DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

A General Characterization of Optimal Income Taxation and Enforcement

Parkash Chander
Indian Statistical Institute
California Institute of Technology

Louis Wilde
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 791

March 1992
Revised January 1994

A General Characterization of Optimal Income Taxation and Enforcement

Parkash Chander Louis Wilde

Abstract

This paper develops a general approach to characterizing optimal income tax enforcement. Our analysis clarifies the nature of the interplay between tax rates, audit probabilities, and penalties for misreporting. In particular, it is shown that for a variety of objective functions for the principal the optimal tax schedule is in general concave (at least weakly) and monotonic; the marginal tax rates determine the audit probabilities; and less harsh penalties lead to higher enforcement costs. Our results imply that there exists a tradeoff between equity and efficiency considerations in the enforcement context which is similar to that in the moral hazard context for tax policy.

Keywords: Principal, agent, informational asymmetry, costly verification, income taxes, audit probabilities, penalty for misreporting, revelation principle, efficient schemes, regressive taxes, equity-efficiency tradeoffs.

A General Characterization of Optimal Income Taxation and Enforcement

Parkash Chander Louis Wilde*

1 Introduction

It is well accepted by now that informational asymmetries generate important constraints on the formulation of economic policy. An early and classic example is given by Mirrlees (1971) who considers the problem of optimal income taxation when individuals differ in their ability to transform effort into income. It is assumed that ability and effort of an individual cannot be observed at any cost, whereas income, which is the product of these two unobservables, is perfectly and costlessly observable. Taxes can thus depend only on final income. Under these informational conditions the equity-efficiency tradeoffs that constrain the choice of tax policies arise only because of the presence of moral hazard associated with taxation of income. If there were no moral hazard (i.e. individual efforts and therefore incomes were exogenous), the problem of optimal income taxation would be a trivial one since, by assumption, individual incomes are perfectly and costlessly observable. In the presence of moral hazard the optimal taxes must strike the right balance between efficiency and equity considerations.

In much of the recent literature on optimal income tax enforcement (Reinganum and Wilde (1985), Border and Sobel (1987), and others), on the other hand, individual incomes are treated as exogenous, i.e., there is no moral hazard.[1] However, informational constraints arise from another important consideration; namely: an individual's income cannot be directly observed; it can only be verified through a costly audit. In this setting the question of optimal tax policy takes a different form. Besides tax rates, a policy must include an audit strategy and a scheme of penalties or fines for misreporting. This

*This is a revised version of Chander and Wilde (1992). We thank Kim Border, Luis Corchon, Rajeeva Karandikar, John Ledyard, Dilip Mookherjee, Ignacio Ortuno-Ortin, Thomas Palfrey, Till Requate and Kunal Sengupta for helpful comments and discussion. We have also benefitted from a seminar presentation at Caltech.

[1] An extensive literature on income tax enforcement has developed since the contributions of Allingham and Sandmo (1972) and Srinivasan (1973). Cowell (1990) provides a recent and relatively comprehensive review of that literature.

opens up additional interesting questions of optimal audit strategy, the nature of interaction between tax rates and audit strategy, and the role of penalties for misreporting. Although the traditional costs associated with moral hazard are absent, the issue of equity versus efficiency may not disappear but arise in a different way. Progressive taxes generate stronger incentives for individuals with higher incomes to underreport, thereby necessitating more auditing. Since audit expenditure is a direct resource cost, an optimal policy must weigh the welfare benefits from more progressive taxation against the concomitant increase in audit expenditure. Therefore, as in the optimal tax problem of Mirrlees, the central question is: How progressive should the income tax be?

Reinganum and Wilde (1985) propose a general formulation of the problem of optimal income taxation in the presence of costly enforcement. They compare a "random" audit strategy in which audits are independent of reported income to an "audit cutoff" strategy in which a taxpayer triggers an audit if and only if reported income is below some critical level.[2] By restricting attention to linear tax schedules, they show that a deterministic audit cutoff strategy in which all income reports below a certain cutoff level are audited with probability one, and higher income reports are not audited is optimal.[3] Border and Sobel (1987) analyze a model which is identical to Reinganum and Wilde's general formulation except that Reinganum and Wilde assume that income is continuously distributed over some non-negative interval and Border and Sobel assume that income takes a finite number of possible values. They show that optimal taxes are nondecreasing with income whereas optimal audit probabilities are nonincreasing.[4] They emphasize that in general optimal audit strategies involve random auditing. Similarly, Mookherjee and Png (1989) are interested particularly in finding general conditions under which random auditing is optimal. In a related paper, Melamud and Mookherjee (1989) show that the principal may implement any optimal random audit strategy by delegating responsibility for audits to an independent auditor. More recently, Cremer, Marchand and Pestieau (1990) analyze a model which is characterized by a linear income tax schedule and two-part audit probability strategy.[5] Sanchez and Sobel (1993) analyze a model which generalizes Cremer, Marchand and Pestieau to arbitrary audit strategies. In summary, most of the existing literature on optimal income tax enforcement emphasizes properties of optimal audit strategies as opposed to those of optimal tax schedules. This is perhaps because only limited qualitative results concerning the nature of the optimal tax schedule have been obtained. In particular, unlike in the case of the optimal tax problem of Mirrlees, the question of progressivity of the optimal tax schedule has not been resolved.

[2]As noted by these authors, the foundation of their analysis was provided by Townsend's (1979) analysis of optimal verification strategies in insurance contexts.

[3]Such audit strategies resemble debt contracts, provided the act of auditing is identified with bankruptcy, and have been shown to be optimal in other contexts also. See, for example, Dye (1986) and Morton (1986). Also see Border and Sobel's (1987) discussion of some specialized schemes and the issue of partial commitment.

[4]As noted by Krasa and Villamil (1993) these monotonicity results of Border and Sobel are important because they are consistent with certain stylized facts in insurance markets.

[5]The model analyzed by Cremer, Marchand and Pestieau is closely related to the audit cutoff model analyzed by Reinganum and Wilde (1985), their primary innovation being the embedding of the optimal tax and audit problem into a conventional social welfare maximization problem.

In this paper we will characterize the optimal tax schedule in the presence of costly enforcement. We will also clarify the nature of the interplay between optimal tax rates, audit probabilities, and penalties for misreporting.

We adopt the following model. The tax authority (principal) and taxpayers (agents) are risk neutral. The principal knows the probability distribution of income, but does not know who has what income, while each individual knows his own income. The principal may audit an agent in order to verify his income, but this is costly. The principal collects taxes from the agents through the following mechanism. The principal chooses schedules for pre-audit payment (tax) and post-audit payment (which for convenience we call penalty) and probability with which a message from an agent will trigger an audit. Each agent, treating the policy of the principal parametrically, sends a message so as to minimize his expected payment. We do not specify the objective function for the principal. Instead, as in Border and Sobel (1987), we introduce the notion of an efficient scheme such that for a variety of objective functions for the principal an optimal scheme must be efficient. Objectives of the principal could be to maximize gross revenue subject to a given limited auditing budget; to minimize audit costs subject to meeting some gross revenue target; to maximize revenue net of audit costs; or to maximize a Rawlsian social welfare function. For the case when the objective of the principal is to maximize a social welfare function, which reflects a concern for both equity and efficiency, an optimal scheme may be efficient, though this is not true in general. We thus characterize efficient schemes.

Our main results are as follows. As in most analyses of the principal-agent problem, the revelation principle (Myerson (1979); Harris and Townsend (1981)) holds. This means that without loss of generality we can restrict attention to incentive compatible direct revelation schemes. In an efficient scheme, the payment function is nondecreasing and concave, and the tax function is nondecreasing with nonincreasing average tax rates. The audit probabilities are nonincreasing. Only the downward incentive constraints are binding and the marginal payment rates determine the audit probabilities. An efficient scheme may involve either deterministic or random auditing, and there is really no qualitative difference between the two. We are also able to show that uniform lump-sum taxes are never optimal.

These results are robust with respect to two significant extensions also considered explicitly in the paper. The motivation for these comes from the fact that in reality penalties are often constrained more than allowed in our model. We are able to show that the results stated above are unaffected if we impose on penalties exogenous constraints of 'horizontal equity' or 'punishment should fit the crime.'

The major assumptions that limit our analysis are the following. As mentioned earlier, and, as in most of this literature, we rule out supply side effects of income taxation. Though supply side effects are important, we do not think that introducing them would reverse the qualitative nature of our results. As the results of Mirrlees (1971) show, these will only reinforce the regressivity of the tax/payment schedules. Cremer and Gahvari

(1993) analyze a model of optimal tax enforcement with moral hazard, but under the restriction that there are only two types of agents: skilled and unskilled. It is seen that their results, though limited, are not inconsistent with our results. Absence of supply side effects enables us to focus sharply on the enforcement problem and leads to a more transparent analysis.

The other major assumption is that the individual agents are risk neutral. Mookherjee and Png (1989) consider a model in which agents are risk averse, and the principal selects taxes and audit probabilities to maximize a utilitarian welfare function. They show that the revelation principle holds and obtain results that are weaker than the monotonicity results of Border and Sobel (1987), leaving open the question of monotonicity of optimal taxes and audit probabilities for the case with three or more possible income levels. Though this is the subject matter of a forthcoming paper by us, we describe the effect of risk aversion on our results later in the paper.

The paper proceeds as follows: Section 2 presents the model; Section 3 introduces the notion of efficiency; Sections 4 and 5 derive the results stated above; Section 6 makes the concluding remarks.

2 The Model

We consider a population consisting of a continuum of taxpayers (agents) who are distinguished by a one-dimensional measure of income. The incomes of the taxpayers are distributed over a nonnegative interval, which for convenience we take to be $[0, \bar{y}], \bar{y} > 0$, according to a probability density function g with $g(y) > 0$ for $y \in (0, \bar{y})$.

We shall sometimes refer to a 'taxpayer with income y' simply as 'taxpayer y' or 'agent y'.

A *mechanism* for the principal consists of a set M of messages; a function $t : M \to R$, to be called the *tax function*; a function $p : M \to [0, 1]$, to be called the *audit function*; and a function $f : M \times [0, \bar{y}] \to R$, to be called the *penalty function*. A taxpayer who reports the message m to the principal is audited with probability $p(m)$. In the event that an audit does not occur, his payment to the principal is $t(m)$. In the event that an audit occurs, the true income of the taxpayer is discovered, and his payment to the principal is $f(m, y)$ where y is the true income of the taxpayer.

We assume that a taxpayer can never pay more than his true income.[6] Therefore, in order to be feasible a mechanism must satisfy certain requirements. To begin with, a taxpayer of any given income must be able to submit a report that requires a payment which is not larger than his income. Define $M(y) = \{m \in M : t(m) \leq y\}$ to be the set of feasible messages for taxpayer y. Then the first feasibility requirement on a mechanism

[6] We later consider the implications of weakening this assumption.

is that for all $y \in [0, \bar{y}], M(y) \neq \phi$ and $f(m, y) \leq y$ for all $m \in M(y)$.[7]

We also assume that each taxpayer, treating the mechanism of the principal parametrically, submits a message to the principal so as to minimize his expected payment. Thus, taxpayer y minimizes $(1 - p(m))t(m) + p(m)f(m, y), m \in M(y)$. The second feasibility requirement on a mechanism is that the problem: minimize $[(1 - p(m))t(m) + p(m)f(m, y)], m \in M(y)$, has a solution for each $y \in [0, \bar{y}]$.

A mechanism is said to be a *direct revelation mechanism* if its message space is $[0, \bar{y}]$, and a direct revelation mechanism is said to be *incentive compatible* if it is optimal for each taxpayer to report his income truthfully.

The Revelation Principle: Let (t, p, f) be some feasible mechanism. Then there exists an incentive compatible direct revelation mechanism (t', p', f') which is equivalent from the point of view of the principal and each agent, when each agent reports his income truthfully.

Even though the set of feasible messages for each agent depends on his true income, it is easily seen that this principle holds. Consequently, we can confine our attention to the class of incentive compatible direct revelation mechanisms that are described by the set of schemes (t, p, f) that satisfy for each y

$$t(y) \leq y; \tag{1}$$

$$f(x, y) \leq y \text{ for all } x \text{ with } t(x) \leq y; \tag{2}$$

and

$$(1 - p(y))t(y) + p(y)f(y, y) \leq (1 - p(x))t(x) + p(x)f(x, y) \text{ for all } x \text{ with } t(x) \leq y. \tag{3}$$

These inequalities say that each taxpayer's expected payment is minimized if he reports his income truthfully.

As in most analyses of the principal-agent problem, we assume henceforth that an agent reports truthfully if reporting truthfully is optimal. Thus, given a scheme (t, p, f) that satisfies the inequalities in (1), (2) and (3), the expected gross revenue for the principal is

$$\int_0^{\bar{y}} r(y)g(y)dy, \tag{4}$$

[7] Given that $y \in [0, \bar{y}]$, an implication of this requirement is that there exists an $m \in M$ such that $t(m) \leq 0$.

where $r(y) \equiv (1 - p(y))t(y) + p(y)f(y,y)$ is the *payment function*. Observe that if $f(y,y) > t(y)$ then we can increase $t(y)$ and lower $f(y,y)$ such that $r(y)$ is unaffected. This will not affect the incentive constraints in (1) for agent y, and only weaken them for agents other than y. This means that $f(y,y)$ should be as small as possible. Thus, without loss of generality we can take $f(y,y) \leq t(y)$. Similarly, $f(x,y), x \neq y$ with $t(x) \leq y$ should be as large as possible which, in view of (3), is accomplished by setting $f(x,y) = y$ for $x \neq y$ with $t(x) \leq y$. With some notational inconsistency let f denote the function defined as $f(y) \equiv f(y,y)$. Then in view of the observations just made, we may consider only those schemes (t,p,f) that satisfy for each y the following inequalities:

$$0 \leq p(y) \leq 1; \tag{5}$$

$$f(y) \leq t(y) \leq y; \tag{6}$$

and the incentive constraints

$$r(y) = (1 - p(y))t(y) + p(y)f(y) \leq (1 - p(x))t(x) + p(x)y \text{ for all } x \text{ with } t(x) \leq y. \tag{7}$$

Note that the set of schemes (t,p,f) that satisfy these inequalities is closed but not compact. We know from the literature on economics of crime prevention (Becker (1968); Stigler (1970)) that if the feasible set is not compact then an optimal scheme may not exist. One familiar way to get around this problem is to impose some exogeneously given bounds on the feasible set so as to make it compact. This is accomplished in the present context by requiring the penalty function f to be bounded below. Thus, (6) may be replaced by

$$0 \leq f(y) \leq t(y) \leq y \text{ for all } y. \tag{8}$$

This inequality together with (7) says that if a taxpayer reported truthfully and was audited, then he may be given a reward which is not more than the full tax rebate.[8] This rules out the possibility of making the enforcement costs arbitrarily small by offering the agents large rewards with small probabilities and thus inducing them to do what the principal prefers.

Let Q_1 denote the set of all schemes that satisfy inequalities (5), (7), and (8). We shall call Q_1 the *set of all feasible schemes*.

[8]More generally (6) may be replaced by $d \leq f(y) \leq t(y) \leq y$, where $-\infty < d \leq 0$ is some constant. None of our results below depends qualitatively on the value of d. Leaving aside technicalities, other social and economic argument can be given for which the penalty function f must be bounded. See, for example, Cremer, Marchand and Pestieau (1990) for a further discussion.

We now describe a subset of schemes that are of an independent interest. The motivation for these comes from the fact that a scheme in Q_1, may not satisfy the principle of horizontal equity–the payments of otherwise identical taxpayers might be different.[9] The subset of schemes that satisfy the principle of horizontal equity is obtained by imposing an additional feasibility requirement on the principal's mechanism; namely: $f(m, y) \geq t(m)$ for all m and y. The revelation principle continues to hold and the set of incentive compatible direct revelation schemes reduces to

$$Q_2 = \{(t, p, f) \in Q_1 : f = t\}. \tag{9}$$

If $(t, p, f) \in Q_2$, then as seen from (7), $r(y) = t(y)$ for all y, i.e., the payment and tax functions are identical.

3 The Efficient Schemes

What is the appropriate objective function for the principal? This is an empirical question for which no clear-cut answer is readily available. We thus consider a variety of objective functions for the principal.

Let S be some subset of the feasible set Q_1, i.e., $S \subseteq Q_1$. We assume that other things being equal, the principal always prefers smaller audit probabilities as they reduce audit costs.

A scheme $(t, p, f) \in S$ is *efficient* in S if there is no other scheme $(t', p', f') \in S$ such that $p' \leq p, r' \geq r$, and $r' \neq r$ or $p' \neq p$, where r and r' are the payment functions corresponding to (t, p, f) and (t', p', f'). That is, it is not possible to not increase the audit probabilities and raise the payment of a taxpayer without lowering the payment of some other taxpayer, and it is not possible to not lower the payments of taxpayers and decrease an audit probability without increasing some other audit probability.

Note that the definition of efficiency does not involve the density function g. Thus if a scheme is efficient, it is efficient with respect to every g.

If the objective of the principal is to maximize gross (expected) revenue subject to a given limited auditing budget, then clearly an optimal scheme must be efficient.

A scheme $(t, p, f) \in S$ is *audit efficient* in S if there is no other scheme $(t', p', f') \in S$ such that $\int_0^{\bar{y}} r'(y)g(y)dy \geq \int_0^{\bar{y}} r(y)g(y)dy$, $p' \leq p$, and $p' \neq p$ where r and r' are the payment functions corresponding to (t, p, f) and (t', p', f'). That is, it is not possible

[9] See Stiglitz (1982) for a discussion of horizontal equity in the context of optimal income tax problem. Ortuno-Ortin and Roemer (1990) forcefully argue in favor of horizontal equity on legal and ethical grounds.

to collect at least the same gross revenue by decreasing an audit probability without increasing some other audit probability.

Border and Sobel (1987) note that for a variety of objective functions for the principal an optimal scheme must be audit efficient including when the principal may wish to maximize the sum of taxpayers' utilities subject to a net revenue constraint. Since, as can be shown, audit efficient schemes are efficient, it follows that for all those objective functions for the principal an optimal scheme must also be efficient.

A scheme $(t, p, f) \in S$ maximizes *revenue net of audit costs* in S if there exists no other scheme $(t', p', f') \in S$ such that $\int_0^{\bar{y}} r'(y)g(y)dy - c\int_0^{\bar{y}} p'(y)g(y)dy > \int_0^{\bar{y}} r(y)g(y)dy - c\int_0^{\bar{y}} p(y)g(y)dy$, where r and r' are the payment functions corresponding to (t, p, f) and (t', p', f') and $c > 0$ is the cost per audit.[10]

Clearly, a net revenue maximizing scheme in S must be efficient in S.[11]

We now come to the case when the objective of the principal is to maximize a social welfare function. To keep matters simple we shall restrict attention to schemes in the set Q_2, i.e., $S = Q_2$, where Q_2 is as defined in (9).

Let us first consider the case when the welfare function is additive and the social utility of an individual's disposable income is strictly increasing, strictly concave and the same for all individuals. If the incomes of the taxpayers were costlessly observable, the optimal policy would be a hundred percent tax on everyone, the proceeds of which are equally distributed in a lump-sum manner. This would lead to the most egalitarian distribution of after-tax income. In the presence of costly enforcement, however, welfare gains from a more equal distribution must be weighed against the loss in welfare due to the corresponding enforcement costs. Thus, we define:

$$\begin{array}{ll} \underset{(t,p,f)\in Q_2, \alpha\geq 0}{\text{maximize}} & \int_0^{\bar{y}} u(y - t(y) + \alpha)g(y)dy \\ \text{subject to} & \int_0^{\bar{y}} t(y)g(y)dy - c\int_0^{\bar{y}} p(y)g(y)dy \geq \alpha, \end{array}$$

where u is strictly increasing and strictly concave.[12]

A scheme $(t, p, f) \in Q_2$ is *Pareto efficient* in Q_2 if there is no other scheme $(t', p', f') \in Q_2$ such that $t'(y) - \alpha' \leq t(y) - \alpha$ for all y where $\alpha' \leq \int_0^{\bar{y}} t'(y)g(y)dy - c\int_0^{\bar{y}} p'(y)g(y)dy$ and $\alpha = \int_0^{\bar{y}} t(y)g(y)dy - c\int_0^{\bar{y}} p(y)g(y)dy$.

[10] More generally, we may have $c(y) > 0$ as the cost of verification of income of taxpayer y. As will be clear from below, such a generalization does not affect our analysis.

[11] Reinganum and Wilde (1985) describe conditions under which a nontrivial net revenue maximizing scheme exists and is unique. Scotchmer (1988) and Cremer, Marchand and Pestieau (1990) assume that the objective of the 'tax administration' is to maximize net revenue.

[12] This is very much like the welfare maximization problem analyzed by Cremer, Marchand, and Pestieau (1990) and Sanchez and Sobel (1993) except that the tax function is not assumed to be linear.

This simply says that it should not be possible to increase a taxpayer's disposable income without decreasing some other taxpayer's disposable income. Since $0 \leq t(y) \leq y$ for $y \in [0, \bar{y}]$, a net revenue maximizing scheme, if unique, is Pareto efficient. Of course, which Pareto efficient scheme is optimal depends on the welfare function. The more egalitarian the social welfare function, the larger might be the role of redistributive taxation. Therefore, we may also consider optimal schemes by assuming a Rawlsian social welfare function.[13] That is,

$$\begin{array}{ll} \underset{(t,p,f)\in Q_2, \alpha \geq 0}{\text{maximize}} & [\min_y \{u(y - t(y) + \alpha)\}] \\ \text{subject to} & \int_0^{\bar{y}} t(y)g(y)dy - c\int_0^{\bar{y}} p(y)g(y)dy \geq \alpha. \end{array}$$

Since $0 \leq t(y) \leq y$ for $y \in [0, \bar{y}]$, this is equivalent to maximizing net revenue α. Thus, an optimal scheme must be a net revenue maximizing scheme. It was noted earlier that a net revenue maximizing scheme is efficient. Thus, we characterize efficient schemes in Q_1, and Q_2.

4 A General Characterization

From inequalities in (7) and (8) it is seen that $0 \leq r(y) \leq t(y) \leq y$ for all y. From inequalities in (7) it is seen that if $p(y) = 1$ then $r(y) = f(y)$ and taking $t(y) = y$ will have no effect on the incentive constraints in (7) for taxpayer y. However, the incentive constraints in (7) for taxpayers other than y would be weaker if $t(y)$ is larger. Thus, in view of inequalities in (8), there is no loss of generality in assuming that

$$\text{if } p(y) = 1, \text{ then } t(y) = y. \tag{10}$$

If $t(y) < y$, and $p(y) > 0$, we can raise $t(y)$ and lower $f(y)$ unless $f(y) = 0$ such that $r(y)$ is not affected. This will not affect the incentive constraints in (7) for taxpayer y but will weaken the incentive constraints for taxpayers other than y because $t(y)$ is higher. If $p(y) = 0$, then as seen from (7), $f(y)$ does not matter. Thus as in the case of (10), there is no loss of generality in assuming that

$$\text{if } t(y) < y, \text{ then } f(y) = 0. \tag{11}$$

In order to fix our arguments and avoid discussion of many alternative cases, it is

[13] See Stiglitz (1987) for a discussion of the Rawlsian social welfare function in the context of the optimal income tax problem of Mirrlees and for emphasizing that results may vary greatly with the particular welfare function employed.

henceforth assumed that if $(t, p, f) \in Q_1$, then in addition to (5), (7), and (8) it also satisfies (10) and (11).

Theorem 1 : *A scheme $(t, p, f) \in Q_1$ with payment function r is efficient in Q_1 only if r and t are nondecreasing.*

Proof: We first show that r is nondecreasing. Suppose contrary to the assertion that there exist y and y' such that $y' > y$ and $r(y') < r(y)$. Since $r(y) \leq y$, from the incentive constraints for taxpayer y we have

$$p(x) \geq \frac{r(y) - t(x)}{y - t(x)} \text{ for all } x \text{ such that } r(y) > t(x).$$

Since $r(y') < r(y)$ and $y' > y$, the above inequalities imply that

$$p(x) > \frac{r(y') - t(x)}{y' - t(x)} \text{ for all } x \text{ such that } r(y') > t(x).$$

This means that the audit probabilities are such that the incentive constraints for taxpayer y' are nonbinding. We can therefore raise the tax and payment of taxpayer y' such that his incentive constraints continue to be nonbinding. Raising the tax and payment of taxpayer y' will not affect the incentive constraints in (7) for other taxpayers. Therefore, keeping the audit probabilities the same, the payment of a taxpayer can be raised without lowering the payment of some other taxpayer. This contradicts our supposition that (t, p, f) is efficient. Hence r must be nondecreasing.

That r is nondecreasing implies that t is nondecreasing. Suppose not, i.e., there exist y and y' such that $y' > y$ and $t(y') < t(y)$. As seen from (8), this means that $t(y') < y'$ and thus, in view of (11), $f(y') = 0$ and $r(y') = (1 - p(y'))t(y')$. Since $r(y') \geq r(y)$ as shown above, $(1 - p(y'))t(y') \geq (1 - p(y))t(y) + p(y)f(y)$. And since $t(y') < t(y)$ by supposition and $f(y) \geq 0$, we must have $p(y') < p(y)$. This is, however, a contradiction, since as seen from the inequalities in (7), $p(y')$ and $p(y)$ must be such that

$$p(y') \geq \frac{r(z) - t(y')}{z - t(y')} \text{ for all } z \text{ such that } z > t(y')$$

and

$$p(y) \geq \frac{r(z) - t(y)}{z - t(y)} \text{ for all } z \text{ such that } z > t(y),$$

from which it follows that $p(y') \geq p(y)$ (because $r(z) \leq z$ for all z and $t(y') < t(y)$ by supposition). Hence t must be nondecreasing. This completes the proof. ∎

Corollary 1.1: *If $(t,p,f),(t',p',f') \in Q_1$ are efficient in Q_1 and $p = p'$, then $t = t'$ and $f = f'$.*[14]

Proof: Let r and r' be the corresponding payment functions. If $r = r'$, then in view of (11) and $p = p'$, $t = t'$ and $f = f'$. If $r \neq r'$, then first note that in view of (11) and $p = p'$, $r(y) \geq (\leq) r'(y) \Rightarrow t(y) \geq (\leq) t'(y)$. Define for each $y, t''(y) = \max\{t'(y), t(y)\}$ and $r''(y) = \max\{r'(y), r(y)\}$. It is easily seen that the scheme (t'', p, f'') with payment function r'' belongs to Q_1 and is such that $r'' \geq r, r'' \neq r$ (since $r \neq r'$) which contradicts that (t,p,f) is efficient in Q_1. This completes the proof. ∎

Corollary 1.2: *If $(t,p,f) \in Q_2$ is efficient in Q_2, then t is nondecreasing.*

Proof: The proof is analogous to that of Theorem 1 and hence it is omitted. ∎

Theorem 2 : *A scheme $(t,p,f) \in Q_1$ with payment function r is efficient in Q_1 only if r is concave, t is continuous and $t(y)/y$ is nonincreasing, and p is nonincreasing.*

Proof: We first show that r must be concave. Suppose not. Let r' be the smallest concave function above r. Since by supposition r is not concave, $r' \geq r$ and $r' \neq r$.

Since r is nondecreasing (as shown in Theorem 1), r' is nondecreasing (see Lemma 0 in the Appendix). Furthermore, since $0 \leq r(z) \leq z$ for all z, by definition of r', $0 \leq r'(z) \leq z$ for all z. Thus, $0 \leq r(y) \leq r'(y) \leq y$ for all y. Define t' and f' from r' such that $r'(y) = (1 - p(y))t'(y) + p(y)f'(y), t'(y) \geq f'(y) \geq 0$, and $f'(y) = 0$ if $t'(y) < y$. Clearly, $t'(y) \geq t(y)$ and $0 \leq t'(y) \leq y$ for all y, since $0 \leq r(y) \leq r'(y) \leq y$. Also, $t'(y) \geq r'(y)$ for all y.

Since $(t,p,f) \in Q_1, p$ must be such that for each x

$$p(x) \geq \sup_{y > t(x)} \frac{r(y) - t(x)}{y - t(x)} \tag{12}$$

Since $t'(x) \geq t(x)$ and $y \geq r(y)$, for each x

$$\frac{r(y) - t'(x)}{y - t'(x)} \leq \frac{r(y) - t(x)}{y - t(x)} \text{ for all } y \text{ such that } y > t'(x).$$

[14] As noted by Border and Sobel (1987), Corollary 1.1 is in essence similar to the result that in an optimal auction, the transfers among the players are completely determined by the probabilities of winning the prize (Myerson (1981)).

Therefore, for each x

$$\sup_{y>t'(x)} \frac{r(y)-t'(x)}{y-t'(x)} \leq \sup_{y>t'(x)} \frac{r(y)-t(x)}{y-t(x)}. \tag{13}$$

We claim that for each x

$$\sup_{y>t'(x)} \frac{r'(y)-t'(x)}{y-t'(x)} \leq \sup_{y>t'(x)} \frac{r(y)-t'(x)}{y-t'(x)}. \tag{14}$$

Suppose contrary to the assertion that the expression on the left of (14) is greater than β_x for some x fixed, where β_x denotes the expression on the right. Since $r(y) \leq y$ for all y, $\beta_x \leq 1$. Define the line $l_x(y) = t'(x) + \beta_x(y - t'(x))$. Clearly, r lies below the line l_x for $y > t'(x)$. Since $\beta_x \leq 1$ and $0 \leq r(y) \leq y, r$ also lies below l_x for $y \leq t'(x)$. Thus all of r lies below l_x. But this contradicts that r' is the smallest concave function above r since if β_x is less than the expression on the left of (14) then there must be some $z > t'(x)$ such that $r'(z) > l_x(z)$, where l_x is a linear (and therefore concave) function above r. Hence we must have

$$\beta_x \geq \sup_{y>t'(x)} \frac{r'(y)-t'(x)}{y-t'(x)},$$

which proves (14). Let t', f' be as defined above from r'. It follows from (12), (13), and (14) that (t', p', f') with $p' = p$ and payment function r' belongs to Q_1. Furthermore, $r' \geq r$, and $r' \neq r$. This contradicts that (t, p, f) is efficient in Q_1. Hence the payment function r must be concave.

Next we show that p must be nonincreasing. Since $(t, p, f) \in Q_1$ is efficient in Q_1, p must be such that

$$p(x) = \sup_{r(y)>t(x)} \frac{r(y)-t(x)}{y-t(x)} \tag{15}$$

Since t is nondecreasing, the above equality clearly implies that p is nonincreasing. If for some x there is no y such that $t(x) < r(y)$, then since t is nondecreasing we have, in view of (7), $p(z) = 0$ for all $z \geq x$ which again shows that p is nonincreasing.

From the concavity of r and that p is nonincreasing, we show that $t(y)/y$ is nonincreasing. If t is such that $t(y) = y$ for all y then $t(y)/y$ is nonincreasing. On the other hand, let some y be such that $t(y) < y$. Then, in view of (11), $f(y) = 0$ and $r(y) = (1 - p(y))t(y)$. Since p is nonincreasing, $(1 - p(z)) \geq (1 - p(y))$ for all $z > y$. Since r is concave, $r(z)/z$ is nonincreasing with z. Therefore, for $y' > y$

$$\frac{r(y')}{y'} = (1 - p(y'))\frac{t(y')}{y'} + p(y')\frac{f(y')}{y'} \leq (1 - p(y))\frac{t(y)}{y} = \frac{r(y)}{y}$$

and

$$(1 - p(y')) \geq (1 - p(y)).$$

Since $f(y') \geq 0$, these inequalities imply that $t(y')/y' \leq t(y)/y$. Thus, if $t(y) < y$ for some y, then $t(y') < y'$ for all $y' > y$. Since $[0, \bar{y}]$ is compact, there exists a $\tilde{y} \in [0, \bar{y}]$ such that $t(y) < y$ only for all $y > \tilde{y}$ (or $t(y) = y$ only for all $y \leq \tilde{y}$). This proves that $t(y)/y$ is nonincreasing with y.

Since t is nondecreasing

$$\lim_{h \to 0} t(y - h) \leq \lim_{h \to 0} t(y + h);$$

and since $t(y)/y$ is nonincreasing

$$\lim_{h \to 0} \frac{t(y - h)}{y - h} \geq \lim_{h \to 0} \frac{t(y + h)}{y + h}$$

Therefore,

$$\lim_{h \to 0} t(y - h) = \lim_{h \to 0} t(y + h),$$

that is, t is continuous. This completes the proof of Theorem 2. ∎

Corollary 2.1: *If $p(x') = p(x)$ for some $x' > x$, then either $p(x') = p(x) = 1$ or $p(x') = p(x) = 0$.*

Proof: Since t is nondecreasing, it is seen from (15) that $p(x') = p(x)$ only if either $r(y) = y$ for some y such that $r(y) > t(x') \geq t(x)$ or if $t(x') = t(x)$. In the first case, (15) implies $p(x') = p(x) = 1$. In the second case, $r(x') = r(x)$, since r is nondecreasing and $x' > x \geq t(x) = t(x')$. Since r is concave and nondecreasing $r(z) = r(x)$ for all $z \geq x$. Since $t(x) \geq r(x)$, there is no $y > x$ such that $r(y) > t(x)$ which as noted earlier implies $p(x) = 0$. Since p is nonincreasing $p(x') = 0$. Hence the corollary. ∎

Corollary 2.2: *For each x there exists a $\hat{y} \in [0,\bar{y}]$ such that $p(x) = (r(\hat{y}) - t(x))/(\hat{y} - t(x)) \geq (r(y) - t(x))/(y - t(x))$ for all $y > t(x)$.*

Proof: Since r is concave, it is continuous over the compact set $[0, \bar{y}]$. Thus, the expression on the right of (15) must attain its supremum at some $\hat{y} \in [0, \bar{y}]$. Hence the corollary. ∎

Corollary 2.3: *A scheme $(t, p, f) \in Q_2$ is efficient in Q_2 only if t is concave and p is nonincreasing.*

In order to interpret our results we display in Figure 1 a scheme which is efficient in Q_1.

[FIGURE 1 ABOUT HERE]

The payment function r is shown to be concave and nondecreasing. The tax function t is shown to be nondecreasing with nonincreasing average tax rates. The tax function is shown to be above the payment function, since $t(z) \geq r(z)$ for all z. The audit probability $p(x)$ at x is as defined in (15), and as seen from the figure equal to the marginal payment rate at a higher point y. Hence, the marginal payment rates determine the audit probabilities.

Figure 2 similarly displays a scheme which is efficient in Q_2. The tax and payment function are shown to be identical.

[FIGURE 2 ABOUT HERE]

In order to understand the roll of rewards (or "rebates") for truthful reporting, the point x and the payment function r in Figure 2 are drawn identically to those in Figure 1. It is seen that the audit probability $p(x)$ is higher. It is higher because the tax function is not above the payment function. Thus, other things being equal, horizontal equity constraints lead to higher audit costs.

The efficient schemes displayed in Figures 1 and 2 involve random auditing since the marginal payment rates are strictly between zero and one. Figure 3 displays a scheme in which the marginal payment rate is either zero or one. It is seen that for this case the audit probabilities as defined in (15) are deterministic, i.e., $p(x) \in \{0, 1\}$. It is also easily seen that this scheme is efficient in both Q_1 and Q_2: it is not possible to decrease an audit probability without lowering some taxpayer's payment; and it is not possible to increase a taxpayer's payment without increasing some audit probability. As in the case of efficient schemes that involve random auditing, the audit probabilities in this case are also determined by the marginal payment rates. Thus, efficient schemes may involve random or deterministic auditing.

[FIGURE 3 ABOUT HERE]

The result that the payment and the tax are nondecreasing with income is along the

expected lines. The result that the low income tax payers are more likely to be audited may seem counter intuitive. However, when interpreted carefully it seems reasonable. First it is low income reports and not low income taxpayers that are audited more intensively; low income taxpayers get 'adversely selected'. Secondly, the result applies to a population of taxpayers who are otherwise indistinguishable by occupation, residential location, source of income, etc. A more careful interpretation of the result is that within each category of taxpayers (defined according to exogenously available information on other characteristics) a taxpayer is more likely to be audited if he reports low income.

The result that the payment function r is concave (at least weakly) is explained intuitively as follows. Only the downward incentive constraints are relevant and the sole purpose of auditing is to deter strategic misreporting. Keeping the taxes high for the low income taxpayers deters high income taxpayers from underreporting and thus saves audit expenditure. Similarly, relatively less high payments and taxes for high income taxpayers weakens their incentives to underreport and save audit expenditure. A concave payment function is particularly suitable in both these respects.

We assumed implicitly that individuals can survive with any positive level of income, no matter how small. If there exists a minimum income level $\bar{\alpha} > 0$ necessary for survival, we can suitably redefine the feasibility requirements for the principal's mechanism such that the after-tax income of a taxpayer is never less than $\bar{\alpha}$. The revelation principle holds and an incentive compatible direct revelation scheme (t, p, f) under the new feasibility requirements is technically equivalent to a scheme $(t', p, f') \in Q_1$ such that $t = t' - \bar{\alpha}$ and $f = f' - \bar{\alpha}$. Since subtracting a constant from a function does not alter its qualitative properties in an essential way, the efficient schemes under the new feasibility requirements have more or less the same characteristics.

We assumed that the agents are risk neutral. We now describe how our results are affected when the agents are risk averse, i.e., when all agents have the same von Neumann-Morgenstern utility function. As noted earlier, the revelation principle holds and we can define the corresponding sets of schemes Q_1 and Q_2 by suitably rewriting the incentive constraints in (7). For the efficient schemes in the so-defined set Q_2, we obtain a result that parallels Theorem 1 or Corollary 1.2. Furthermore, for a class of utility functions that satisfy nonincreasing absolute risk aversion and that includes those with constant relative risk aversion, we obtain a result that parallels Theorem 2 or Corollary 2.3. The techniques of proof are, however, different. The results for efficient schemes in Q_1 are weaker in that the monotonicity results hold for the case with three possible income levels, but that concavity in general does not. The results are weaker in this case because risk aversion introduces an important qualitative difference between the schemes in Q_1 and Q_2. For a scheme in Q_2, an agent's utility (because of the horizontal equity constraints) depends only on his after-tax income. Whereas for a scheme in Q_1 an agent's (expected) utility, unlike the risk neutral case, depends not only on his (expected) after-tax income, but also on the probability of being audited. This introduces an additional complexity in the relationship between the audit probabilities and the payment rates and weakens the results.

We assumed that the audits are perfect, and the principal discovers an agent's true income without error. However, as noted by Mookherjee and Png (1990), continuity arguments can be used to generalize the results to the case where there are "small" errors in auditing.

We assumed that a taxpayer can never pay more than his true income. We now consider the implications of weakening this assumption. We assume instead that a taxpayer can pay up to $\bar{y}$ (the highest income level) but the mechanism for the principal is such that each taxpayer can submit a report that requires a payment which is not larger than his true income and the penalty is never more than the true income of a taxpayer. It is easily seen that the revelation principle holds. The inequalities in (8) are modified as

$$0 \leq f(y) \leq t(y) \leq \bar{y}; \tag{8'}$$

and that $0 \leq r(y) \leq y$ for all y. The inequalities in (7) are affected only to the extent that '$t(x) \leq y$' in the statement is replaced by '$t(x) \leq \bar{y}$'. Without loss of generality we can assume that (10) holds with y on the right replaced by $\bar{y}$, and that

$$\text{if } t(y) < \bar{y}, \text{ then } f(y) = 0. \tag{11'}$$

The generalization proposed above is of interest not because it is more realistic to assume that a taxpayer can pay more than his income, but because as seen from (8′) it allows us to consider schemes with nontrivial uniform lump-sum taxes.

Let Q denote the set of schemes that satisfy (6), suitably modified (7) and (10), (8′), and (11′).

We characterize schemes that are efficient in Q.[15]

Theorem 3 : *A scheme $(t, p, f) \in Q$ with payment function r is efficient in Q only if (a) r is nondecreasing and concave, t is nondecreasing, and $t(y)/y$ and p are nonincreasing; and (b) $r(y) = (1 - p(y))t(y)$ for all $y \in [0, \bar{y}]$.*

Proof: The proof for statement (a) follows from similar arguments as in Theorems 1 and 2, and hence it is omitted. We give proof for statement (b). If $(t, p, f) \in Q$ is efficient in Q, then as in (15)

$$p(y) = \sup_{r(z) > t(y)} \frac{r(z) - t(y)}{z - t(y)} \tag{16}$$

[15] It may be noted that Corollary 3.1 below is unaffected if we replace (8′) by $d \leq f(y) \leq t(y) \leq \bar{y}$, where $-\infty < d < 0$ is some constant.

Note that if there is no z such that $r(z) > t(y)$ then we must have $p(y) = 0$, and therefore $r(y) = t(y)$. In particular, we must have $p(\bar{y}) = 0$, since t is nondecreasing.

In view of (11′), we need to show that $t(y) < \bar{y}$ for all $y < \bar{y}$. Suppose not; i.e., $t(y) = \bar{y}$ for some $y < \bar{y}$. Since t is nondecreasing, $t(x) = \bar{y}$ for all $x \in [y, \bar{y}]$. But this implies $p(x) = 0$ for all $x \in [y, \bar{y}]$ and $r(y) = \bar{y}$. This is a contradiction, since (as noted earlier) $r(y) \leq y$. Hence, $t(y) < \bar{y}$ for all $y < \bar{y}$. It was noted earlier that $p(\bar{y}) = 0$. Hence, the theorem. ∎

Corollary 3.1: *Let $(t, p, f) \in Q$ be such $t(y) = \bar{\alpha}$ for all y where $0 < \bar{\alpha} < \bar{y}$. Then (t, p, f) is not efficient in Q.*

Proof: Since $t(y) = \bar{\alpha}$ for all y, it follows from (16) that $p(y) = p(y')$ for all $y, y' \in [0, \bar{y}]$. Therefore, as in Corollary 2.1, either $p(y) = 1$ for all y or $p(y) = 0$ for all y. In the first case, $r(y) = (1 - p(y))t(y) = 0$ which shows that (t, p, f) is not efficient. In the second case, $r(y) = \bar{\alpha}$ for all $y \in [0, \bar{y}]$ which contradicts that $r(y) \leq y$ for all $y \in [0, \bar{y}]$. Hence the corollary. ∎

5 Marginal Deterrence and the Efficient Schemes

In our analysis so far, penalties, like taxes and audit probabilities, are determined endogeneously subject only to an upper and a lower bound. We know from the literature on economics of crime prevention (Becker (1968); and Stigler (1970)) that endogenous determination of penalties gives rise to what is often called the 'Becker conundrum'–it is optimal to increase penalties as far as possible and minimize the probability of costly auditing. Accordingly, in the present context we have $f(x, y) = y$ for all $x \neq y$. That is, the penalty for misreporting, no matter how insignificant, is extreme. This is clearly at odds with actual practice in most democratic societies. Stigler (1970) provides a possible argument as to why penalties may not always be extreme; namely: offenses vary in severity and extreme penalties for all offenses would provide no marginal deterrence for an offender not to commit a more serious offense.

From another perspective also it has been argued that it is unrealistic to assume that penalties are a choice variable of the 'tax authorities'–penalties, unlike taxes and audit probabilities, are determined by social convention and by principles of law which constrain them by the common ethical norms of horizontal equity and letting the punishment fit the crime.

We have already analyzed the consequences of imposing the norm of horizontal equity. The considerations above motivate us to now assume that the penalty function is given exogenously. More specifically, we introduce a particular exogenously given penalty

function, which emphasizes marginal deterrence or the norm of letting the punishment fit the crime. That penalty function is defined as follows:[16]

$$f_t(x,y) = t(x) + \max\{0, y-x\} \text{ for all } x \text{ and } y. \tag{17}$$

This says that a taxpayer with true income y who reports x and is audited pays the tax $t(x)$ and all of the underreported income, and there is no reward for overreporting income.

Let $R = \{(t,p,f) \quad f = f_t, 0 \leq t(x) \leq x, 0 \leq p(x) < 1 \text{ for all } x\}$. We characterize efficient schemes in R.

Lemma 4: *Let $(t,p,f) \in R$ be some scheme. Then there exists an incentive compatible scheme $(t',p',f') \in R$ such that t' is nondecreasing and (t',p',f') is equivalent from the point of view of the principle and each agent when each agent reports his income truthfully.*

This version of the revelation principle does not follow from standard arguments. Therefore, a proof, which involves a constructive argument, is given in the Appendix.

In view of Lemma 4, we confine our attention to schemes that satisfy for each y

$$\text{if } y' > y, \text{ then } t(y') \geq t(y); \tag{18}$$

$$0 \leq p(y) \leq 1; \tag{19}$$

$$0 \leq t(y) \leq y; \tag{20}$$

and incentive constraints

$$\begin{aligned} r(y) \equiv t(y) &\leq (1-p(x))t(x) + p(x)(t(x)+y-x) \\ &= t(x) + p(x)(y-x) \text{ for all } x < y. \end{aligned} \tag{21}$$

Let S denote the set of schemes $(t,p,f) \in R$ that satisfy (18) to (21).

[16] More appropriately, the penalty function is defined as $f_t(x,y) = \min\{t(x), y\} + \max\{0, y-x\}$ for all x and y. But this distinction, as can be seen from below, does not matter. One could have also considered $f(x,y) = t(x) + \max\{0, (1+\pi)(t(y)-t(x))\}$ for all x and y where $\pi > 0$ is the exogeneously given 'penalty rate'. But this one implies a penalty which is more than the true income of a taxpayer for at least the tax function defined as $t(x) = x$ for all x.

Theorem 5: *A scheme $(t, p, f) \in S$ is efficient in S only if t is concave and nondecreasing and p is nonincreasing.*

Proof: Since $(t, p, f) \in S, t$ is nondecreasing, and for each x

$$p(x) \geq \sup_{y>x} \frac{t(y) - t(x)}{y - x} \geq 0. \tag{22}$$

Suppose contrary to the assertion that (t, p, f) is efficient but t is not concave. Let t' be the smallest concave function above t such that $t' \geq t, t' \neq t$. By Lemma 0, t' is nondecreasing, and by definition of $t', 0 \leq t'(z) \leq z$ for all z. Define

$$p'(x) = \sup_{y>x} \frac{t'(y) - t'(x)}{y - x}. \tag{23}$$

We claim that

$$p'(x) \leq \sup_{y>x} \frac{t(y) - t(x)}{y - x}. \tag{24}$$

Suppose not, i.e., $p'(x) > \beta_x$ for some x fixed, where β_x denotes the expression on the right of (24). Consider the line $l_x(y) = t'(x) + \beta_x(y - x)$. Since $t'(x) \geq t(x)$, it follows from the definition of β_x that $l_x(y) \geq t(y)$ for all $y \geq x$. On the other hand, concavity of t' implies that $l_x(y) \geq t'(y)$ for all $y < x$ and $l_x(y) < t'(y)$ for at least one $y > x$, since by supposition $\beta_x < p'(x)$. Thus, $l_x(y) \geq t'(y) \geq t(y)$ for all $y < x$. Hence, l_x is a linear function above t, and $l_x(y) < t'(y)$ for at least some $y > x$. Let k denote the function defined as $k(y) \equiv \inf\{t'(y), l_x(y)\}$. Then, k is a concave function above t such that $k(y) < t'(y)$ for some $y > x$. But this contradicts the fact that t' is the smallest concave function above t. Hence (24) must be true.

Consider the scheme (t', p', f'), where $f' = f_{t'}$. Then $(t', p', f') \in S$. Thus, $(t', p', f') \in S$ is such that $r' \geq r, r' \neq r$, and $p' \leq p$. But this contradicts our supposition that (t, p, f) is efficient in S. Hence t must be concave.

From concavity of t we show that p is nonincreasing. Since $(t, p, f) \in S$ is efficient in S, p must be such that for each x

$$p(x) = \sup_{y>x} \frac{t(y) - t(x)}{y - x}. \tag{25}$$

Since t is concave, the supremum is attained at $y = x$, i.e. $p(x) = D^+t(x)$–the right hand derivative of t at x–which is equal to the derivative if t is differentiable at x which it is

at almost all x because it is concave. Since t is concave, p must be nonincreasing. This completes the proof. ∎

Corollary 5.1: *If $(t, p, f) \in S$ is efficient in S, then for each x, $p(x) = D^+t(x)$.*

Let $Dt(x)$ denote the derivative of t at x whenever it exists. Let G denote the distribution function, i.e., $G(x) = \int_0^x g(z)dz$. Let $c > 0$ be the cost per audit.

Corollary 5.2: *If $(t, p, f) \in S$ is efficient in S, then the net revenue corresponding to (t, p, f) is $\int_0^{\bar{y}} Dt(x)(1 - G(x) - cg(x))dx$.*

Proof: Since $r = t$, in view of Corollary 5.1 the net revenue corresponding to (t, p, f) is

$$\begin{aligned}
&\textstyle\int_0^{\bar{y}} t(x)g(x)dx - c\int_0^{\bar{y}} Dt(x)g(x)dx \\
&\text{(integrating by parts, which is legitimate since} \\
&t \text{ is concave and therefore absolutely continuous)} \\
&= t(\bar{y}) - \textstyle\int_0^{\bar{y}} Dt(x)G(x)dx - c\int_0^{\bar{y}} Dt(x)g(x)dx \\
&= \textstyle\int_0^{\bar{y}} Dt(x)(1 - G(x) - cg(x))dx.
\end{aligned}$$

This completes the proof. ∎

That the supremum of the expression on the right of (25) is attained at $y = x$ means that the local downward incentive constraints are in general binding in an efficient scheme in S.[17] As noted earlier, only downward incentive constraints are in general binding in an efficient scheme in Q_1 or Q_2. What is the crucial difference between schemes in S and those in Q_1 or Q_2 that gives rise to this result? It seems that it comes from the marginal deterrence aspect of the penalty function–if the audit probabilities are high enough for deterring misreporting by small amounts then they are also high enough for deterring misreporting by large amounts.

[FIGURE 4 ABOUT HERE]

An efficient scheme in S is displayed in Figure 4. For the sake of comparison the point x and the payment/tax function are drawn identically to those in Figures 1 and 2. It is seen that audit probability $p(x)$ is higher. Thus, penalties that emphasize marginal deterrence lead to higher audit costs. It is easily seen that the scheme displayed in Figure 3 is efficient in S also. Therefore, it is not surprising that deterministic schemes, as noted in footnote 3, have shown up in a variety of contexts.

Corollary 5.2 shows how the net revenue corresponding to an efficient scheme $(t, p, f) \in S$ may be calculated. Therefore, numerical calculations for comparison purposes are

[17] This is similar to results in the auction design problem (Maskin and Riley (1984); and Myerson (1981)) and in the theory of bilateral trade (Myerson and Satterthwaite (1983)).

possible. This corollary can also be used to prove the existence and uniqueness of net revenue maximizing schemes, which as noted earlier are Pareto efficient.

6 Concluding Remarks

In this paper we have developed a general approach to characterizing optimal income tax enforcement. Our results clarify the nature of the interplay between tax rates, audit probabilities, and penalties for misreporting. For a variety of objective functions for the principal, the optimal income tax schedule is in general concave (at least weakly) and monotonic. The marginal payment rates determine the audit probabilities; and less harsh penalties lead to higher enforcement costs. In many ways our results mirror Mirrlees' results, particularly the seemingly ubiquitous regressivity of optimal income taxes.

What are we to conclude from these results? We do not think that regressive taxes can ever be advocated or voted for in a democratic society. However, our analysis does show that there exists a tradeoff between equity and efficiency considerations; a tradeoff that must be taken into account in the choice of tax policy. Our results provide less support for progressive taxes than what one might have hoped.

We restricted our analysis to the case when the agents are risk neutral. However, as already mentioned, in a later paper we will extend some of our results to the case when the agents are risk averse.

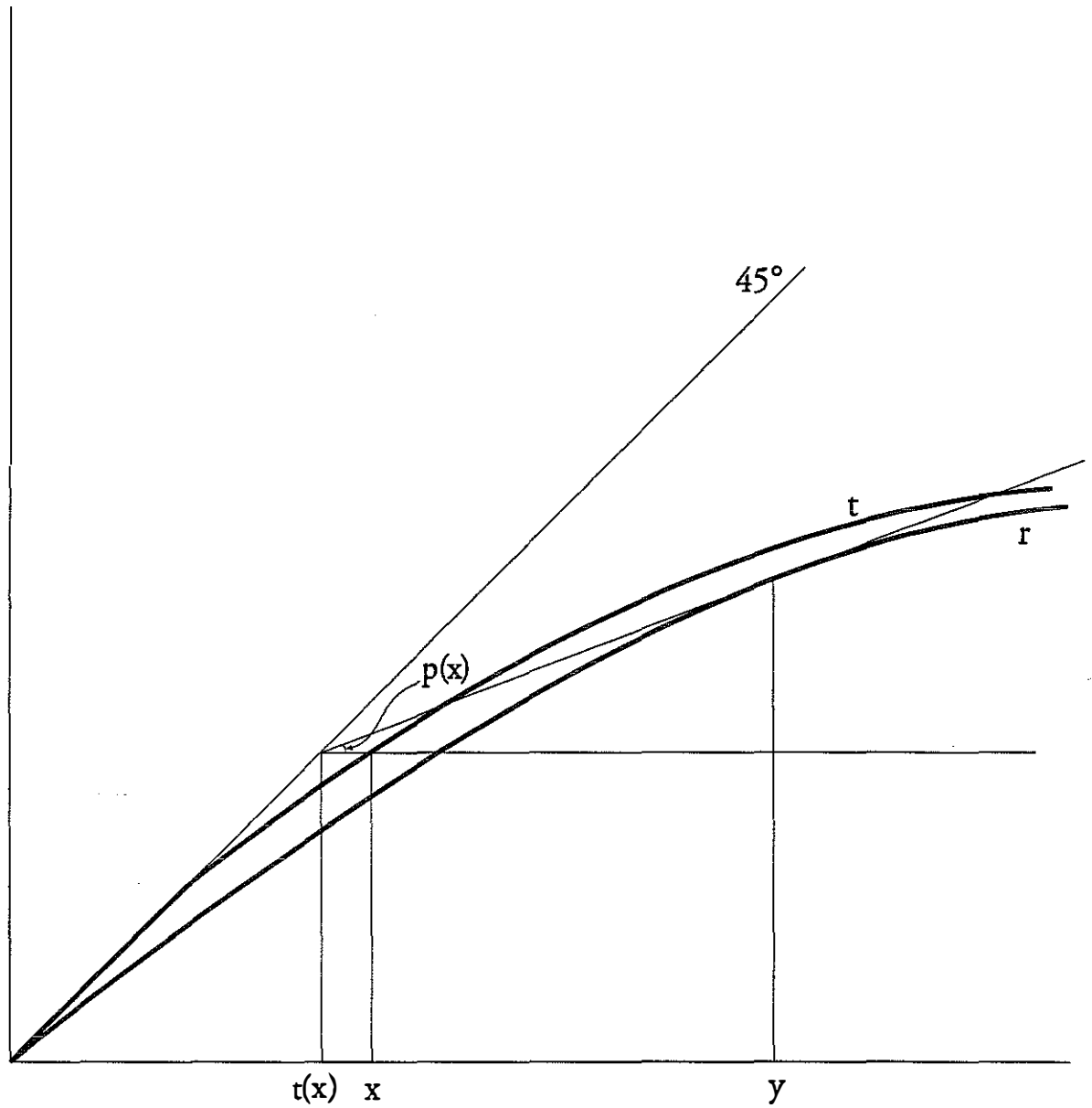

Figure 1

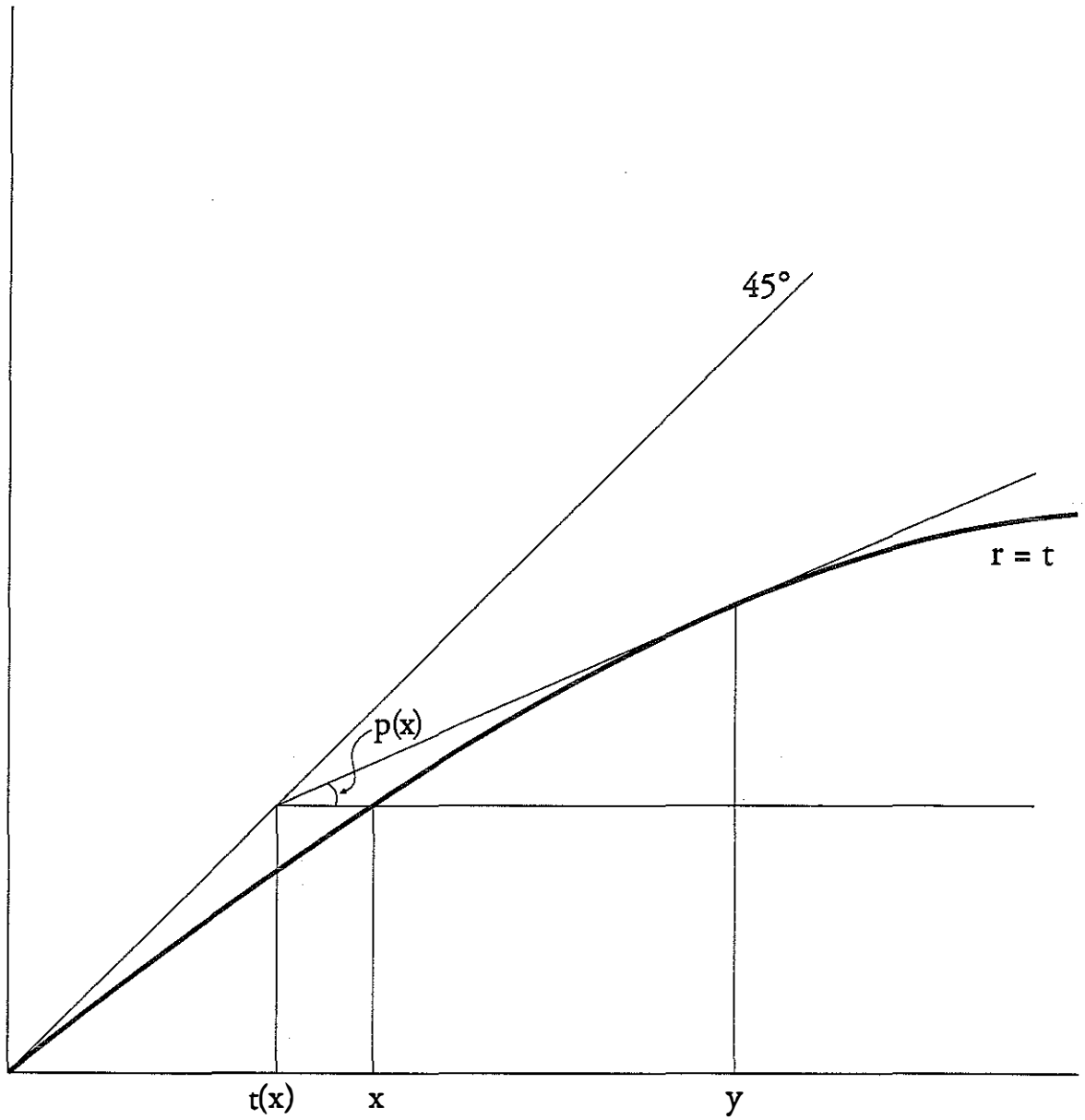

Figure 2

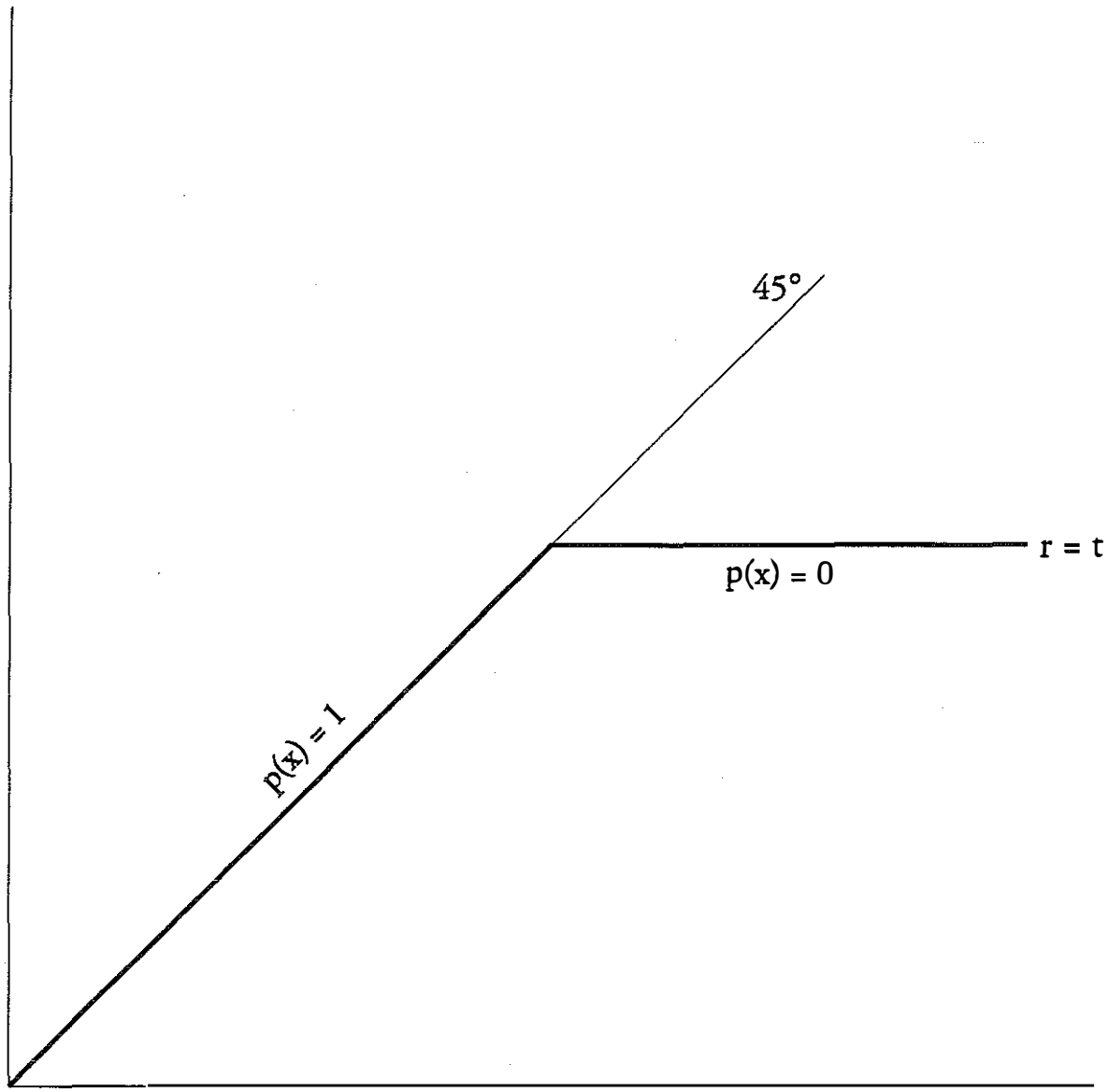

Figure 3

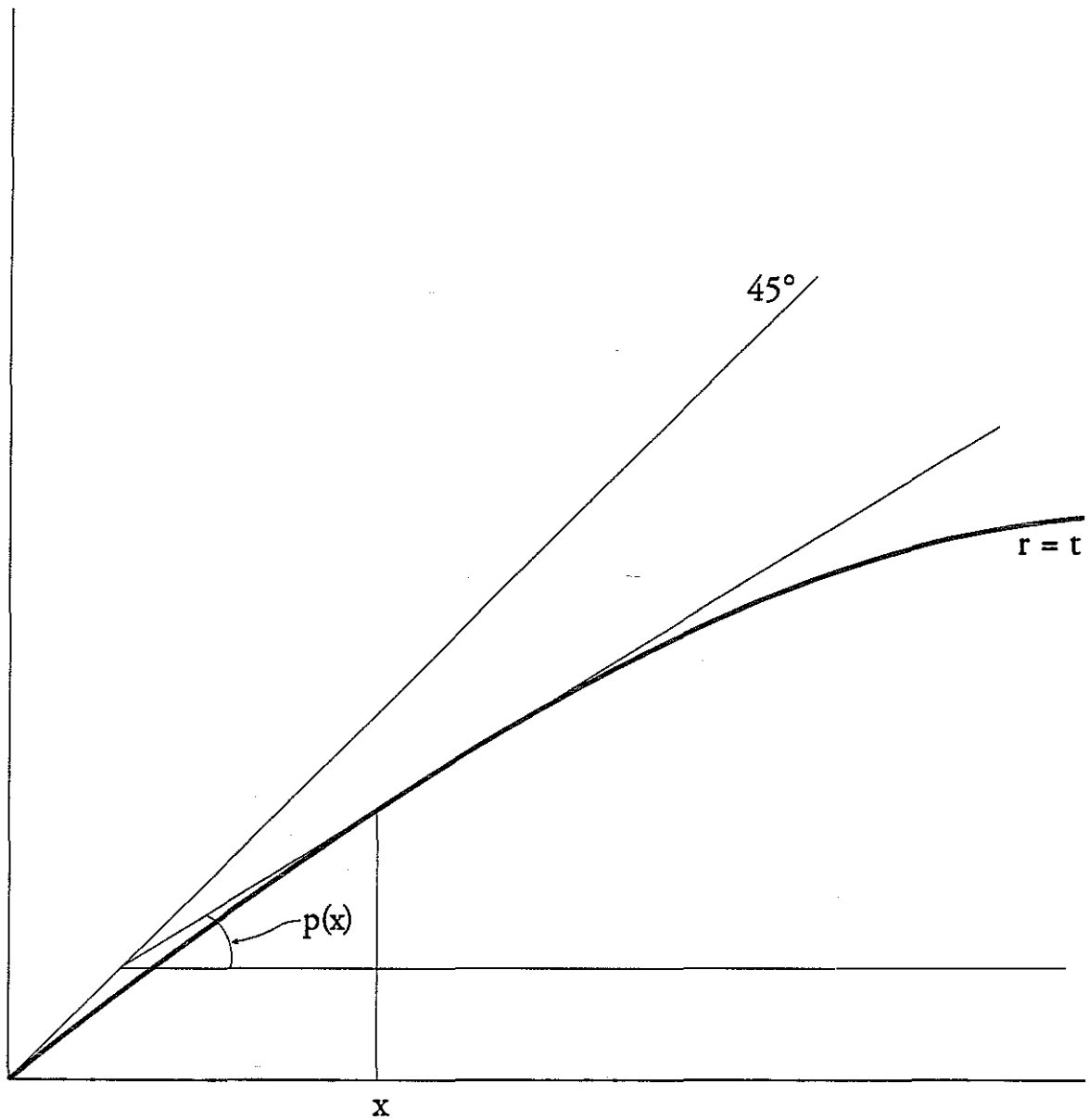

Figure 4

Appendix

We establish the two lemmata that were used in the analysis above.

Lemma 0: Let $r : [0, \bar{y}] \to R$ be a nondecreasing function. Let $r' : [0, \bar{y}] \to R$ be the smallest concave function above r. Then r' is nondecreasing.

Proof: Since r' is concave,

$$r'(x) = \int_o^x \gamma(z)dz \tag{26}$$

for a nonincreasing function γ. In fact, $\gamma(z) = Dr'(z)$–the derivative of r' at z–whenever the derivative exists. Suppose for $x_1 < x_2, r'(x_1) > r'(x_2)$. Then, using (26) (i.e., concavity), it follows that $r'(y) \leq r'(x_2)$ for all $y \geq x_2$ because $\gamma(x_2) \leq 0$ implies $\gamma(y) \leq 0$ for all $y > x_2$ and $r'(y) \leq r'(x_2)$ by (26). Now let $\delta = r'(x_2)$. We claim $r(z) \leq \delta$ for all z: for $z \leq x_2, r(z) \leq r(x_2) \leq r'(x_2) = \delta$; and for $z > x_2, r(z) \leq r'(z) \leq r'(x_2) = \delta$. Hence, the function k defined as $k(x) \equiv \inf \{r'(x), \delta\}$ is concave and such that $r \leq k \leq r'$ with $k \neq r'$. This contradicts the fact that r' is the smallest concave function above r. This proves the lemma.

Proof of Lemma 4: We first show that without loss of generality we can take t to be nondecreasing. Let $y, z \in [0, \bar{y}]$ be such that $y < z$ and $t(y) > t(z)$. Then, since $f = f_t$, no agent will report income to be y. Suppose the tax and audit probability corresponding to y are changed such that they are equal to those corresponding to z. Such a change in the tax and audit probability corresponding to y will not induce agents above y to change their optimal reports, and will make agent y and agents below y indifferent between y and z. Since the tax and audit probability at y are now the same as those at z, this will neither affect the payments of the agents nor the audit costs. Thus, we can take t to be nondecreasing.

Given $(t, p, f) \in R$ and t nondecreasing, we define a scheme (t', p', f') from (t, p, f) as follows: $t'(y) = t(\alpha(y)) + p(\alpha(y))(y - \alpha(y)), p'(y) = p(\alpha(y))$, and $f' = f_{t'}$, where $\alpha(y)(\leq y)$ is an optimal report of agent y.

We first show that t' is nondecreasing. By definition of $t', t'(z) = t(\alpha(z) + p(\alpha(z))(z - \alpha(z))$ and $t'(y) = t(\alpha(y)) + p(\alpha(y))(y - \alpha(y))$. Suppose $z > y$. Then, (a) if $\alpha(z) \geq y$, then clearly $t'(z) \geq t(\alpha(z)) \geq t(y) \geq t(\alpha(y)) + p(\alpha(y))(y - \alpha(y))$ (by optimality of $\alpha(y)$ and that t is nondecreasing); and (b) if $\alpha(z) < y$, then $t(\alpha(z)) + p(\alpha(z))(z - \alpha(z)) > t(\alpha(z)) + p(\alpha(z))(y - \alpha(z)) \geq t(\alpha(y)) + p(\alpha(y))(y - \alpha(y))$ (from the fact that $\alpha(y)$ is optimal for agent y). This proves that t' is nondecreasing.

Clearly, $0 \leq t'(y) \leq t(y) \leq y$ and $0 \leq p'(y) \leq 1$ for each y. Thus, $(t', p', f') \in R$. We show that (t', p', f') is incentive compatible. Since t' is nondecreasing (as shown), we need to show that for each $y, t'(y) \leq t'(x) + p'(x)(y - x)$ for all $x < y$. That is, for each

y and $x < y, t(\alpha(y)) + p(\alpha(y))(y - \alpha(y)) \leq t'(x) + p'(x)(y - x) = t(\alpha(x)) + p(\alpha(x))(x - \alpha(x)) + p(\alpha(x))(y - x) = t(\alpha(x)) + p(\alpha(x))(y - \alpha(x))$. This is indeed true, since $\alpha(y)$ is optimal for agent y, and $\alpha(x) \leq x$.

This proves that $(t', p', f') \in R$ is an incentive compatible scheme. It is easily seen that (t', p', f') and (t, p, f) are equivalent from the point of view of the principle and each agent.

References

Allingham, M. and Sandmo, A. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 1 (1972):323-338.

Becker, Gary. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (1968):169-217.

Border, K. and Sobel, J. "Samurai Accountant: A Theory of Auditing and Plunder." *Review of Economic Studies* 54 (1987):525-540.

Chander, Parkash and Wilde, Louis. "Corruption in Tax Administration." *Journal of Public Economics* 49 (1992):33-349.

Chander, Parkash and Wilde, Louis. "A General Characterization of Optimal Income Taxation and Enforcement." Caltech Social Science Working Paper 796, March 1992.

Cowell, Frank A. *Cheating the Government: The Economics of Evasion.* Massachusetts: The MIT Press, 1990.

Cremer, H.; Marchand, M.; and Pestieau, P. "Evading, Auditing and Taxing: The Equity-Compliance Tradeoff." *Journal of Public Economics* 43 (1990):67-92.

Cremer, H. and Gahvari, F. "Tax Evasion and the General Income Tax," mimeo, presented at The 49th Congress of International Institute of Public Finance, August 1993.

Dye, R. "Optimal Monitoring Policies in Agencies." *Rand Journal of Economics* Autumn (1986):339-350.

Harris, M. and Townsend, R. "Resource Allocation Under Asymmetric Information." *Econometrica* 49 (1981):33-64.

Krasa, S. and Villamil, A. "Optimal Multilateral Contracts." Forthcoming in *Economic Theory* 1993.

Maskin, E. and Riley, J. "Optimal Auctions with Risk Averse Buyers." *Econometrica* 52 (1984):1473-1518.

Mirrlees, J. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies* 38 (1971):175-208.

Melamud, N. and Mookherjee, D. "Delegation as Commitment: The Case of Income Tax Audits." *Rand Journal of Economics* Summer (1989):139-163.

Mookherjee, D. and Png, I. "Optimal Auditing, Insurance and Redistribution." *Quarterly Journal of Economics* 104 (1989):399-415.

Morton, Sanford. "Strategic Auditing for Fraud." mimeo, Pennsylvania State University, November 1988.

Myerson, Roger B. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47 (1979):61-73.

Myerson, Roger B. "Optimal Auction Design." *Mathematics of Operations Research* 6 (1981):58-73.

Myerson, Roger B. and Satterthwaite, Mark A. "Efficient Mechanisms for Bilateral Trading." *Journal of Economic Theory* 29 (1983):265-281.

Ortuno-Ortin, I. and Roemer, J. "Implementation with Inspection." Working Paper, Department of Economics, University of California Davis (1990).

Reinganum, J. F. and Wilde, L. L. "Income Tax Compliance in a Principal-Agent Framework." *Journal of Public Economics* 26 (1985):1-18.

Sanchez, Isabel and Sobel, Joel. "Hierarchical Design and Enforcement of Income Tax Policies." *Journal of Public Economics* 50 (1993):345-369.

Sandmo, A. "Income Tax Evasion, Labor Supply and the Equity-efficiency Tradeoff." *Journal of Public Economics* 16 (1981):265-288.

Scotchmer, S. "The Regressive Bias in Tax Enforcement: A Note." Mimeo. 1988.

Srinivasan, T. N. "Tax Evasion: A Model." *Journal of Public Economics* 2 (1973):339-346.

Stigler, George. "The optimum Enforcement of Laws." *Journal of Political Economy* 78 (1970):526-536.

Stiglitz, J. "Utilitarianism and Horizontal Equity: The Case for Random Taxation." *Journal of Public Economics* 18 (1982):1-33.

Stiglitz, J. "Pareto Efficient and Optimal Taxation and the New Welfare Economics." Handbook of Public Economics, Vol II, (eds.) A. J. Auerbach and M. Feldstein, (1987), Elsevier Science Publishers, B.V. North Holland.

Townsend, R. M. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory* 21 (1979):265-293.