U.S. tax system, like most income tax systems, taxes a firm on its net income and, therefore, generally allows deductions for all ordinary and necessary expenses paid or incurred in carrying on a trade or business.¹ The Internal Revenue Service and taxpayers have disagreed about whether amounts paid as monetary sanctions, usually fines or penalties imposed by federal and state agencies, qualify as ordinary and necessary business expenses. Courts faced two competing considerations: the pressure to measure net income accurately, regardless of moral and legal considerations, and a notion that to allow such deductions would frustrate public policy.

In the major decision examining the deductibility of fines and penalties, the U.S. Supreme Court denied the deduction of amounts paid as fines by a trucking concern for violation of a state's maximum highway weight laws.² It held that allowing such deduction would reduce the odds of compliance and would frustrate sharply defined state policy. The Tax Reform Act of 1969 codified this decision

Address for correspondence: I. P. L. Png, John E. Anderson Graduate School of Management, University of California, Los Angeles, CA 90024-1481.

This paper was the basis of testimony on organizational sanctions before the United States Sentencing Commission on December 2, 1988. We thank M. Asimow, D. Hirshleifer, S. Munzer, M. Polinsky, M. Ramseyer, E. Rasmusen, D. Rubinfeld, G. Schwartz, S. Shavell, and S. Yeazell for their generous advice.

¹Internal Revenue Code, §162(a)(1982) and predecessor, §23(a)(1)(A)(1939).

²*Tank Truck Rentals, Inc. v. Commissioner*, 356 U.S. 30 (1958).

as Section 162(f), which provides “[n]o deduction shall be allowed . . . for any fine or similar penalty paid to a government for the violation of any law.”³ By disallowing these deductions, the courts and Congress created an exception to the broad principle that the federal income tax was meant as a tax on net income and not as a sanction against wrongdoing.⁴

The courts and Congress were correct that disallowing deductions will, at the margin, increase the odds of compliance. Disallowing deductions increases the burden of the sanction by the amount of the additional tax liability. We contend, however, that the courts and Congress failed to focus on the issue of how much compliance would be desirable.

Not all amounts paid as monetary sanctions are subject to disallowance. Deductions are generally allowed for most damages paid, even as a result of fraud or breach of fiduciary duties, for fines for offenses that are not generally enforced, and for those penalties that are compensatory. Failure to recognize that monetary sanctions are often treated differently for tax purposes prevents the government from designing sanctions to achieve efficient deterrence.

II. MODEL

In this section, we model the socially efficient outcome in an industry where one of the inputs into production generates external harm. We adopt the partial equilibrium approach taken by Polinsky (1980) in his comparison of the tort rules of strict liability and negligence in a setting of competitive firms.

Suppose that there are n identical firms engaged in some business that requires inputs of labor and some other input, for example, air pollution, that generates external harm. Each firm has the same technological capabilities, and can produce $q(l, a)$ units of output with inputs l of labor and a of air pollution. Let the production technologies exhibit diminishing marginal returns and let the marginal product of labor be strictly greater the higher the amount of air pollution, or formally, let $q(l, a)$ be increasing and strictly concave in both arguments and $q_{la} > 0$. The supply of the labor is infinitely elastic at wage w . For simplicity, we assume that the private cost of the pollution input is zero and that each firm is risk neutral.

For a given demand for the product of the industry, let the price at which a quantity Q will be demanded by the market be $p(Q)$. As n identical firms each produce a quantity q , the total quantity produced will be $Q = nq$ and the price will be $p(nq)$. Assume that there are no income effects, so that consumer surplus may be measured by the area between the demand curve and the market price. Let the external harm from the pollution be $H(a)$, and the marginal harm be increasing in the amount of air pollution, or formally, let $H(\bullet)$ be increasing and strictly convex.

In this setting, one can determine the efficient amount of labor, l^* , amount of pollution, a^* , and number of firms, n^* , by maximizing

$$W = \int_0^{nq} p(s) ds - n [wl + H(a)]. \quad (1)$$

³Public Law No. 91-172, §902(a), 83 Stat. 487 (1969).

⁴See Comment, “Business Expenses, Disallowance, and Public Policy: Some Problems of Sanctioning with the Internal Revenue Code,” *Yale Law Journal*, 72, 108 (1972), for a description of the legislative history of the early revenue acts.

The efficient input of labor by each firm is a quantity such that the benefit of the marginal product of labor equals the wage:

$$p(nq) \frac{\delta q}{\delta l} = w. \quad (2)$$

Second, the efficient input of air pollution by each firm is a level such that the benefit of the marginal product of air pollution equals the marginal harm caused by air pollution:

$$p(nq) \frac{\delta q}{\delta a} = \frac{\delta H}{\delta a}. \quad (3)$$

Finally, the efficient number of firms is characterized by the condition that the benefit yielded by the production of the marginal firm be equal to the social cost in terms of wages and external harm from air pollution:

$$p(nq) = \frac{wl + H(a)}{q}. \quad (4)$$

Production in the industry is socially efficient if the labor input, air pollution, and number of firms satisfy conditions (2), (3), and (4). Dividing (3) by (2), we note that social efficiency requires that the marginal rate of technical substitution between air pollution and labor be equal to their relative marginal social costs:

$$\frac{\delta q}{\delta a} \bigg/ \frac{\delta q}{\delta l} = \frac{\delta H}{\delta a} \bigg/ w. \quad (5)$$

This provides a convenient benchmark by which we may compare the competitive equilibrium of the industry under alternative tax rules and schemes of fines.

The socially efficient output by each firm is $q(l^*, a^*)$. The socially efficient number of firms in the industry is⁵

$$n^* = \frac{1}{q(l^*, a^*)} p^{-1} \left[\frac{wl^* + H(a^*)}{q(l^*, a^*)} \right]. \quad (4')$$

We assume that the costs of private negotiation and enforcement are prohibitive, hence the government must intervene to achieve the socially efficient outcome. By the very nature of harmful activity, the government may have difficulty observing each firm's air pollution without enforcement effort. The government can determine a firm's level of air pollution by inspection. Since enforcement is costly, it will be optimal to do so randomly.⁶ In general, the probability of enforcement, e , may be a function of the level of air pollution by all firms, but for simplicity, we assume that e is a constant. We assume that fines are the only sanctions available to the government. The fine $F(a)$ may vary with the level of air pollution emitted by the offending firm.

We depart from the preceding literature on efficient deterrence by supposing that the government taxes the income of each firm. Our aim is to determine how

⁵We ignore the difficulty that the solution to (4') may not be an integer.

⁶See Becker (1968) and Stigler (1970).

the government may achieve efficient deterrence in this setting. We assume that the government taxes the income of each firm at a proportionate rate t and allows firms to deduct their expenditure on labor.

III. COMPETITIVE EQUILIBRIUM: FINES NOT DEDUCTIBLE

We first consider the competitive equilibrium of the industry in which fines for air pollution are not deductible. As a firm can deduct its expenditure on labor, the taxable income of the firm will be its revenue less expenditure on labor. Hence, the firm's expected after-tax income will be the difference between (1) a factor $(1 - t)$ multiplied by its taxable income and (2) the firm's expected fines.

We suppose that the firm chooses labor, l , and air pollution, a , to maximize its expected after-tax income:

$$\Pi^{ND} = (1 - t) [pq - wl] - e F(a). \quad (6)$$

The firm maximizes expected after-tax income by choosing an input of labor such that the marginal revenue product of labor equals the wage:

$$p \frac{\delta q}{\delta l} = w, \quad (7)$$

and a level of air pollution such that the after-tax marginal revenue product of air pollution equals the marginal expected fine:

$$(1 - t)p \frac{\delta q}{\delta a} = e \frac{\delta F}{\delta a}. \quad (8)$$

The industry as a whole will be in equilibrium if the expected after-tax income of each firm is zero, and hence no firm has incentive to enter or leave the industry:

$$\Pi^{ND} = (1 - t) [pq - wl] - e F(a) = 0,$$

or

$$p(nq) = \frac{wl}{q} + \frac{e}{1 - t} \frac{F(a)}{q}. \quad (9)$$

We first show the effect of setting fines according to a simple Becker formula without adjustment for taxes. Suppose that the government sets fines according to

$$F(a) = \frac{1}{e} H(a). \quad (10)$$

Substituting this rule in conditions (7), (8), and (9), we find that the labor and air pollution of each firm and the size of the industry are characterized by

$$p \frac{\delta q}{\delta l} = w, \quad (7)$$

$$p \frac{\delta q}{\delta a} = \frac{1}{1-t} \frac{\delta H}{\delta a}, \quad (11)$$

and

$$p(nq) = \frac{wl}{q} + \frac{H(a)}{(1-t)q}. \quad (12)$$

Dividing (11) by (7), the competitive firm maximizes expected after-tax income by choosing an input of labor and a level of air pollution such that the marginal rate of technical substitution is expressed by:

$$\frac{\delta q}{\delta a} \bigg/ \frac{\delta q}{\delta l} = \frac{1}{1-t} \frac{\delta H}{\delta a} \bigg/ w.$$

Comparing this with the condition for efficient use of inputs, (5), we observe that the expected after-tax marginal cost of air pollution relative to labor is higher than their relative marginal social costs.

In particular, comparison of (11) with (3) reveals one basic distortion: the expected after-tax marginal cost of air pollution to the competitive firm is higher than the marginal social cost by a factor $1/(1-t)$. Consequently, the competitive firm will maximize expected after-tax income by choosing a level of air pollution below the socially efficient level.

Furthermore, by comparing (12) with (4), we find the second basic distortion: the expected after-tax average cost of air pollution to the firm is higher than the average social cost by a factor $1/(1-t)$. Hence, the average cost of production and the market price in the industry equilibrium will be higher than the socially efficient level.

The government can correct these distortions by reducing the expected after-tax marginal and average cost of air pollution to the firm. It can achieve socially efficient use of inputs and production in the industry by *reducing* the fine for air pollution relative to the harm caused by a factor reflecting the offender's tax rate. The specific formula is:⁷

$$F(a) = \frac{1-t}{e} H(a). \quad (13)$$

Setting the fine according to (13) ensures that the expected after-tax marginal cost of air pollution relative to labor equals the relative social marginal costs of the two inputs. The higher the tax rate, the more attractive the deductible input becomes relative to the input of air pollution, hence the more the government

⁷In the case of complete enforcement, $e = 1$, and this rule simplifies to $F(a) = (1-t)H(a)$.

should reduce the fine. If the government sets fines in this manner, it may retain the tax rule of no deductions for amounts paid as fines and still achieve a socially efficient outcome.

Substituting the formula (13) in (7), (8), and (9), we see that the competitive equilibrium of the industry is defined by:

$$p \frac{\delta q}{\delta l} = w,$$

$$p \frac{\delta q}{\delta a} = \frac{\delta H}{\delta a},$$

and

$$p(nq) = \frac{wl + H(a)}{q}.$$

These conditions are identical to the conditions for social efficiency—(2), (3), and (4). Thus, if the government sets fines according to (13), then the equilibrium of the industry will be socially efficient.

IV. COMPETITIVE EQUILIBRIUM: FINES DEDUCTIBLE

The government could alternatively allow fines to be deductible. Under this regime, both inputs receive the benefit of a tax deduction and a firm's expected taxable income will be revenue less expenditure on labor and less expected fines paid. So the firm's expected after-tax income will be a factor $(1 - t)$ multiplied by the firm's expected taxable income. Again, we assume that the firm chooses labor, l , and air pollution, a , to maximize expected after-tax income:

$$\Pi^D = (1 - t) [pq - wl - e F(a)].$$

Under this approach, expenditures on both inputs are deductible. So if the government sets sanctions so that the expected sanction equals the harm caused, the relative marginal cost of the inputs on an after-tax basis will be identical to the relative social marginal cost. Formally, let fines be set according to

$$F(a) = \frac{1}{e} H(a). \quad (10)$$

Then the firm's expected after-tax income becomes

$$\Pi^D = (1 - t) [pq - wl - H(a)]. \quad (14)$$

The firm maximizes expected after-tax income by choosing labor such that the marginal revenue product of labor equals the wage:

$$p \frac{\delta q}{\delta l} = w, \quad (15)$$

and air pollution such that the marginal revenue product of air pollution equals the marginal expected fine:

$$p \frac{\delta q}{\delta a} = \frac{\delta H}{\delta a} . \quad (16)$$

No firm will enter or leave the industry if the expected after-tax income of each firm is zero:

$$\Pi^D = (1 - t) [pq - wl - H(a)] = 0,$$

or

$$p(nq) = \frac{wl + H(a)}{q} . \quad (17)$$

Conditions (15), (16), and (17) are identical to conditions (2), (3), and (4). Therefore, if the government sets fines according to the rule (10) and allows deductibility of fines, the equilibrium of the industry will be socially efficient.

Under the regime allowing deductibility of monetary sanctions, the schedule of fines that achieves efficient deterrence is identical to the schedule that is efficient in the absence of taxation. In particular, if enforcement is complete, $e = 1$, then the efficient fine is

$$F(a) = H(a),$$

which is the Pigouvian excise tax on an externality.

V. DEDUCTIBLE VS. NONDEDUCTIBLE APPROACH

We have outlined two approaches to adjust the current design and tax treatment of monetary sanctions to achieve efficient deterrence. The choice between allowing deductions for monetary sanctions or reducing the amount of the sanctions may depend on administrative considerations, particularly if sanctions and tax rates are set by different federal or state agencies. A policy of allowing deductions for sanctions may be less costly to administer. It allows the same sanction to apply in instances where the sanctions are applicable to tax-exempt entities and individuals engaged in personal, nondeductible activities as well as firms subject to income taxation. This approach enables the government to avoid maintaining two schedules of monetary sanctions: (1) sanctions for tax-exempt entities and personal, nondeductible activities determined according to the Becker rule, and (2) sanctions for firms subject to the tax system set at a lower level, adjusted for tax rates.

The deductible regime also does not require the government to adjust monetary sanctions when tax rates change. The rate reductions in the U.S. Tax Reform Act of 1986, for example, may significantly reduce the total costs of nondeductible sanctions for certain taxpayers.⁸ If the government sets sanctions considering the

⁸Public Law No. 99-514, §§101(a) and 601 (1986). Whether the 1986 Act changes reduce a taxpayer's total cost of fines and penalties depends on the change in the offender's marginal tax rates. The impact of the 1986 Act on a firm's effective marginal tax rate depends on two factors: first, the lower nominal tax rates introduced by the 1986 Act, and second, the effect of a broadening of the tax base on a firm's taxable income.

tax consequences to the firm based on a certain marginal rate, then it would be required to adjust the sanctions to reflect rate changes.

Finally, allowing deductions for all monetary sanctions makes the tax system more neutral by removing an exception to the broad principle that the tax system constitutes a tax on net income and not as a sanction against wrongdoing. It further eliminates the need to distinguish in designing sanctions between those sanctions that are deductible under current tax rules and those sanctions that are not deductible.

VI. CONCLUDING REMARKS

We have described the distortions resulting from failing to consider the tax consequences in designing monetary sanctions. Such sanctions distort a firm's efficient choice of input mix by providing an additional cost in the form of a disallowed deduction. Firms will substitute away from inputs that result in nondeductible sanctions toward inputs that are fully deductible. The inefficient mix results in higher social costs of production and consequently a higher price for the product of the industry. We therefore contend that either the government should allow deductibility for all amounts paid as monetary sanctions or it should reduce the amount of the sanction to reflect taxes. Either approach results in an efficient choice of inputs, scale of production, and size of the industry.⁹

Circumstances may exist, however, where the current system of disallowing deductions for amounts paid for some monetary sanctions closely approximates the optimal fine or penalty. A policy of nondeductibility of monetary sanctions may make sense if the government sets sanctions too low from either underestimating the amount of the external harm or not increasing the fine or penalty high enough to account for incomplete enforcement. Some contend, for example, that fines for air and water pollution are systematically too low, considering the amount of harm and the probability of enforcement.¹⁰

But this may begin to change. In its draft guidelines for organizational sanctions, the U.S. Sentencing Commission has proposed that courts apply monetary sanctions according to the Becker rule. For environmental offenses, the guidelines provide for monetary sanctions equal to the sum of (1) the harm caused subject to a minimum of \$5000 multiplied by a factor of at least 2.0 to reflect the difficulty of detecting and punishing the offender, and (2) the costs of enforcement.¹¹ In any event, if sanctions are set too low to achieve efficient deterrence, it seems preferable to increase the amount of the sanctions directly rather than use the tax system as a countervailing instrument.

Several important limitations apply to our conclusions. First, we assume a system of proportional income taxation. A progressive tax system affects the industry equilibrium in two ways: the marginal rate of tax distorts the firm's choice of inputs by making the labor input cheaper relative to air pollution; the average rate of tax distorts the firm's decision to enter or leave the industry. Because a firm's marginal and average rates differ under a progressive income tax, the government cannot achieve efficient deterrence by the single step of either ad-

⁹In related works, Rea (1989) analyzes the efficient tax treatment of tort awards.

¹⁰See Russell, Harrington, and Vaughn (1986), chapter 2.

¹¹See United States Sentencing Commission, *Discussion Draft of Sentencing Guidelines and Policy Statements for Organizations*, July 1988, pp. 8.2-8.30. The Commission proposed similar guidelines for food, drug, and agricultural offenses.

justing the sanction or allowing deductibility.¹² Second, we assume that all firms are risk-neutral. If the management of firms is risk-averse, sanctions must be further adjusted for their degree of risk aversion.¹³ Finally, we assume a constant probability of enforcement. In practice, the government chooses a probability that balances the marginal increase in social welfare from more enforcement with the marginal cost of increased enforcement.¹⁴

Although our analysis adopts the harm-based deterrence approach, we believe that failure to consider the tax consequences of monetary sanctions is a problem common to many deterrence theories. Thus, a proponent of a gain-based theory of deterrence should also be concerned that sanctions that ignore tax consequences will fail to approximate the gain from the offense. One who favors fashioning punishment that varies with the gravity of the offense, however gravity is measured, may object to a penalty scheme whose costs vary with the offender's tax bracket.¹⁵ Finally, one who adopts a standard of like punishment for like offenders should criticize the disparate treatment of offenders after considering tax consequences.

REFERENCES

- Becker, G. "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76 (1968):169-217.
- Polinsky, A. M. "Strict Liability vs Negligence in Market Setting," *American Economic Review*, 70 (1980):363-367.
- , and S. Shavell. "The Optimal Trade-off Between the Probability and Magnitude of Fines," *American Economic Review*, 69 (1979):880-891.
- Rea, S. "Taxation and Damage Awards," manuscript, Department of Economics, University of Toronto, July 1989.
- Russell, C., W. Harrington, and W. Vaughn. *Enforcing Pollution Control Laws*. Washington, D.C.: Resources for the Future (1986).
- Stigler, G. "The Optimum Enforcement of Laws," *Journal of Political Economy*, 78 (1970):526-536.
- U.S. Sentencing Commission. *Discussion Draft of Sentencing Guidelines and Policy Statements for Organizations*. Washington, D.C.: July 1988.

¹²A further consideration arises under the deductible approach: the firm's marginal rate of tax may be a function of the firm being successfully prosecuted. If the government fines the firm, the firm's taxable income is reduced and its marginal tax rate may be lower than if the government does not fine the firm.

¹³See Polinsky and Shavell (1979).

¹⁴See Becker (1968) and Stigler (1970).

¹⁵An argument exists, however, that this additional penalty for higher-bracket taxpayers may be necessary to deter higher-bracket taxpayers to the same degree as those in lower brackets.