Optimal non-linear income taxation when income from elastic labour supply is costly to monitor*

Fred Schroyen

Report 94/312
October 1994

First version: February 17, 1993
This version: September 15, 1994

* This paper is a revision of the third essay of my PhD thesis (Schroyen, 1994). I am grateful to Maurice Marchand and Erik Schokkaert, for comments and long discussions, and to Agnar Sandmo for sending his remarks on a previous version of the paper. The paper has also benefited from comments of participants at the 1994 Royal Economic Society meeting (Exeter) and at ESEM 94 (Maastricht). The responsibility for remaining flaws is entirely mine.

* SESO-UFSIA, University of Antwerp, Prinsstraat 13, B-2000 Antwerp, Belgium.
(tel: 32-3-220 40 50, fax: 32-3-220 40 26, e-mail: dse.schroyen.f@alpha.ufsia.ac.be)

Universitaire Faculteiten St.-Ignatius
Prinsstraat 13 - B 2000 Antwerpen
D/1994/1169/29
Abstract

The paper assumes a two-class economy with an official and black labour market. Labour supply is not inelastic. Official income is taxed nonlinearly, while unofficial income is only observable after a costly audit upon which it is taxed at an exogenous penalty rate. After characterization of the Pareto efficient audit and tax policies, it is verified how these policies react to a lowering of the penalty rate. The effect on the marginal tax and the audit rate is shown to decompose into a deadweight loss and substitution (or iso-deadweight loss) effect. This exercise informs about the way the government should adjusts its audit and tax policy when evasion opportunities in the economy grow. A numerical example illustrates the relation between first, second and third best frontiers.

JEL classification: H21, H26, K42.

Key words: redistributive taxation, tax evasion
1. **Introduction.**

Since Mirrlees (1971), the standard literature on optimal income taxation has focused on the problem of redistributing wealth among people when the personal characteristics which provide people with economic opportunities are unobservable to the government. This is a second-best problem because the inability to observe characteristics generally precludes the use of non-distortionary taxation. Rather, taxes are levied on endogenous variables, like income, which are thought of as perfectly and costlessly observable. But in recent years, this 'cheap observability' assumption has been questioned more and more. High levels of tax evasion, together with considerable budgets allocated to the Inland Revenue Service indicate that the second-best problem is a hypothetical one. This seriously undermines optimal tax theory as a reliable guide to action and makes it susceptible to similar criticisms of delusions as first-best theory. The purpose of this paper is then two-fold. First I want to inquire about the nature of the optimal tax schedule when labour income can only be perfectly observed through a random, but costly, audit technology. Second, I want to relate the optimal marginal tax rate to its counterpart in standard income tax models where observation of income is costless. In this way I hope to provide more structure to the debate on the best way to cope with evasion opportunities in a welfare state economy.

The model I shall use is an immediate extension of the one presented by Sandmo (1981). The economy is populated by two classes of agents who differ from one another by their productive ability. Like in Sandmo (1981), I assume that labour supply is not totally inelastic, and that only one class of agents has access to a black labour market where they can escape their tax liabilities; I will assume this to be the class of high ability agents. The main difference, however, is that in this paper the tax schedule is not restricted to linearity; in this respect, the paper is also an extension of Stiglitz (1982). Besides being more general, this has the desirable by-product that the behavioural responses of a tax evader are easier to sign. A second implication of the non-linearity which distinguishes both models, is that tax evasion, and therefore the conviction and sanctioning of evaders, occurs only out of equilibrium under an optimal policy. Along the equilibrium path, horizontal equity therefore obtains, identical tax payers being treated identically both ex ante and ex post (the ex post differential treatment by the random audit policy being confined to an out-of-equilibrium event). On the other hand, since increasing the penalty on evading activities will entail a social benefit at zero social cost, penalties on evading activities cannot be fully endogenized, for otherwise, enforcement of the tax code could be trivially
achieved by setting the penalty high enough (possibly by introducing capital punishment). To preclude such a trivial outcome, I shall assume that the penalty rates inflicted on evaders fall outside the discretion of the government.¹

The feature that tax evasion only occurs out of equilibrium is a common result in the costly-state-verification approach to tax evasion, used e.g. by Reinganum & Wilde (1985), Mookherjee & Png (1990), Border & Sobel (1987). All these papers take off from the assumption that individual labour supply is totally inelastic, implying that the distribution of gross incomes in society is fixed. A fixed gross income distribution also features in the model of Cremer, Marchand & Pestieau (1990), although they restrict the income tax schedule to linearity.

An exception to this approach is the work by Mookherjee & Png (1989) who introduce moral hazard. In their model ex ante all agents are alike. A finite number of income realizations are possible. The probability with which an agent will receive a particular income depends on the (unobservable) effort the agent undertakes. This effort level is chosen optimally, taking the tax, audit and penalty policy as given. Because all agents are ex ante alike, they all choose the same effort level, and by the law of large numbers, the relative number of people in each income category coincides with the prior probabilities evaluated at the optimal effort level. Ex post, an agent belonging to income category i can only pretend to belong to category j by misrepresenting his tax file. The sequence of events is as in figure 1a.

In contrast, in the present model, the sequence of events is partially reversed (see figure 1b): at the initial stage of the game, nature determines the productive ability of the agent who is privately informed about it; subsequently, the agent determines his total income by his (unobservable) effort level. Being endowed with a high productivity, an agent may report a low income on his tax file, either by refraining from working hard and reporting truthfully, or by earning a higher total income which he subsequently underreports, or a combination of both.²

— insert figure 1 here —

¹ Some authors follow a slightly different approach by assuming that even an out-of-equilibrium evader should be guaranteed a minimal living-standard. A recent contribution which tries to explain why we do not observe capital punishment of tax evaders is provided by Pestieau et al (1992).

² Related to this, is the informativeness of the audit. In Mookherjee & Png (1989) the audit reveals the characteristic which distinguishes one agent from another. On the other hand, in the model of this paper, an audit provides (perfect) information on an endogenous variable (income) which is only imperfectly related to the intrinsic characteristic of an agent.
In this paper, I show that as far as its structure is concerned, the optimal tax schedule coincides with the one derived by Stiglitz (1982): the no-distortion-at-the-top result carries over, and the expression for the optimal marginal tax rate on low incomes does not exhibit a correction term accounting for the distortion of the black labour market—unlike in Sandmo’s (1981) linear income tax model. Moreover, auditing should be confined to low incomes. Second, I investigate how reforms in the (exogenous) penalty system affect the optimal fiscal policy of the government. I show that these effects decompose into a deadweight loss effect (an adverse government revenue effect) and a substitution effect. These effects are signed, and it is argued that both the government’s urge to redistribute as well as the sensitivity of both the relative efficiency and relative cost of its instruments will determine to which extent more intensive use should be made of them. Because the Stiglitz model is encompassed by the present one for a sufficiently severe penalty system, this exercise informs about the way tax evasion opportunities affect the optimal tax policy of the government, compared to situations where such opportunities do not exist.

The paper is organized as follows. Section 2 describes production possibilities, preferences and the working of the labour markets. In section 3, the information set of the government is defined, as well as the audit technology and the penalty system. Next, the set of informationally and budgetary feasible policies is delineated, and the problem of the planner is formulated. In section 4, I derive black market behaviour and give a first impression on the effects of such behaviour on the incentive compatibility of a fiscal policy. Section 5 characterizes Pareto efficient tax and audit policy when evasion opportunities are present. In sections 6 and 7, I look at the comparative statics effects of a reform of the penalty system on the optimal policy. This sheds a light on the relation between a second and third best solution, and in particular on the effect of tax evasion opportunities on the optimal level of the marginal tax rate. Section 8 provides a numerical example, and concluding remarks are collected in section 9.

2. Production possibilities, preferences, and labour markets.

The economy is composed of a large number of individuals which are divided into two classes of equal size according to their productive ability. Agents of type 1 have an ability to supply \( w_1 \) efficiency units of labour during one hour; for type 2 agents, this amounts to \( w_2 (> w_1) \) units. The production sector
is competitive and transforms efficiency units of labour into a consumption good at a rate which is fixed and normalized to unity. Also the price of the consumption good is normalized to unity, so that the productive abilities can readily be interpreted as real wage rates.

Both types of agents have an initial time endowment equal to unity, and are assumed to have similar preferences over consumption \((c)\) and leisure \((\ell)\), as well as over any lotteries among consumption-leisure bundles. These preferences are representable by the von Neumann-Morgenstern utility function \(u(c, \ell)\) which is continuous and continuously differentiable at least two times. Furthermore, this utility function shares the following properties: (U1) monotonicity (U2) concavity, and (U3) Edgeworth complementarity \((u_{c\ell} > 0)\). In this cardinal context, these three assumptions together imply both normality of leisure \((\partial(u_i/u_j)/\partial c > 0)\) and consumption \((\partial(u_i/u_j)/\partial \ell > 0)\). Furthermore, I shall assume that (U4) the standard labour supply schedule implied by the preference ordering is forward bending.

There are two labour markets in the economy: an official market where the wage bill an agent receives is recorded and reported to the government by the employer, and a black labour market where such reporting behaviour does not exist. In both cases, the employer makes use of the same production technology. When an agent of type \(i\) (agent \(i\), for short) earns an income of \(Y\) on the official labour market, he spends \(Y/w_i\) hours on that labour market. This income may be complemented by unofficial earnings. These are denoted by \(X\), meaning that another \(X/w_i\) hours are supplied on the black labour market.

As in Sandmo's (1981) model, I assume that only one class of agents have access to the black labour market. In particular, I assume this to be the class of high ability agents. Low ability agents do not participate on the black labour market, for instance because they lack information about such evasion opportunities, or because they do not wish to take advantage of them, like the habitual compliers in Graetz et al (1986).

3. Informational and budgetary feasibility of fiscal policies.

In this economy, the government (later also referred to as the planner) is imperfectly informed, in that it knows the distribution of agents over types but is unable to assess who is of which type. On the other hand, the government can observe gross income. In the standard income taxation models, this observation is costless. In the present model, however, the government can only observe officially earned income at zero cost. Income which is earned on the black labour
market may be perfectly assessed, but only after the conduct of a costly audit. The audit policy is taken be of the random type, to be error-proof, and to be conditional on observable officially earned income. That is, a fraction of the population is paid a visit by the tax inspector who is able to assess without error the amount of illegally earned income. It is fair to assume that the tax inspector reads first the official income report, before deciding whom to pay a visit.

Formally, upon receiving an official income report of $Y$ by an agent, the planner subjects the agent to a tax treatment, leaving him with an official net income of $Z(Y)$. To the extent that this tax treatment distorts the decision of the agent, efficiency losses occur. These losses constitute the social cost of the tax treatment, as will be made precise below in section 5. On the basis of the official income report $Y$, the planner in addition decides on further investigation of this agent with probability $0 \leq \Pi(Y) \leq 1$. Let me refer to this as the audit treatment of the income report. Once audited, an agent who is found having earned $X$ on the black labour market, is inflicted a penalty amounting to $fX$, where the marginal (and average) penalty rate $f$ is stated in the country's criminal law, and is exogenous to the planner. By the law of large numbers, the probability with which an income report $Y$ is audited, corresponds to the fraction of $Y$-reporters actually audited. This auditing is costly, and I shall assume that the amount of resources drawn away from the rest of the economy, when auditing income group $Y$, is given by the function $K(\cdot)$ which has the following properties for some small, but strictly positive level $\pi^\circ$: $K(\pi) = 0$, $K'(\pi) > 0$ and $K''(\pi) > 0$, (A3) $\lim_{\pi \rightarrow 0} K(\pi) = \infty$, (A4) total audit costs are given by $\Sigma Y K(\Pi(Y))$. In other words, the audit technology is strictly separable, strictly increasing and strictly convex for audit rates above $\pi^\circ$. Up to this level, auditing is free. Finally, auditing the entire class of $Y$-earners is prohibitively costly. As will be made precise below, the possibility to carry out some costless auditing will ensure the existence of a finite marginal penalty rate, the combination of which deters people sufficiently so as not to participate on the black market at all. However, without any problem the level $\pi^\circ$ can be very small.

Suppose now that an agent of ability $w_3$ has chosen to earn $Y$ and $X$ on the official and black labour market, respectively. The expected utility provided by these choices is then given by $E_{\[Y,Y\]_r}\{Z(Y) + (1-\tau)X, 1-(Y+X)/w_3\}$, where $\tau$ is a stochastic penalty rate taking on the values 0 and $f$ with probability $1-\Pi(Y)$ and $\Pi(Y)$, respectively. For a given choice of $Y$, the optimal choice of $X$ is given by
\[ X_i[Y,Z(Y),\Pi(Y), f] = \arg\max_{X \leq 0} E_{\Pi(Y)} u[Z(Y)+1-(1-\tau)X, 1-\frac{Y+X}{w_2}] \quad (3.1) \]

Let \( v^i[Y,Z(Y),\Pi(Y), f] \) denote the associated value of the expected utility function. Similarly, let \( v^1[Y,Z(Y),\Pi(Y), f] \) denote the utility of a type 1 agent who earns \( Y \) franc officially (and whose choice of \( X \) is by definition constrained to 0, i.e. \( X_1 = 0 \) and \( v^1[Y,Z(Y),\Pi(Y), f] = u(Z(Y), 1-Y/w_i) \)).

All agents of type \( i \) will now choose the official income \( Y \) which maximizes the utility function \( v^i[Y,Z(Y),\Pi(Y), f] \). Because there are only two types of agents, at most two values of official income will be encountered in equilibrium; these optimal values will depend on the ability of the agents and are denoted as \( Y_i \) for short \( (i=1,2) \). Thus, when the corresponding net income levels and audit probabilities are \( z_i = Z(Y_i) \) and \( \pi_i = \Pi(Y_i) \), respectively, the following set of weak inequalities immediately follows:

\[ v^i[Y_i,z_i,\pi_i, f] \geq v^j[Y_j,z_j,\pi_j, f], \quad i,j=1,2. \quad (3.2) \]

Since the net revenue the government collects from class \( i \) is given by \( \rho_i = Y_i z_i + \pi_i f X_i K(\pi_i) \) \( (i=1,2) \), the fiscal policy will then be budgetary feasible when total revenues cover an (exogenous) amount of public expenditure \( R \), viz \( \rho_1 + \rho_2 \geq R \). Alternatively, one can think of \( R \) as the fixed cost of the audit technology.

The fiscal policy induces in effect an allocation mechanism which maps truthful announcements of ability levels into levels of gross and net official earnings and of audit probability. By the revelation principle, we may reason in reverse order as well: suppose we start from a budgetary feasible allocation mechanism, then there exists a budgetary feasible fiscal policy to which this mechanism corresponds provided the mechanism respects the above set of weak inequalities (3.2), that is, provided the agents are incited to make an honest announcement of their ability level.\(^3\) Whence, if the planner is interested in looking for the optimal fiscal system, she can do so by searching for the optimal budgetary and informationally feasible allocation mechanism. Since the latter problem is easier, the optimal implementing fiscal system may be derived from it \( \text{ex post} \). But before analyzing the planner’s problem, I shall first explain why the solution to the standard optimal taxation problem is no longer incentive compatible and how this compatibility can be restored.

---

\(^3\) Refer e.g. to Hammond (1979, Theorem 6) or Guesnerie (1981, Theorems 1 and 2).
4. Evading behaviour producing incentive incompatible tax policies.

Suppose that a type 2 agent is allocated by the mechanism the treatment \((z, \pi, Y)\) after having announced (possibly in a false way) his type. The optimal choice of black labour market earnings, \(X(z, Y, \pi, f)\), necessarily satisfies the conditions \(Eu_t(1-\tau) - Eu_t/w_2 \leq 0\) and \([Eu_t(1-\tau) - Eu_t/w_2] \cdot X = 0\). Whence, this agent will find it lucrative to participate on the black labour market if and only if \(MRS^2 \mid_{x = 0} < 1 - \pi f\) where \(MRS^2\) denotes his marginal rate of substitution in the \((z, Y)\)-space, i.e.

\[
MRS^2 \overset{\text{def}}{=} \left. \frac{dz}{dY} \right|_{dEu^2, 0} = \frac{Eu_t^2}{w_2 Eu_e^2}.
\]

(4.1)

For an interior solution, it may be verified that the following expressions describe the comparative statics effects (the signing of which will be motivated below):

\[
\frac{\partial X}{\partial z} = \frac{1}{\Delta} \left[ \frac{1}{w_2} Eu^2_e - Eu^2_{ce}(1-\tau) \right] - \frac{1}{\Delta} \frac{\partial Eu^2_t}{\partial X} < 0,
\]

(4.2)

\[
\frac{\partial X}{\partial Y} = \frac{1}{\Delta} \frac{1}{w_2} (Eu^2_e(1-\tau) - \frac{1}{w_2} Eu^2_t) - \frac{1}{\Delta} \frac{1}{w_2} \frac{\partial Eu^2_t}{\partial X} < 0,
\]

(4.3)

\[
\frac{\partial X}{\partial \pi} = \frac{1}{\Delta} \left( Du^2_e(1-\tau) - \frac{1}{w_2} Du^2_t \right) - \frac{1}{\Delta} \frac{\partial Du^2_t}{\partial X} < 0,
\]

(4.4)

\[
\frac{\partial X}{\partial f} = \frac{1}{\Delta} \left[ \frac{\partial^2}{w_2^2} - \frac{1}{w_2} \frac{\partial^2}{u^2_e} - u^2_{ce}(1-f)X \right] \pi < 0.
\]

(4.5)

Here, \(u\) (\(\bar{u}\)) and derivatives denote the utility level and marginal utilities evaluated in the audit (non-audit) state, \(Dx\) is a shorthand for \(x - \bar{x}\), and \(\Delta\) is a negative expression in view of the second order condition for an interior solution.

When leisure (consumption) is a normal good in both states of nature, the first (second) expression above is negative; when in the standard labour supply model under certainty, the labour supply curve is forward (backward) bending, the last two expressions are negative (positive).\(^4\)

\(^4\) The proofs of these statements go along exactly the same lines as in Cowell (1986, pp 30-31). The comparative statics can be easily signed because in a discrete class economy, a non-
The black market earnings function $X(\cdot)$ can be plugged into the expected utility function to produce the indirect utility function in terms of the observables $z, Y, \pi$ and $f: v^*(Y, z, \pi, f)$. By the envelope theorem, the slope of the corresponding indifference curve in the $(z, Y)$-space is still given by expression (4.1). But compared to $\text{MRS}^2 \mid_{X=0}$, this slope will take on a higher value since black market participation both increases consumption opportunities and reduces the availability of leisure, and under the normality assumptions, this increases the willingness to substitute consumption for leisure.⁵

In figure 2, I have drawn a typical solution to the optimal income tax problem for a two class economy with absence of black labour markets and when redistribution takes place from the high to the low ability agent (as, e.g., described by the Stiglitz (1982) model). The thin solid curves depict the indifference fields of the two types. The field for the high ability agent is complemented with two expansion paths corresponding to the slopes 1 and $1-\pi f$; these paths connect all bundles the indifference curves through which exhibit the same slope.⁶ The high ability agent receives the undistorted bundle $A=(z_2^*, Y_2^*)$ and does not find this bundle strictly worse than bundle $C=(z_1^*, Y_1^*)$ designed for the type 1 agent. Without tax evasion possibilities, this allocation is incentive compatible and budgetary feasible. Now imagine that the black labour market opens, and that the audit and penalty policy of the government leads to an expected marginal penalty of $\pi f$. In this new situation, how will agent 2 compare bundles $C$ and $A$? To answer this question, it suffices to draw in agent 2's indifference field conditional on his labour supply on the black market being optimally chosen. These are the bold dashed lines. For allocations North-East of the $(1-\pi f)$-expansion path nothing changes since the agent does not wish to participate on the black labour market. On the other hand, at any bundle below

linear tax schedule has the effect of imposing a rationing constraint on the consumer's official labour supply (which explains the congruence with Cowell's analysis of black market labour supply when hours restrictions apply on the official labour market).

⁵ Imagine that for some reason a quantity constraint applies on the black labour market and that $\bar{X}$ denotes the maximal amount of income that an agent could earn on that market. The agent then needs to solve again problem (3.1) but now subject to the additional constraint that $X \leq \bar{X}$. Let me now move $\bar{X}$ from 0 to $X(z, Y, \pi, f)$ (assuming that this latter figure is strictly positive). Clearly, the optimal choice of $X \ w.r.t.$ this new problem will follow $\bar{X}$. How will the marginal rate of substitution in the $(z, Y)$ space evolve for a given bundle $(z, Y, \pi, f)$? From (4.1), (4.2) and (4.3) it may be established that $\partial \text{MRS}^2 / \partial X = (\Delta / \Delta z)[(\partial X / \partial z) \cdot \text{MRS}^2 + \partial X / \partial Y] > 0$ where the inequality sign follows from the normality assumptions. Therefore it may be concluded that the $\text{MRS}^2$ is raised by the opportunity to evade taxes when this opportunity is taken advantage of.

⁶ Under normality of consumption and leisure such expansion paths in the $(z, Y)$ space are downward sloping.
this path, and in particular at bundle $C$, there is an incentive to supply labour the black labour market; this raises the agent's (expected) utility level (to level $u^*_i$, say) and his willingness to transform gross into net income. It then becomes immediately apparent that the original redistribution scheme is no longer incentive compatible for agent 2 will now strictly prefer $C$ to $A$. How could incentive compatibility be restored? Ignoring for the moment the implications on the government's budget constraint, the following options are open:

- either the planner could further increase the marginal tax rate on the low ability agent (moving the bundle designed for the low ability agent further South-West along $\bar{a}^1$, at least as far as $C'$);
- or it could increase the audit probability rate sufficiently to make $C$ no longer desirable from agent 2's point of view (which has the graphical effect that the expansion path below which evasion takes place is moved at least as far as $C'$);
- or a combination of both.

— insert figure 2 here —

The optimal policy response will of course also need to satisfy budgetary feasibility. Its characterization will be the subject of the remainder of this paper. In particular I will try to answer the question whether tax evasion possibilities give rise to an optimal marginal tax rate which is higher than the one prescribed by the standard Stiglitz (1982) model.

To restore incentive compatibility, there is a third option as well: shifting the $(1-\pi_f)$-expansion path sufficiently downwards by increasing the penalty rate. Although I have ruled out the possibility for the planner to change this penalty rate, the fictitious idea of moving $f$ will be helpful later on. In particular, let $f^0$ denote the smallest penalty rate necessary to induce full compliance at the bundle prescribed by the standard Stiglitz model for agent 1, when the audit frequency is $\pi^0$, i.e. $f^0 = \min\{f|X(z^1, y^1, \pi^0, f) = 0\}$. Then, with the penalty rate set as high as $f^0$, the solution to the Stiglitz model also constitutes the optimal policy for the present model. By lowering the marginal penalty rate from this full compliance level to the level given by law, it is possible to trace out how the optimal policy adjusts. In particular, this reform exercise will enable me to identify the motivations for adjustments in optimal marginal tax rate. This will be the subject of sections 6 and 7. In the next section, I will characterize the optimal policy for less extreme values of the penalty rate. Before doing so, one technicality needs to be sorted out.

Observe that I have drawn in figure 2 the dashed curve crossing the $\bar{a}^1$-curve
only once. In the standard model of optimal income taxation, the Spence-Mirrlees condition imposes that at any \((z, y)\) bundle the indifference curve of the low ability agent is always steeper than the indifference curve of a highly able individual. It is well known that normality of consumption is a sufficient condition for this feature to obtain. When there is scope for tax evasion, however, normality of consumption alone is no longer sufficient. The reason is that the high ability agent will have opportunities to participate on the black labour market, and as I have just argued above, choosing a positive level of \(X\) raises the \(MRS\) in the \((z, y)\)-space when leisure and consumption are normal goods. In the sequel of the paper, I will therefore have recourse to the following assumption:⁷

(SCP) The Single Crossing Property: At any point in the \((z, y)\)-space, the inequality \(MRS^1(z, y) > MRS^2(z, y)\) is verified.

The advantage of invoking SCP is that I can safely ignore the self-selection constraint for the low ability type in the planning problem.

5. Pareto efficient taxation and auditing.

In this section I look for the structure of tax and audit policies which maximize the welfare of the high ability agents, but at the same time guarantee the low ability persons a standard of living \(\bar{u}^1\) which is higher then the utility they would obtain under laissez faire. The programme which the planner seeks to solve may then be formulated as:

\[
\begin{align*}
\max_{(z_2, z_1, \pi_2, \pi_1, f)} & \quad v^2(Y_2, z_2, \pi_2, f) \\
\text{s.t.} & \quad u(z_1, 1 - \frac{Y_1}{w}) \geq \bar{u}_1 \quad (\mu) \\
& \quad v^2(Y_2, z_2, \pi_2, f) \geq v^2(Y_1, z_1, \pi_1, f) \quad (\lambda) \\
& \quad Y_1 - z_1 + Y_2 - z_2 + \pi_2 f X_2 \geq K(\pi_1) + K(\pi_2) + R \quad (\gamma)
\end{align*}
\]

The first constraint \((\mu)\) guarantees agent 1 a utility level of \(\bar{u}^1\); it will be binding under monotonicity \((U1)\). The self-selection constraint \((\lambda)\) prevents agent 2 to disassemble as a low ability type, while the budget constraint \((\gamma)\) imposes that net

⁷ In appendix A, the possibility of black labour market supply to produce double-crossing is related to the \(z\)-compensated elasticity of the mimicking high ability agent's black labour supply \(X(z, Y, \pi, f)\) w.r.t. \(Y\). Under the normality assumptions, this elasticity is negative and single crossing will obtain when it is not too negative.
tax receipts and fines collected in equilibrium should cover total audit costs and an amount \( R \) of government consumption. Both the genuine and mimicking agent 2 take an optimal decision w.r.t. black labour market earnings, as is apparent from the use of the indirect utility function \( v^2(\cdot) \).

The following lemma characterizes the Pareto efficient fiscal treatment of high ability agents. Essentially, it says that either the tax treatment of type 2 agents should be undistorted (in which case they will have no incentive to participate on the black labour market), or, if distorted, these types of agents should be provided with maximal incentives (in the sense of not bearing any audit risk) to remove this distortion by black labour market participation. This result is an extension of the no-distortion-at-the-top result of Stiglitz (1982) to the case of costly monitoring: the effective marginal tax rate on top incomes should be zero.

**Lemma 1:** The planner is indifferent between fiscal treatments \((z_2, \pi_2, Y_2)\) that satisfy any of the following combinations:

\[
MRS^z_{x_2=0} < 1 \land \pi_2 = 0 , \text{ or } MRS^z_{x_2=0} = 1 \land \pi_2 \in [0, \pi^o] .
\] (5.1)

All other combinations are Pareto inferior.

The intuition for this result goes as follows.\(^8\) Since the low ability type will never have an incentive to dissemble as a high ability agent,\(^9\) a small reform of the fiscal treatment \((z_2, \pi_2, Y_2)\) will only have repercussions for the government budget constraint and the high ability agents welfare level in equilibrium. Suppose now that the planner considers a fiscal treatment which does not satisfy the properties in the lemma. In that case she could extract the same amount of tax revenue from the high ability agent, while at the same time improving this agent’s utility level and/or saving on audit costs. Whence the original fiscal treatment could not have been Pareto optimal. The intuition for the indifference result is simple. If, in the \((z, Y)\) space, the planner provides high ability agents with a bundle \((z_2, Y_2)\) on their 1-expansion path, it will collect from them an amount of taxes \(Y_2-z_2\). Now suppose that the planner provides the high ability type an alternative bundle \((z'_2, Y'_2)\) with smaller gross and net income levels but which yields the same amount of taxes \(z'_2 < z_2, Y'_2 < Y_2, Y'_2-z'_2 = Y_2-z_2\). Such a bundle is distorted and will initially provide high ability agents with a lower utility level,

---

\(^8\) The proof of lemma 1 is given in appendix B.

\(^9\) A similar constraint for agents of type 1 has been left out for the reason that by the single crossing property at most one self-selection constraint will be binding and that under redistribution from class 2 to class 1, this will be the constraint \((\lambda)\).
but by refraining from auditing \(Y_2\)-income levels, these agents will make use of the black labour market to move towards the same consumption and leisure allocation which was implied by the original bundle \((z_1, Y_2)\).

Thus, as far as the fiscal treatment of high ability agents is concerned, the planning problem has multiple optima. From now on I will make the following

**Convention:** The planner provides type 2 agents with an undistorted \((z, Y)\) bundle, that is \(\text{MRS}^2|_{x=0} = 1\) and \(\pi_2 = 0\).

Before I will characterize the fiscal treatment of low ability agent, observe that because the living-standard constraint \((\mu)\) is binding, I can 'invert' this agent's utility function to write \(z_1\) as a function of \(Y_1\) and the parameter \(\alpha^l\), viz \(z_1 = F(Y_1, \alpha^l)\). In fact, the graph of \(F(\cdot)\) in the \((z, Y)\)-space is precisely agent 1's indifference curve and it is easy to check that it shares the following properties:

\[
F_z - \text{MRS}^1 > 0, \quad F_{\alpha^l} - \frac{1}{\alpha^l} > 0, \quad F_{Y_1} - \frac{\partial \text{MRS}^1}{\partial Y_1} \bigg|_{\alpha^l = 0} > 0, \quad \text{and} \quad F_{Y_1} - \frac{\partial \text{MRS}^1}{\partial z_1} F_{\alpha^l} > 0, \quad (5.2)
\]

where \(\text{MRS}^1\) is the low ability agent's marginal rate of substitution, i.e.

\[
\text{MRS}^1 = \frac{d_z}{dY_1} = \frac{u^1_{\alpha^l}}{w_1 \mu^1_{\alpha^l}}, \quad (5.3)
\]

and \(\alpha^l\) is a shorthand for his marginal utility of consumption (the positivity of \(F_{\alpha^l}\) follows from the normality of leisure). Below, I shall use a hat ("\(^\wedge\)") to remind the reader that the effect on an expression of a change in \(Y_1\) is always consists of a \(z_1\)-compensated change.

The planner's problem has now been considerably simplified. She essentially operates two instruments, \(Y_1\) and \(\pi_i\). The choice of \(Y_1\) is tantamount to a choice of the marginal tax rate on low income levels because lowering \(Y_1\) along the indifference curve \(\alpha^l\) is the same as increasing the marginal tax rate on low income (which is \(1 - \text{MRS}^1\)). Both instruments have the desirable effect of lowering the expected utility of the mimicking high ability agent, thereby relaxing the self-selection constraint. Using the abridged notations \(v^{(1)}\) for \(v(Y_1, z_1, \pi_1 f)\) and \(\alpha^{(1)}\) for \(\partial v^{(1)}/\partial z_1\) (=\(E_{\mu^1_{(1)}}\)), the marginal benefit of each instrument can be defined as
\begin{align}
\beta_t & \equiv \frac{\partial \phi^{2(1)}}{\partial (-Y_t)} - \alpha^{2(1)} [MRS^1 - MRS^{2(1)}] > 0, \\
\beta_x & \equiv \frac{\partial \nu^{2(1)}}{\partial \pi_1} - Du^{2(1)} - \hat{u}^{2(1)} \hat{u}^{2(1)} \begin{pmatrix} \gamma \\ \gamma \end{pmatrix} 0 \rightarrow X \begin{pmatrix} \gamma \\ \gamma \end{pmatrix} 0, \tag{5.4}
\end{align}

where \( MRS^{2(1)} \) is as in (4.1) with \( z = F(Y_i, \hat{u}^1) \) and \( Y = Y_i \).

On the other hand, the use of both instruments is costly. A higher marginal tax rate on low incomes (a lower \( Y_i \)) means that the transfer made to low ability persons should be higher then the lump sum tranfer which could guarantee those agents \( \hat{u}^1 \) under perfect information (which I denote by \( LSTr \)). Auditing also draws resources away from the rest of the economy. Thus the amount of resources lost due to imperfect information is given by \( \Delta(Y_i, \hat{u}^1, \pi_1) = F(Y_i, \hat{u}^1) - Y_i + K(\pi_1) - LSTr \); I shall call this amount the deadweight loss (\( DWL \)) in the economy. The marginal costs of operating the two instruments are then given by

\begin{align}
\kappa_t & \equiv \frac{\partial \Delta}{\partial (-Y_t)} = 1 - MRS^1 > 0, \\
\kappa_x & \equiv \frac{\partial \Delta}{\partial \pi_1} = K'(\pi_1) > 0 \text{ for } \pi_1 \geq \pi^* \text{ and } 0 \text{ otherwise.} \tag{5.5}
\end{align}

Thus, the planner should select both instruments in such a way that the self-selection constraint is verified but at the same time that the \( DWL \) in the economy is minimized, because this amount of lost resources—in addition to government expenditure under perfect information, \( R + LSTr \)—will have to be financed out of lump sum taxes on the high ability class. As the following lemma shows, a balancing of marginal costs and benefits characterizes the optimal policy:

**Lemma 2:** The Pareto efficient fiscal treatment of low ability agents should satisfy the following policy rules: \( \gamma \kappa_t = \lambda \beta_t \) and \( \gamma \kappa_x \geq \lambda \beta_x \) with strict equality obtaining whenever the optimal audit policy is 'intensive' (i.e. \( \pi_1 > \pi^* \)).

Rearranging the first of these policy rules produces the familiar optimal marginal tax formula, viz

\[ 1 - MRS^1 = \frac{\lambda \alpha^{2(1)}}{\gamma} [MRS^1 - MRS^{2(1)}]. \tag{5.6} \]
Using (SCP) it follows immediately that a strictly positive marginal tax rate on the low ability agent will be part of an optimal policy: \(1-MRS^1 > 0\). Thus, as far as its characterization is concerned, the optimal tax policy is not affected by tax evasion opportunities. The reason is—and this is the main differences w.r.t. Sandmo’s (1981) model—the absence of tax evasion in equilibrium. This can be imputed to the non-linearity of the tax schedule. Any distortion of agent 2's decision to supply labour on the official labour market is wasteful and might create an incentive to participate on the black labour market. The flexibility of a non-linear tax schedule allows the planner to get rid of such distortions. By the same token, the distortion of the black labour market due to the penalty system does not bear directly upon the optimal marginal tax rate on official income. This explains the similarity of eq (5.6) to its counterpart in the Stiglitz model. On the other hand, the optimal tax formula derived by Sandmo (1981, eq 52), consists of the standard linear tax formula plus a correction term added by evasion opportunities, the latter accounting for the fact that the marginal tax rate partially restores the distortion of the black market labour supply due to the penalty system.

Whether the planner should audit at a rate \(\pi^0\) or higher, depends on the marginal audit cost at this rate and on the level of the penalty rate which courts apply. If this rate is equal to or slightly lower than \(f^\circ\), black labour market earnings will be zero or negligible; the marginal benefit of auditing will be likewise and will not make up for the strictly positive marginal cost at \(\pi^0\). On the other hand, if the ruling penalty rate is lenient, the marginal benefit of auditing at a rate \(\pi^0\) will be high, which motivates a more intensive policy up to the level where (socially evaluated) marginal cost and benefit match one another.

Because (5.6) is a characterization result and cannot in general be solved explicitly for the marginal tax rate, an obvious question remains unanswered: Do tax evasion opportunities give rise to larger marginal tax rates? Intuitively, one would say yes. The role of a strictly positive marginal tax rate is to deter mimicking behaviour. If such behaviour is encouraged by access to a black labour market, then would one not expect an even higher tax rate? In the next section, I will attempt to address this question.

6. Pareto efficient policy responses to a reduction in the penalty rate.

To learn the relationship between the optimal marginal tax rate under evasion opportunities and its counterpart in the standard Stiglitz model, I will
exploit the fact that that model is nested in the present one if the penalty rate takes the value $f^0$ or higher. In this case, the planner can enforce truthful income reporting by the mimicker at zero cost [by (A1)] and the optimal tax rate of the Stiglitz model suffices to satisfy the self-selection constraint (but only just). Starting from this solution, we may then lower the penalty rate from its fictitious level $f^0$ to the actual level given by law, and record how optimal tax and audit rates adjust.

In this section I will describe how such adjustment rules should look like when the government is committed to guaranteeing low ability agents a living standard $u^l$, even when a more lenient penalty system results in a more costly redistribution policy; in other words I shall look into Pareto efficient adjustment rules. However, when the government sets its policy according to a Bergsonian welfare function, the increased cost of redistributing income is more likely to be shared by both classes, rather than being borne entirely by the high ability class. In the next section, I shall develop Bergsonian adjustment rules which take such cost sharing into account.

As the optimality rule in lemma 2 made clear, when carrying out this thought experiment of lowering the penalty rate from the level $f^0$, it will be necessary to distinguish between two regimes. Initially, when the penalty rate $f$ is equal or slightly below $f^0$, the net relative marginal cost of the audit rate is negative and the optimal audit policy is 'passive' (regime I: $\gamma_{h_l} > \lambda \beta_\gamma$ & $\pi_l = \pi^o$). Any deterrence of mimicking behaviour will then be taken care of by the marginal tax rate on low incomes. Indeed, within this regime, for each value of $f$ there will be just one value of $Y_l$ which equates the utility level of the high ability type with a lump sum tax liability of $F(Y_l, u^l) - Y_l + R$ to the utility level under mimicking, viz. $\hat{V}^l(Y_l, F(Y_l, u^l), \pi_l, f)$. The Pareto efficient adjustment rule and the effect on the DWL may be summarized as ($\alpha^2$ is a shorthand for $\partial^2 \hat{V}^l / \partial z_2^2$, i.e. $\alpha^2$):

**Lemma 3**: With a passive audit policy, the optimal adjustment of the marginal tax rate to a fall in $f$ induces the following adjustment in the income earned by low ability agents:

$$\frac{d(-Y_l)}{d(-f)} = (1 + \lambda) \frac{\partial^2 \hat{V}^l}{\partial z_2^2} \frac{\partial \hat{V}^l}{\partial f} = \frac{\partial(-Y_l)}{\partial R} \frac{\partial^2 \hat{V}^l}{\partial z_2^2} \frac{\partial(-Y_l)}{\partial(-f)} \geq 0.$$ (6.1)

---

$^{10}$ The formal proofs of the lemma's in sections 6 and 7 are provided in appendix C.
In addition, the effect on the DWL in the economy is given by
\[
\frac{d\Delta}{d(-f)} = \lambda \frac{1}{\alpha^2} \frac{\partial v^{(2)}}{\partial (-f)} \geq 0.
\] (6.2)

Expression (6.1) is quite intuitive. Initially, when \(f = f^o\) no adjustment in \(Y_1\) and therefore in the marginal tax rate will occur because the mimicking agent does not participate on the black labour market.\(^{11}\) For further marginal reductions in \(f\), \(Y_1\) needs to be lowered to restore expected utility under mimicking to its original level. By how much, is given by the ratio in the middle expression. But this lowering of \(Y_1\) means that the original tax revenues are no longer sufficient to finance the transfer to type 1 people; thus more lump sum taxes have to be raised. As these in turn tighten the self-selection constraint, \(Y_1\) needs to be lowered further. Hence a 'multiplier' effect \((1+\lambda)\) bigger than unity. The far RHS term in (6.1) shows that the optimal adjustment rule in \(Y_1\) w.r.t. a fall in \(f\), is partially controlled by the optimal adjustment in \(Y_1\) to an increase \(R\). For this reason, I will say that the adjustment rule in \(Y_1\) is characterized by a revenue effect. The effect of a lower fine on the total DWL in the economy is given by expression (6.2). For high levels of \(f\), black market activity by the mimicker will be low and the relaxation of the penalty system on the DWL will be negligible.

As the penalty rate is brought down further, the marginal efficiency of the audit policy will catch up with its relative marginal cost. From then on we enter in another regime where auditing is 'intensive' and pursued up to the level where its relative marginal cost matches the relative marginal benefit (regime II: \(\gamma\kappa_\pi = \lambda\beta_\pi \& \pi_1 > \pi^o\)). In this regime, there is a degree of freedom and the optimal adjustment in the instruments is now governed by the way a lower penalty bears upon their respective marginal costs and benefits. Moreover, in this regime, we need to distinguish between a direct effect, and an indirect effect. The latter follows because changes in tax, audit and penalty treatments will trigger adjustments in the mimicker's black market behaviour, which in turn will affect the marginal benefits of the two instruments.

Before taking a look at the responses to \(f\), it is useful to inquire about the nature of the revenue effects. These are given by the following lemma where \(\delta\) is a positive expression which is (negatively) related to the second order condition

\[^{11}\partial v^{(2)}/\partial (-f) = \pi^* \beta_\pi X \geq 0 \iff X \geq 0.\]
of the planning problem:  

**Lemma 4:** An increase in wasteful government consumption has the following effects on the policy instruments

\[
\frac{d\pi_1}{dR} = \frac{\gamma}{\delta} \left( \frac{d\log x_i}{d(-Y_i)} - \frac{d\log \hat{\beta}_i}{d(-Y_i)} \right) + \frac{d\log \pi_1}{d(-Y_i)}, \tag{6.3}
\]

\[
\frac{d(-Y_i)}{dR} = \frac{\gamma}{\delta} \left( \frac{d\log x_i}{d\pi_1} - \frac{d\log \beta_i}{d\pi_1} \right) \frac{d\log \hat{\beta}_i}{d(-Y_i)}. \tag{6.4}
\]

**Furthermore, the effect on the DWL is**

\[
\frac{d\Delta}{dR} = \kappa \frac{d(-Y_i)}{dR} + \kappa \frac{d\pi_1}{dR} = \lambda > 0. \tag{6.5}
\]

When wasteful government consumption increases, this will be financed by a larger lump sum tax on agent 2, which will reduce his utility level below the level when mimicking. To restore self-selection, the planner adjusts both instruments $Y_i$ and $\pi_1$. The adjustment of an instrument is controlled by two measures: (i) the extent to which the net marginal cost (in relative terms) of the other instrument is increasing (this is the term in square brackets); and (ii) the extent to which the instrument enhances the marginal deterrence efficiency of the other instrument.

The marginal cost of both instruments is always increasing due to the convexity of both the audit cost function and the low ability agent’s preferences. The marginal efficiency of $\pi_1$ is decreasing in $\pi_1$ (the direct effect is nil by the linearity of $\nu^{(i)}$ in $\pi_1$, but the indirect effect via $X$ is negative)\(^{13}\). As to the marginal efficiency of the marginal tax rate, an increase in $Y_i$ has a direct effect on $\beta$, which bears an ambiguous sign and an indirect effect (via the optimal adjustment $X$) which is negative.\(^{14}\) In the remainder of this essay, I will make the assumption

---

\(^{12}\) The expression for $\delta$ is $\beta_i d\log(\beta_i^\pi) / d(-Y_i) - \beta_i d\log(\hat{\beta}_i^x) / d(-Y_i) - \beta_i d\log(\hat{\beta}_i^\pi) / d(-Y_i) + \beta_i d\log(\hat{\beta}_i^x) / d(-Y_i)$. In Appendix D it is shown that the standard assumptions and condition (E) below in the text make the marginal cost of each instrument increasing, and the marginal efficiency of an instrument decreasing in its own use but increasing in the use of the other instrument. This ensures that the problem is well behaved and therefore that $\delta > 0$.

\(^{13}\) Using the definition of $\beta_i$ and result (4.4) we get:

\[
\frac{\partial \beta_i}{\partial X} \cdot \frac{\partial X}{\partial \pi_i} = \frac{\partial X}{\partial \pi_i} - \Delta(\Delta X/\partial \pi_i)^2 < 0.
\]

\(^{14}\) The direct effect is given by $\frac{\partial \beta_i}{\partial X} \cdot \frac{\partial X}{\partial \pi_i}$ which can be shown to expand in a difference of two negative quadratic forms and cannot be signed—see (D.3). The indirect effect is given by $\frac{\partial \beta_i}{\partial X} \cdot \frac{\partial X}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial X}$. Using expressions (4.2)-(4.3), this indirect effect becomes $-\Delta(\Delta X/\partial \pi_i \cdot MRS_i + \Delta X/\partial Y_i - \Delta X/\partial \pi_i)^2 < 0$.  

17
\[
\frac{d\hat{\beta}_t}{d(-Y_t)} - \frac{\partial \hat{\beta}_t}{\partial(-Y_t)} + \frac{\partial \beta_t}{\partial X} \frac{dX}{d(-Y_t)} < 0; 
\]

that is, the direct effect of a higher marginal tax rate on its own efficiency never offsets the indirect effect. Therefore, on grounds of rising net marginal costs, the use of both instruments is intensified as \( R \) goes up. Moreover, it can be shown that under the standard assumptions the spillover effects mentioned under (ii) will be positive. Consider for instance \( d\hat{\beta}_t/d(-Y_t) \): under Edgeworth complementarity a reduction in \( Y_t \) will raise the utility difference between audit/non audit state, and under a forward bending labour supply schedule this same reduction will incite the mimicker to earn more on the black market which in turn will raise this utility difference.\(^{15}\)

Expression (6.5) in lemma 4 shows that the DWL in the economy will increase to the extent that the self-selection constraint is binding. In this sense, an increase in \( R \) necessitates the planner to pursue a more "active" policy.

These findings may be summarized as

**Proposition 1**: Let the audit policy be intense. Then under Edgeworth complementarity and a forward bending standard labour supply schedule, the revenue effect on the marginal tax rate is always positive. If in addition assumption (E) holds, the same is true for the revenue effect on the audit rate. The revenue effect on the DWL is always positive.

With the revenue effect being well-defined, the characterization of the penalty effects follows in a straightforward way. The impact of a marginal lowering of the penalty rate \( f_t \) on both instruments, may be decomposed into a revenue and a substitution effect, as the following lemma asserts:

**Lemma 5**: The effect of a marginal reduction in the penalty rate on the instruments decomposes as

\[
\frac{d\pi_1}{d(-f)} - \frac{d\pi_1}{dR} \frac{1}{\alpha^2 \partial(-f)} + \frac{\beta_t}{\delta} \frac{d\log(\beta_t/\beta_0)}{d(-f)}, \tag{6.6}
\]

\(^{15}\) \( d\hat{\beta}_t/d(-Y_t) = \partial \hat{\beta}_t/\partial(-Y_t) + \partial \hat{\beta}_t/\partial X \partial X/\partial(-Y_t) \). Using the definition of \( \beta_t \) and expression (4.4), the RHS can be written as \([-Du_{21} + MR_{11} + Du_{21}] + \Delta \cdot \partial X/\partial \pi_t \cdot \partial \hat{\pi}_1/\partial(-Y_t) \). Edgeworth complementarity ensures that the second term in square brackets is positive, while a upward sloping standard labour supply curve and normality of consumption make the final term positive (cf 4.2-3).
\[
\frac{d(-Y_c)}{d(-f)} - \frac{d(-Y_s)}{d(-f)} \frac{1}{\alpha^2} \frac{\partial \nu^{(2)}}{\partial(-f)} - \frac{\beta_s}{\delta} \frac{d \log(\beta_s/\beta)}{d(-f)}. \tag{6.7}
\]

The final RHS terms satisfy the homogeneity restriction:
\[
\kappa_s \left[ -\frac{\beta_s}{\delta} \frac{d \log(\beta_s/\beta)}{d(-f)} \right] + \kappa_s \left[ \frac{\beta_s}{\delta} \frac{d \log(\beta_s/\beta)}{d(-f)} \right] = 0. \tag{6.8}
\]

By an appropriate fall in government consumption, the penalty effects on the instruments can be limited to the final RHS terms of (6.6) and (6.7). The homogeneity restriction implies that these remaining effects do not bear upon the DWL in the economy. Whence these effects are iso-DWL effects, or substitution effects, and it is clear that they are of opposite sign.\(^{16}\)

Since I have signed already the revenue effects, let me concentrate on the substitution term. How does a lowering of the penalty rate affect the ratio \(\beta_s/\beta_r\)? Again, a direct effect has to be distinguished from an indirect one. The former effect is given by
\[
\frac{\partial \log(\beta_s/\beta)}{\partial(-f)} = \frac{1}{\beta} \left[ -\mu^{(2)} X \right] - \frac{1}{\beta} \left[ \pi_1 (\mu^{(2)}) MRS^1 - \frac{1}{w_2} \mu^{(2)} X \right]. \tag{6.9}
\]

Under Edgeworth complementarity the second square brackets term is negative so that the direct effect cannot be signed. But whatever its sign, observe that its importance is negatively related to the degree of compliance.

The indirect effect works via \(X\). A first observation is that under a forward bending standard labour supply schedule, an increase in black market activity raises \(\beta_s\) and lowers \(\beta_r\); whence it raises their ratio.\(^{17}\) Second, we know from section 4 that black market activity raises after a fall in \(f\): the income effect will work in the same direction as the positive substitution effect (cf (4.5)). The indirect effect on \(\beta_s/\beta_r\) is therefore positive. Combining these results we have:

\(^{16}\) Because in regime II \(\kappa_s/\kappa_s = \beta_r/\beta_r\), the homogeneity restriction also implies that the substitution effects of the penalty rate do not affect the utility of the mimicking high ability agent. In fact, the substitution effects can be shown to follow from the dual problem where the planner has to minimize the expected utility of the mimicker subject to the constraint that the DWL in the economy cannot exceed a predetermined amount.

\(^{17}\) Upon use of (4.2)-(4.4) it can be established that \(\partial \beta_r / \partial X = -\Delta (\partial X / \partial x_r^1, MRS^1 + \partial X / \partial y_s) = \Delta (\partial X / \partial (-y_s)) < 0\), and \(\partial \beta_s / \partial X = \Delta (\partial X / \partial x_r^1) \geq 0\). Whence \(\partial (\beta_s / \beta_r) / \partial X \geq 0\).
\[
\frac{d(\beta_s/\beta_r)}{d((-f))} = -\frac{1}{\beta_r} \left[ \frac{\partial \beta_r}{\partial X} \frac{\partial X}{\partial (-f)} \beta_s \left( \frac{\partial \beta_s}{\partial X} + \frac{\partial \beta_r}{\partial X} \right) \right].
\] (6.10)

Two different sufficiency conditions may now formulated for a fall in \( f \) to exert a positive influence on \( \beta_r/\beta_s \). The first condition is simply that \( d\beta_s/d(-f) \) is positive; in other words, that the negative direct effect on \( \beta_r \) does not offset the positive indirect effect. This conforms with the idea that a more lenient penalty system makes auditing more efficient. *Ceteris paribus* this favours a substitution of the tax instrument for the audit instrument. The second condition is on the degree of compliance. Above, it was observed that the direct effects will be insignificant when \( X \) is small: only the positive indirect effect will remain significant (to be precise, only the substitution term of that effect). Thus, when a high degree of compliance prevails, the positive indirect effect on \( \beta_r/\beta_s \) will dominate. This discussion is summarized in

**Proposition 2:** Let the audit policy be intense. When either the rate of compliance is high, or when \( d\beta_s/d(-f) > 0 \), the substitution effect due to a lower penalty rate leads to more intensive auditing and a lower marginal tax rate.

Because under the conditions of Proposition 2, the revenue and substitution effects on the audit rate work in the same direction, the following corollary immediately follows:

**Corollary 1:** Let assumption (E) hold and let the standard labour supply schedule be forward bending. Then under the same conditions of proposition 2, a more lenient penalty system gives rise to a more intensive audit policy.

On the other hand, the revenue and substitution effects on the marginal tax rate work in different directions, and no a priori conclusions can be drawn on the net effect. Lemma 5 and proposition 2 are the main results of this paper. They show that growing evasion opportunities should incite the government to adjust tax and audit rates on the basis of two different arguments, and that these arguments need not necessarily work in the same direction. In fact, under (E), they cannot work in the same direction for both instruments at the same time. Thus, if in public debates it is argued that marginal tax rates should be lowered in a response to soaring evasion levies, one is implicitly assuming a dominating substitution effect.
7. Standard of living effects and Bergsonian responses to a lower penalty rate.

As argued in the beginning of the previous section, when tax and audit policies are designed to maximize a Bergsonian welfare index, the government is likely to impute part of the costs of its redistribution policy to the low ability class. This is illustrated in figure 3. With $f = f^{\circ}$ the second best frontier (SB) depicts the utility combinations under full compliance. When the penalty rate drops to its level given by law, the Pareto efficient adjustment rules describe how policy ought to change when moving for instance from $A$ to $B$. But when a strictly concave welfare function inspires policy, the optimal policy adjustment will be the one taking the economy from $A$ to $D$. Ideally, one would therefore like to carry out a comparative statics excercise on the solution to a welfare maximization problem, rather than on a Pareto problem. But such an exercise is very cumbersome because unlike the latter problem, the former is no longer recursive. I will therefore proceed in a more heuristic way by breaking down the move $A \rightarrow D$ into $A \rightarrow B$ and $B \rightarrow D$. The former move has already been described, so attention can now be confined to the policy adjustments when the living standard on low ability agents changes; I will call these standard of living effects.

--- insert figure 3 here ---

Again, the nature of these effects will hinge on the regime in which they take place. In regime I the audit policy is 'passive' ($\pi_{i} = \pi^{\circ}$) and only the income tax policy is considered for adjustment:

**Lemma 6:** With a passive audit policy, the optimal adjustment of the marginal tax rate to a rise in $a^{1}$ induces the following adjustments in the income earned by low ability agents:

\[
\frac{d(-Y_{i})}{da^{1}} = \frac{d(-Y_{i})}{dR} \frac{1}{a^{2}} \frac{\alpha^{2(1)} + \alpha^{2}}{\alpha^{1}} - (1+\lambda) \frac{(\alpha^{2(1)} + \alpha^{2})/\alpha^{1}}{\beta_{t}} > 0. \tag{7.1}
\]

The effect on the DWL in the economy is given by

\[
\frac{d\Lambda}{da^{1}} = \lambda \frac{1}{\alpha^{2}} \frac{\alpha^{2(1)} + \alpha^{2}}{\alpha^{1}} > 0. \tag{7.2}
\]
According to (7.1), guaranteeing low ability agents a higher utility level requires making them earn less on the official labour market. Whether this also gives rise to a higher marginal tax rate is a priori uncertain since under normality of leisure and for given $Y_l$, the implicit marginal tax rate falls as $\bar{u}^l$ (and hence $z_l$) goes up.\(^{18}\) In any case—as (7.2) teaches—guaranteeing low ability agents a higher welfare level boosts the DWL in the economy; redistribution is costly, a standard result.

In the more interesting regime II where both policies are 'active', the utility effects decompose in way similar to the penalty effects:

**Lemma 7:** The marginal effects on the policy instruments of a higher standard of living constraint on the low ability agents decompose as

$$
\frac{d\pi_{1}}{d\bar{u}^l} - \frac{d\pi_{1}}{d\bar{u}^1} \frac{1}{\alpha_1} \frac{\alpha_2}{\alpha_2 + \alpha_1} + \frac{\beta_{\bar{u}}}{\delta} \left( \frac{d\log(\beta_{\bar{u}}/\beta_{\bar{u}})}{d\bar{u}^1} - \frac{d\bar{u}^1}{d\bar{u}^1} \right),
$$

(7.3)

$$
\frac{d(-Y_l)}{d\bar{u}^1} - \frac{d(-Y_l)}{d\bar{u}^1} \frac{1}{\alpha_1} \frac{\alpha_2}{\alpha_2 + \alpha_1} - \frac{\beta_{\bar{u}}}{\delta} \left( \frac{d\log(\beta_{\bar{u}}/\beta_{\bar{u}})}{d\bar{u}^1} - \frac{d\bar{u}^1}{d\bar{u}^1} \right).
$$

(7.4)

The second RHS terms satisfy the homogeneity condition

$$
\kappa_{1} \left[ -\frac{\beta_{\bar{u}}}{\delta} \left( \frac{d\log(\beta_{\bar{u}}/\beta_{\bar{u}})}{d\bar{u}^1} - \frac{d\bar{u}^1}{d\bar{u}^1} \right) \right] + \kappa_{1} \left[ -\frac{\beta_{\bar{u}}}{\delta} \left( \frac{d\log(\beta_{\bar{u}}/\beta_{\bar{u}})}{d\bar{u}^1} - \frac{d\bar{u}^1}{d\bar{u}^1} \right) \right] = 0.
$$

(7.5)

For similar reasons as in section 6, the second RHS terms in (7.3-4) may be identified as substitution effects. Let us examine the sign of their common component. The influence of $\bar{u}^l$ on the marginal efficiency and cost of the instruments takes place through the higher level of $z_l$. Normality of leisure guarantees a negative sign for $d\log k_{\bar{u}}/d\bar{u}^l$ (cf eqs (5.2) and (5.5)). If in addition the standard labour supply schedule is forward bending, $d\log \beta_{\bar{u}}/d\bar{u}^l$ will be negative.\(^{19}\) This leaves us to sign the term $d\log \beta_{\bar{u}}/d\bar{u}^l$.

Recall that $\beta_{\bar{u}}$ is given by $\alpha_{21}[\text{MRS}^{1}-\text{MRS}^{21}]$. Under normality of leisure, the increase in consumption possibilities through a higher $z_1$ (by means of which a higher $\bar{u}^l$ is implemented) directly raises both marginal rates of substitution (cf (U2)), while it lowers the mimicker's expected marginal utility of income. But

---

\(^{18}\) Differentiation of $1-F_{\bar{u}}(Y_{1}, \bar{u}^l)$ yields $F_{\bar{u}}(d(-Y_{1})/d\bar{u}^l) - F_{\bar{u}}$. By lemma 5, the first term is positive, while normality of leisure ensures a positive sign for $F_{\bar{u}}$ (see (5.2)).

\(^{19}\) Using (5.2), (5.6), and (4.4) it can be checked that the effect of $\bar{u}^l$ on $\beta_{\bar{u}}$ decomposes as $[D_{\alpha}^{21} + \Delta \cdot X \cdot \delta \pi_{1}/\partial \pi_{1} \cdot \partial X / \partial z_{1}] / \alpha_{1} < 0$. The first term is the direct effect. The second term accounts for the indirect effect and will be negative under normality of leisure and a forward bending standard labour supply schedule.
as $z_i$ is raised, the mimicker adjusts his optimal black market earnings downwards (under normality of leisure, cf (4.2)). This, in turn, negatively affects his marginal rate of substitution (cf footnote 5), and positively affects his marginal utility of income (see again (4.2)). Thus, while the direct effect of a higher $\bar{A}_i^1$ on $\beta_i$ is ambiguous, the indirect effect works in favour of a higher marginal deterrence efficiency of the tax rate. In the following proposition, I will assume that the combined effect is indeed positive.

**Proposition 3:** Let the audit policy be intense and suppose that $d\beta_i/d\bar{A}_i^1 > 0$. Under a standard forward bending labour supply schedule, the substitution effect due to a more stringent standard of living constraint will lead to less intensive auditing and a marginal tax rate on low income earners such that these agents decide to earn less.

Observe that proposition 3 does not make any direct statement about the optimal adjustment of the marginal tax rate for similar reasons as lemma 6 did not. In view of the revenue effects, this proposition leads to the following

**Corollary 2:** Suppose $d\beta_i/d\bar{A}_i^1 > 0$ and let the standard labour supply schedule be forward bending. A more stringent standard of living constraint leads to a marginal tax rate on low income earners which induces these people to earn less.

On the other hand, a higher $\bar{A}_i^1$ will only give rise to more intensive auditing when the positive revenue effect is not offset by the negative substitution effect.

Although standard of living effects have an interest on their own, their identification allows me now to say something on welfaristic responses to a more lenient penalty system. Consider a policy instrument $x$, say. Then, for a given living standard $\bar{A}_i^1$, a drop in the penalty rate will lead to the adjustment $dx/d(-f)$. But at the same time, the DWL raises with $d\Delta/d(-f)$. This increased burden on the economy will now be shared by both classes. Suppose that the (possibly varying) share of the low ability class is given by $s_i$, then the ensuing adjustment in the instrument $x$, is given by $(-dx/d\bar{A}_i^1) \cdot a_i^1 \cdot s_i \cdot d\Delta/d(-f)$. Whence, the Bergsonian response of instrument $x$ is given by $dx/d(-f) - (dx/d\bar{A}_i^1) \cdot a_i^1 \cdot s_i \cdot d\Delta/d(-f)$. The information to evaluate this expression is now summarized in table 2 below.
Table 2  The penalty and standard of living effects on instruments and deadweight loss.

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\text{Instrum.}$</th>
<th>$dx/d(-f)$</th>
<th>$dx/d\pi^1$</th>
<th>$d\Delta/d(-f)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$(-Y_1)$</td>
<td>$^* +$</td>
<td>$^* +$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\pi_1$</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
| II     | $(-Y_1)$        | $-(\text{SE dominates})^e$  
$\quad + (\text{RE dominates})^e$ | $^* +$       |                 |
|        | $\pi_1$         | $^*$         | $-(\text{SE dominates})^f$  
$\quad + (\text{RE dominates})^f$ | $^*$           |

(a): lemma 3; (b): lemma 6; (c) proposition 1 and lemma 5; (d): corollary 2;  
(e): corollary 1; (f) proposition 1 and lemma 7; SE (RE) = substitution (revenue) effect.

In regime I, both the penalty and standard of living effects are entirely controlled for by revenue effects. Two opposing forces are at work. The efficiency argument calls for a lower income of low ability agents, while the sharing of the increased burden will lower the redistributitional ambitions of the planner and this is a motivation to make low ability agents earn more. In regime II, the possibility to substitute one instrument for the other complements the revenue effects. A necessary condition for $Y_1$ to fall is that, under the assumptions of corollary 2, the revenue effect of a lower penalty rate dominates the substitution effect. Likewise, when the audit frequency is lowered, then under the conditions of corollary 1, the revenue effect of a higher standard of living constraint on low ability agents must necessarily dominate the substitution effect. On the other hand, at this level of generality, no necessary conditions are available for explaining a rise in $Y_1$ and/or $\pi_1$. In the following section, I shall present a numerical exercise to get some feeling for the effects of evasion opportunities on the design of optimal fiscal policy.

8. A numerical example.

To provide some insight in the working of the model, I have calculated the first, second and third best frontier for an economy where: $w_1=10$ and $w_2=20$; preferences are Cobb-Douglas, $u(c,t)=c^\delta t^{1-\delta}$; the audit cost technology is quadratic from $\tau^*=.01$ onwards, $K(\pi)=5(\pi-\tau^*)^2$; taxation is purely redistributive, $R=0$; the penalty rate on evaded income is $f=1.2$.

Figure 4 partially pictures the three different frontiers and the choices made on those frontiers by a planner with utilitarian (U), Rawlsian (R) and egalitarian
(E) preferences. The first-best frontier (FB) is the locus of Pareto efficient allocations when the planner has perfect information on the abilities \(w_i\). It contains the competitive outcome (C). The second-best frontier (SB) traces out the allocations the planner can implement when she is assignment uncertain, but can costlessly observe labour income. In the neighbourhood of the no intervention point, the frontier—here only shown when redistribution takes place from the high to the low ability class—coincides with the first best frontier; this is where (small) lump sum transfers and taxes are still incentive compatible. Once the planner wants to guarantee low ability agents a utility level of (approximately) 1.39, decentralization requires the introduction of distortionary taxation of the income of these agents. When even information on income becomes costly to collect, the only allocations the planner can implement are those bounded by the third best frontier. Table 3 gives more detailed information on these different allocations.

Inspection of figure 4 teaches that any of the utility levels which agents of type 1 can be guaranteed under costly monitoring of income, can be guaranteed under costless monitoring in an undistorted way. Thus, if the planner's preferences were lexicographic, in that she wants to secure low ability agents a standard of living \(l^1\) in any circumstance, tax evasion opportunities give rise to higher marginal tax rates.

— insert figure 4 here —

This statement is no longer true when the planner's preferences are either utilitarian or Rawlsian. A optimal marginal tax rate of 17.7% (56.3%) under second-best falls to 4.3% (42.4%) under third best, if social preferences are utilitarian (Rawlsian). But at the same time, the planner engages in an intensive audit policy: 12.3% (44.7%). This increases the total deadweight loss in the economy, both in absolute terms as well as relative to GDP. Also observe that both under utilitarian and Rawlsian social preferences the low ability agents earn more at the third best optimum than at the second best optimum. As the discussion at the end of section 7 implies, this can occur either because the

---

20 For Cobb-Douglas preferences, this frontier is has a linear shape (with slope \(-w_i/w_j\)) as long as both agents supply labour. When one of the agents decides no longer to participate on the labour market because he is given too generous transfers, it becomes strictly convex to the origin. In fact, a utilitarian planner would select a point \((U_r)\) on the upper strictly convex section of the frontier (not drawn) (cf Atkinson & Stiglitz, 1980, Ch 12).
substitution effect of the penalty effect dominates, or because the standard of living effect and/or the deadweight loss effect are sufficiently strong. Given the existence of the Rawlsian solution, the egalitarian outcome is clearly inefficient—both under second and third best. When egalitarian motives are at play, the marginal tax rate of 60.7% under second best will be raised to a level of 70.9% when the planner can no longer observe income costlessly. Also the audit rate will be set at 70.9%. Such a high audit rate is even more astonishing since even out of equilibrium, the auditor is not likely to detect evasion activities going on at all. The resulting deadweight loss in the economy is tremendous: 44% of GDP is devoted to make high ability agents reveal their type.

Table 3 Details on some first-, second-, and third-best allocations.

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>U_f</th>
<th>R_f=E_f</th>
<th>U_s</th>
<th>R_s</th>
<th>E_s</th>
<th>U_T</th>
<th>R_T</th>
<th>E_T</th>
</tr>
</thead>
<tbody>
<tr>
<td>u_1</td>
<td>1.28</td>
<td>2.82</td>
<td>1.53</td>
<td>1.43</td>
<td>1.47</td>
<td>1.47</td>
<td>1.32</td>
<td>1.38</td>
<td>1.32</td>
</tr>
<tr>
<td>u_2</td>
<td>1.69</td>
<td>0.56</td>
<td>1.53</td>
<td>1.59</td>
<td>1.49</td>
<td>1.47</td>
<td>1.66</td>
<td>1.51</td>
<td>1.32</td>
</tr>
<tr>
<td>z_1</td>
<td>4</td>
<td>13.3</td>
<td>4.77</td>
<td>3.97</td>
<td>2.80</td>
<td>2.62</td>
<td>4.03</td>
<td>3.09</td>
<td>1.99</td>
</tr>
<tr>
<td>Y_1</td>
<td>4</td>
<td>0</td>
<td>2.85</td>
<td>2.76</td>
<td>.414</td>
<td>0</td>
<td>3.69</td>
<td>1.94</td>
<td>0</td>
</tr>
<tr>
<td>z_2</td>
<td>8</td>
<td>2.67</td>
<td>7.23</td>
<td>7.52</td>
<td>7.05</td>
<td>6.95</td>
<td>7.84</td>
<td>7.16</td>
<td>6.23</td>
</tr>
<tr>
<td>Y_2</td>
<td>8</td>
<td>16.0</td>
<td>9.16</td>
<td>8.73</td>
<td>9.43</td>
<td>9.57</td>
<td>8.24</td>
<td>9.27</td>
<td>10.7</td>
</tr>
<tr>
<td>x</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.34</td>
<td>1.95</td>
<td>.001</td>
</tr>
<tr>
<td>r_1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.177</td>
<td>.563</td>
<td>.607</td>
<td>.043</td>
<td>.424</td>
<td>.709</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>-13.3</td>
<td>-1.93</td>
<td>-1.70</td>
<td>-2.61</td>
<td>-2.62</td>
<td>-4.94</td>
<td>-1.98</td>
<td>-1.99</td>
</tr>
<tr>
<td>y</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.123</td>
<td>.447</td>
<td>.709</td>
</tr>
<tr>
<td>LSTr</td>
<td>0</td>
<td>13.3</td>
<td>1.93</td>
<td>1.16</td>
<td>1.47</td>
<td>1.47</td>
<td>.333</td>
<td>.774</td>
<td>.272</td>
</tr>
<tr>
<td>K(\pi)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.064</td>
<td>.956</td>
<td>2.44</td>
</tr>
<tr>
<td>\Lambda</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.050</td>
<td>.904</td>
<td>1.15</td>
<td>.066</td>
<td>1.34</td>
<td>4.70</td>
</tr>
<tr>
<td>\Lambda/GDP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.004</td>
<td>.092</td>
<td>.120</td>
<td>.006</td>
<td>.119</td>
<td>.441</td>
</tr>
</tbody>
</table>

* GDP = x_1 Y_1 + Y_2.


In this paper, I discussed the Pareto efficient tax and audit policy when the government lacks perfect information on endogenous labour income. The non-linearity of the tax schedule reduces tax evasion to an out-of-equilibrium strategy
of the high ability agents, but the credibility of that threat imposes a large burden on the economy. In this respect, the model is another illustration of Hammond's (1987) proposition that markets may act as constraints on economic policy. More specifically, it was shown that under a variant of the Single Crossing property, the optimal redistribution policy is characterized by a zero marginal tax rate at the top, a strictly positive marginal tax rate on low income earners, and an intensive auditing of those income earners whenever the marginal cost of initiating an audit is not too large.

The modern public finance approach is concerned with the structure of the arguments which should motivate fiscal policy. In this respect, I have shown that in response to growing evasion opportunities, governments should adjust their fiscal policy for two reasons. The first reason is that such opportunities place a higher burden on the economy. For this reason, instruments should be adjusted in the same way as required when, say, funds need to be raised after a natural disaster or when a larger contribution to an supranational government is demanded. The second reason why instruments should be adjusted, is because evasion opportunities affect the marginal efficiency of one fiscal instrument vs the other. Not reacting to such changes in relative efficiency would mean that available opportunities to enforce self-selection are not fully taken advantage of.

Extension of the model into several directions is desirable. First, the important assumption that low income earners have no access to the black labour market should be dropped. If the marginal tax rate on low income is high enough, it is clear that this class will have a strong incentive to participate on that market. This extension would not affect the characterization and decomposition results, but the signing would become more difficult since the marginal cost of the tax instrument \( \kappa \) too will become sensitive to changes in the audit and penalty rate.

Endogenizing the penalty rate would be another extension, though not a straightforward one. In a companion paper (Schroyen, 1994, essay 4), I have shown that in a similar model with the two classes having access to the black labour market; the Pareto efficient penalty policy displays properties which goes against notions of fair retribution.

Finally, note that in the determination of the optimal policy, information about the cost structure technology turned out to be equally important as information on preferences. Empirical research on the properties of that technology is a prerequisite for the implementation of optimal tax systems.
References


Appendix A: The Single Crossing property.

For a mimicking agent with wage rate \( w \) who has access on the black labour market, the marginal rate of substitution in the \( (z, Y) \) space, evaluated at the bundle \( (z, Y) \), is given by

\[
MRS = \frac{Eu(z + (1-\tau)X, 1 - \frac{Y - X}{w})}{wEu(z + (1-\tau)X, 1 - \frac{Y - X}{w})},
\]

(A.1)

where \( \tau \) is the stochastic penalty rate being zero with probability \( (1-\pi) \) and \( f \) with probability \( \pi \).

For such a mimicking agent, the FOC w.r.t. \( X \) is given by \( Eu(1-\tau) - Eu/\tau = 0 \), or \( MRS = (1-\tau) \), where \( \tau \) denotes the certainty equivalent marginal penalty rate, i.e. \( \tau = w_f Eu(1-\tau)/Eu \).

The effect of an increase in \( w \) on the marginal rate of substitution is given by

\[
\frac{\partial MRS}{\partial w} = \frac{\partial MRS}{\partial w} \frac{\partial MRS}{\partial X} \frac{\partial X}{\partial w}.
\]

(A.2)

The direct effect has two components:

\[
\frac{\partial MRS}{\partial w} = -\frac{1}{w} MRS + \frac{1}{Eu(1-\tau) MRS Eu(1-\tau)} \left( \frac{Y - X}{w^2} \right).
\]

(A.3)

Normality of consumption is a sufficient condition to make the second term and therefore the entire direct effect negative.

But in addition, there is an indirect effect via the adjustment in \( X \) to a higher wage rate. Under normality of consumption and leisure, a higher \( X \) will raise the \( MRS \) (see also the expression in footnote 6 in the text):

\[
\frac{\partial MRS}{\partial X} = -\frac{1}{wEu} \left[ (Eu_1(1-\tau) - \frac{1}{w} Eu_w) - \frac{Eu_1(1-\tau) - \frac{1}{w} Eu_1}{Eu_1} \right]
\]

\[-\frac{\Delta}{Eu_1} \left[ \frac{\partial X}{\partial Y_1} \frac{\partial X}{\partial Y_1} \right] = \frac{\Delta}{Eu_1} \frac{\partial X}{\partial Y_1} > 0.
\]

(A.4)

Here \( \Delta \partial X/\partial Y_1 \) represents the effect of an increase in \( Y_1 \) on the mimicker's optimal choice of \( X \) when this agent is compensated by a simultaneous increase in \( z \).

The effect of a higher wage rate on the optimised black labour market earnings works through a higher rate of return on that market, and through a higher availability of leisure since less hours of labour need to be supplied on the official labour market to mimic agent 1's official income (which is equivalent to the effect of a lower \( Y_1 \)):

\[
\frac{\partial X}{\partial w} = -\frac{1}{\Delta} \left[ (Eu_1(1-\tau) - \frac{1}{w} Eu_1) \frac{Y + X}{w^2} - \frac{1}{w} Eu_1 \right].
\]

(A.5)

Under normality of consumption, this expression is positive.

Using (A.3), (A.4), (A.5), and the FOC \( MRS = (1-\tau) \), expression (A.2) can be written as

\[
\frac{dMRS}{dw} = -\frac{1}{w} MRS(1+\frac{\partial X}{\partial Y_1}) + \frac{1}{Eu_1}\left[ (1+\frac{\partial X}{\partial Y_1}) - Eu_1(1-\tau) + (1-\tau) \frac{\partial X}{\partial Y_1} \right] \left( \frac{Y + X}{w^2} \right).
\]

(A.6)
Because under normality of consumption and leisure the term \(\partial \bar{X}/\partial Y\), is negative, the opportunity to participate on the black labour market makes that \(dMRS/dw > \partial MRS/\partial w\), which is an application of the Le Chatelier principle. As stated earlier, normality of consumption is sufficient to make the RHS term of the inequality negative. Whence, for single crossing to obtain when access to a black labour market is possible, the compensated effect \(\partial X/\partial Y\) should not be too negative.

**Appendix B: Proof of Lemma 1**

Nine combinations are possible for the fiscal treatment of high ability agents:

<table>
<thead>
<tr>
<th>(MRS^S\mid x=0 &lt; 1)</th>
<th>(\pi_2=0)</th>
<th>(\pi_2 \in (0, \pi^s])</th>
<th>(\pi_2 &gt; \pi^s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.1</td>
<td>I.2</td>
<td>I.3</td>
<td></td>
</tr>
<tr>
<td>II.1</td>
<td>II.2</td>
<td>II.3</td>
<td></td>
</tr>
<tr>
<td>III.1</td>
<td>III.2</td>
<td>III.3</td>
<td></td>
</tr>
</tbody>
</table>

Suppose now that by making optimal use of the black labour market, each of these fiscal treatments provide high ability agents with the same expected utility level. I will now check in which of nine cases the government can keep providing this expected utility level but can raise net government revenue by reforming the treatment.

First observe that cases I.1, II.1 and II.2 are all equivalent because they provide type two agents with the same utility level and raise the same amount of tax revenue for the government. (II.1 and II.2 raise the same net revenue because audits are free up to \(\pi^s\). The equivalence between I.1 and II.1 was discussed in section 5 of the text)

Cases III.1 and III.2 are inefficient because by the Single Crossing Property, no type 1 agent will envy the allocation designed for a type 2 agent. Therefore, the government could provide type 2 agents with the same utility level by giving them an undistorted \((z_2, Y)\) bundle. This would raise the tax revenue of the government.

Case II.3 is clearly inefficient because upon receiving an undistorted allocation, no agent has an incentive to participate on the black labour market. By auditing \(Y\), income levels at a higher rate than \(\pi^s\), the government would incur unnecessary audit costs. The same is true for case III.3: with an \(MRS^S > 1\) no agent has an incentive earn black income.

Consider now case I.2. First assume that \(\pi_f > 1-MRS^S\mid x=0\). Then \(X=0\) is the optimal choice. But since the agent receives a distorted bundle, the government could provide this agent the same utility level by moving the bundle along its indifference curve in North-East direction. This raises tax revenue (cf figure 5).

— insert figure 5 here —

Next, assume that evasion does occur in case I.2. This situation would also be inefficient. The agent will bear a risk, and the government will collect a strictly positive amount of fines: \(\pi_fX\) (cf point A in figure 5). But as \(\pi_fX=X-E(1-\tau)X\), the government could collect the same amount of resources from agent 2 by allocating the bundle \((z_2+E(1-\tau)X,Y,+,X)\). This requires the same amount of effort, but replaces the risky final consumption outcome by its expected value, making the agent better off (given risk aversion). In other words, point C in figure 5 is Pareto superior to point A. But under normality of leisure, the incentive to evade at point C will be lower than at point A: \(MRS^S\mid x=0_c > [MRS^S\mid x=0_A].\) Then either \(MRS^S\mid x=0_c > 1-\pi_f\), in which case no evasion further occurs and the deadweight loss can be further minimized by moving up to the point where \(MRS^S\mid x=0 = 1\) (figure 6a). Or it is still true that \(MRS^S\mid x=0_c < 1-\pi_f\), in which case the procedure can be started over (figure 6b).

— insert figure 6 here —
Case I.3 is inefficient for the same reasons as case I.2. In addition, unnecessary audit costs are incurred by the government.

Appendix C: Reduction of the planning problem and proofs of the lemmas.

Reduction of the planning problem:

There are two ways to solve the planning problem formulated at the beginning of section 5. The standard way is to set up the Lagrangian of a constrained maximization problem in 6 decision variables, and to derive the system of first order conditions. The policy characterization can then be obtained by manipulating those FOCs, and adjustment rules for the decision variables w.r.t. marginal changes in the parameters of the planning problem (like $R$ or $f$) may be derived by total differentiation of the system of (6) FOC's and (3) side constraints, and inverting the corresponding bordered Hessian. However, making use of lemma 1 and the subsequent convention, it is possible to transform this high dimensional constrained maximization problem into an unconstrained one dimensional problem. The comparative statics exercise then simply amounts to dividing the effect of a parameter change on the (single) FOC by the SOC. It is this latter approach that I will follow.

First observe that by making use of the function $F(Y_i,u^i)$, defined at the bottom of page 12, I get rid of the side constraint ($\mu$). Next, constraint ($\gamma$) is used to compute the tax liability of the high ability agent: $Y_i= F(Y_i,u^i)-Y_i+K(\pi_i)+R$. By lemma 1 and the convention, this tax is to be paid in a lump sum way. Whence, if $\psi(w_2,-T)$ is the indirect utility function informing about the utility level of an undistorted type 2 agent paying a lump sum tax $T$, the self-selection constraint ($\lambda$) can be written as

$$\psi[w_2,-F(Y_i,u^i)-Y_i+K(\pi_i)+R] = \nu [Y_i F(Y_i,u^i), \pi_i, f] + \nu$$

for $\nu=0$. For a given audit probability $\pi_i$, this equality implicitly defines the amount of gross official income $Y_i$ such that the standard of living, the self-selection and the government budget constraints are simultaneously satisfied. It will be denoted as $Y_i(\pi_i;R,f,u^i,\nu)$. By total differentiation of (C.1), it may be checked that:

$$\frac{\partial Y_i}{\partial \pi_i} = \frac{a^2}{a^2 \kappa_i + \beta}, \frac{\partial Y_i}{\partial R} = \frac{\partial Y_i}{\partial \kappa_i}, \frac{\partial Y_i}{\partial \beta}, \frac{\partial Y_i}{\partial f}, \frac{\partial Y_i}{\partial \sigma}, \frac{\partial Y_i}{\partial \sigma^2}, \frac{\partial Y_i}{\partial \sigma^3}, \frac{\partial Y_i}{\partial \sigma^4}.$$  

Whence, the planner needs to solve the following one dimensional maximization problem:

$$\max_{w_2,\pi_i} \psi[w_2,-F(Y_i(\pi_i;R,f,u^i,0),u^i)+Y_i(\pi_i;R,f,u^i,0)-K(\pi_i)-R].$$  

(C.3)

When the relative marginal cost and the relative marginal efficiency of the audit instrument are respectively defined as $k = \kappa_i / \kappa$, and $b = \beta_i / \beta$, the necessary Kuhn-Tucker conditions for an optimal choice are given by

$$\begin{align}
(\pi - \pi^*) &\geq 0, \\
(k - b) &\leq 0, \\
(\pi - \pi^*)(k - b) &\leq 0.
\end{align}$$

(C.4)

---

20 The purpose of the parameter $\nu$ is purely analytical—see (C.5) below. Its value will be identically zero throughout. In fact, $\nu dv$ may be interpreted as a marginal windfall utility gain when mimicking.
Thus, a necessary condition for an 'intensive' audit policy (i.e., $\pi_i > \pi^*$) is the equalization of the relative marginal efficiency of this instrument to its relative marginal cost.

The equilibrium values of the Lagrange multipliers $\gamma$, $\lambda$, and $\mu$ may now be retrieved by perturbing the objective function w.r.t. the parameters $R$, $v$, and $a'$, respectively:

$$
\gamma = \frac{dv^2}{dR} - \frac{\alpha^2 \beta_t}{\beta_t - \alpha^2 \kappa_t}, \quad \lambda = \frac{dv^2}{dv} - \frac{\gamma \kappa_t}{\beta_t}, \quad \mu = \frac{dv^2}{da'} - \gamma \lambda \alpha \alpha^2.
$$

(C.5)

Clearly, these multipliers take on positive values. Notice that

$$
(1 - \lambda) = \frac{\gamma}{\alpha^2}, \quad \text{and} \quad \frac{\partial (-Y)}{\partial R} = \frac{\gamma}{\beta_t},
$$

(C.6)

and that the FOC allows to simplify the first expression in (C.2) as follows:

$$
\frac{\partial Y}{\partial \pi_i} = \frac{\beta_t}{\beta_t}.
$$

(C.7)

These relationships will prove helpful later on.

When an intensive audit policy is optimal, the FOC may also be written as $\log(b/k) = 0$. This may be differentiated w.r.t. $\pi_i$ to yield the necessary SOC:

$$
\frac{d \log(b/k)}{d \pi_i} = \frac{1}{\beta_t} \left[ \beta_t \frac{d \log(b/k)}{d \pi_i} \beta_t \frac{d \log(b/k)}{d(-Y)} \right] - \frac{1}{\beta_t} (-\delta) < 0,
$$

(C.8)

with obvious definition of $\delta$. In appendix D it is argued that the assumptions U1-U4, A2-A3, and E in the text are sufficient for the SOC to be verified.

**Proof of lemma 2:**

Upon using the expression for $\lambda$ in (C.5), and the definitions of $b$ and $k$, the first order conditions may be formulated in more familiar terms as

$$
\gamma \kappa_t = \lambda_2 \beta_t, \quad \gamma \kappa_t \geq \lambda_2 \beta_t, \quad \text{with} \quad \text{when} \quad \pi_i > \pi^*.
$$

(C.4')

with the second expression holding with equality whenever an intensive audit policy is optimal. QED

**Proof of lemma 3:**

With a passive audit policy, the required fall in $Y_i$ due to a fall in $f$ is given by the third expression in (C.2); this is the far RHS of (6.1). Making use of (C.6) produces the middle expression in (6.1). The effect on the DWL is obtained by multiplying $\partial(-Y)/\partial(-f)$ by $\kappa$. Using the second expression in (C.6) and the formula for $\lambda$ in (C5), yields (6.2). QED

---

21 When differentiating, it is kept in mind that $Y_i = Y_i(\pi_i; R, f, \beta', 0)$ and that $\partial Y_i/\partial \pi_i = \beta_t / \beta_t$.

22 Essentially, what the standard assumptions and condition (E) do is to make the marginal cost of each instrument increasing, and to make the marginal efficiency of an instrument decreasing in its own use but increasing in the use of the other instrument.
Proof of lemma 4:
When the audit policy is 'intensive', the effect of a change in $R$ on the optimal audit rate is given by

$$\frac{d\pi_1}{dR} = \frac{\beta_1 d\log(b/k)}{\delta dR},$$

(C.9)

as $-\delta/b_1$ is the SOC. The parameter $R$ does not affect $b/k$ directly, but it does through $Y_1(\pi_1;R,f,a',0)$. Because $\partial(-Y_1)/\partial R = \gamma/b_1$, (C.9) becomes

$$\frac{d\pi_1}{dR} = \frac{\beta_1 d\log(b/k)}{\delta d(-Y_1)} Y_1 \left( \frac{d\log\beta_1}{d(-Y_1)} + \frac{d\log\beta_1}{d(-Y_1)} \right),$$

(C.10)

which is expression (6.3) because $d\beta_1/d(-Y_1) = d\beta_1/d\pi_1$. The equilibrium response in $(-Y_1)$ is given by $\partial(-Y_1)/\partial\pi_1, d\pi_1/dR + \partial(-Y_1)/\partial R$. These terms are respectively given by $-\beta_1/b_1$, expression (C.10), and $\gamma/b_1$. Thus,

$$\frac{d(-Y_1)}{dR} = -\frac{\beta_1 Y_1}{b_1} \left[ \frac{d\log\beta_1}{d(-Y_1)} + \frac{d\log\beta_1}{d(-Y_1)} \right] \frac{\gamma}{\beta_1},$$

(C.11)

Using the expression for $\delta$, the above expression reduces to (6.4). The effect on the DWL is given by $\kappa_1 \partial(-Y_1)/\partial\pi_1, d\pi_1/dR + \partial(-Y_1)/\partial R + \kappa_1 [d\pi_1/dR]$. Rearranging, this reduces to $0 = \pi_1/dR + \kappa_1 \delta(-Y_1)/\partial R$ which equals $\lambda$ (see (C.5)); hence formula (6.5). QED.

Proof of lemma 5:
A fall in $f$ affects $b/k$ directly, but works also through $Y_1(\pi_1;R,f,a',0)$. Thus, the equilibrium response in $\pi_1$ is given by

$$\frac{d\pi_1}{d(-f)} = -\frac{\beta_1 d\log(b/k)}{\delta d(-Y_1)} \frac{\delta (-Y_1) \delta (-f)}{\delta (-Y_1) \delta (-f)}.$$

(C.12)

Using the third expression in (C.2), the first RHS term can also be written as:

$$\left( \frac{\beta_1 d\log(b/k)}{\delta d(-Y_1)} \frac{\delta (-Y_1)}{\delta (-f)} \right) \frac{1}{a^2 \delta (-f)}.$$

(C.13)

Since the round brackets term is equal to $d\pi_1/dR$, (C.12) becomes

$$\frac{d\pi_1}{d(-f)} = -\frac{\beta_1 d\log(b/k)}{\delta d(-Y_1)} \frac{\delta (-Y_1) \delta (-f)}{a^2 \delta (-f)}.$$

(C.14)

Expression (6.6) is then achieved after expansion of the final RHS term. To obtain the equilibrium response in $(-Y_1)$, I proceed as in the proof of lemma 4. $d(-Y_1)/d(-f)$ is given by $\partial(-Y_1)/\partial\pi_1, d\pi_1/d(-f) + \partial(-Y_1)/\partial (-f)$. These terms can be replaced by $-\beta_1/b_1$, expression (C.14), and the RHS of the third expression in (C.2), respectively. Rearranging then yields,

$$\frac{d(-Y_1)}{d(-f)} = \left( \frac{\delta (-Y_1)}{\beta_1} \frac{d\pi_1}{dR} \right) \frac{1}{a^2 \delta (-f)}.$$

(C.15)
Because the round brackets term equals \( d(-Y_i)/dR \), expression (6.7) follows. Finally, observe that

\[
\sum_{\delta} \left[ \frac{\beta_x \cdot d\log(b/k)}{d(-f)} \right] \frac{1}{\delta} \left[ \frac{\beta_y \cdot d\log(b/k)}{d(-f)} \right] = \frac{1}{\delta} \beta_x, \delta(k-b) \cdot \frac{d\log(b/k)}{d(-f)},
\]

which is zero by the FOC. QED

**Proof of lemma 6:**
This proof goes exactly along the same lines as the proof of lemma 3, making use of the fourth expression in (C.2). QED

**Proof of lemma 7:**
A rise in \( d^l \) affects \( b/k \) directly (by raising \( z_h \)), but also indirectly through \( Y_c(\pi; R, f, d^l, 0) \). The equilibrium response in \( \pi \), is therefore given by

\[
\frac{d\pi}{d^l} = \frac{\beta_x \cdot \delta \log(b/k)}{\delta} \cdot \frac{\delta(-Y_c)}{\delta a^l} + \frac{\beta_y \cdot \delta \log(b/k)}{\delta} \cdot \frac{\delta(-Y_c)}{\delta a^l}
\]

Using the fourth expression in (C.2), the first term in square brackets can also be written as:

\[
\left( \frac{\beta_x \cdot \delta \log(b/k)}{\delta} \cdot \frac{\delta(-Y_c)}{\delta a^l} \right) \cdot \frac{1}{a^l} = \frac{a^l}{a^l}
\]

But the round brackets term is equal to \( d\pi / dR \), so (C.17) becomes

\[
\frac{d\pi}{d^l} = \frac{d\pi}{d^l} \cdot \left( \frac{\beta_x \cdot \delta \log(b/k)}{\delta} \cdot \frac{\delta(-Y_c)}{\delta a^l} \right) \cdot \frac{1}{a^l}
\]

Expression (7.3) is then obtained by expanding the final RHS term. The equilibrium response in \( (-Y_c)/d^l \), is given by \( \frac{\delta(-Y_c)}{\delta a^l} \cdot d\pi / d^l + \frac{\delta(-Y_c)}{\delta a^l} \). These terms can be replaced by \( -\beta_x/\beta_y \), expression (C.19), and the RHS of the fourth expression in (C.2). Rearranging then yields,

\[
\frac{d(-Y_c)}{d^l} = \left( \frac{\delta(-Y_c)}{\delta a^l} \cdot \frac{\beta_x \cdot d\pi}{\delta} \right) \cdot \frac{1}{a^l} \cdot \frac{a^l}{a^l}
\]

Because the round brackets term equals \( d(-Y_c)/dR \), expression (7.4) follows. Finally, observe that

\[
\sum_{\delta} \left[ \frac{\beta_x \cdot d\log(b/k)}{d(-f)} \right] \frac{1}{\delta} \left[ \frac{\beta_y \cdot d\log(b/k)}{d(-f)} \right] = \frac{1}{\delta} \beta_y, \delta(k-b) \cdot \frac{d\log(b/k)}{d(-f)}
\]

which is zero by the FOC. QED

**Appendix D: The second order condition**

The SOC associated with problem (C.3) is equivalent to the positivity of

\[
\delta = \frac{\beta_x \cdot d\log(b/k)}{d(-Y_c)} - \frac{\beta_y \cdot d\log(b/k)}{d(-Y_c)} - \frac{1}{b} \frac{d\log(b/k)}{d(-Y_c)} - \frac{1}{k} \frac{d\log(b/k)}{d(-Y_c)}
\]

(D.1)
Because $\Lambda$ is by definition additively separable in $(-Y_1)$ and $\pi_1$, the convexity of preferences and the audit technology, imply that the second and fourth RHS terms are respectively negative and positive.

Consider the derivative in the first term. Using the definition of $b$ and taking into account that a change in the instrument will bring about adjustments in $X$ which in turn affects the marginal efficiencies, this derivative decomposes as

$$
\left(\frac{\partial \beta_i}{\partial X} \frac{\partial x}{\partial (-Y_1)} + \frac{\partial \beta_i}{\partial (-Y_1)} \frac{\partial x}{\partial X} \frac{\partial (-Y_1)}{\partial (-Y_1)} \right).
$$

(D.2)

The second round brackets term informs about the behaviour of the marginal deterrence efficiency of the tax rate as this instrument increases. The direct effect is given by

$$
\frac{\partial \beta_i}{\partial (-Y_1)} = \frac{E_{it}^{(1)}(MRS_t^{(1)}))^2 - 2E_{it}^{(1)}MRS_t^{(1)}}{w_2} + \frac{E_{it}^{(1)}}{u_c}\frac{u_{ct}^{(1)}MRS_t^{(1)}}{w_2} + \frac{u_{ct}^{(1)}MRS_t^{(1)}}{w_1}.
$$

(D.3)

which is a difference between two negative quadratic forms and cannot be signed. Applying expressions (4.2)-(4.3) on $d\beta_i/dX$, the indirect effect reduces to

$$
-\Delta \left( \frac{\partial x}{\partial Y_1} + \frac{\partial x}{\partial (-Y_1)} \right) - \Delta \left( \frac{\partial x}{\partial (-Y_1)} \right)^2 < 0.
$$

(D.4)

Assumption (E) in the text therefore amounts to a strong enough indirect effect.

The first round brackets term in (D.2) measures the spillover effect. It was decomposed in footnote 14 of the text, which is repeated for convenience:

$$
\frac{\partial \beta_i}{\partial (-Y_1)} = \frac{\partial \beta_i}{\partial \pi_1} = \left[ -D_{it}^{(1)}MRS_t^{(1)} + \frac{1}{w_2} \right] \Delta \frac{\partial x}{\partial (-Y_1)} \partial \pi_1.
$$

(D.5)

Under Edgeworth complementarity, the direct effect is positive, while normality of consumption and leisure, together with an upward sloping labour supply schedule guarantee a similar sign for the indirect effect (cf (4.3) and 4.4) in the text.

Finally, consider the third term in (D.1). Its derivative decomposes as follows:

$$
\left(\frac{\partial \beta_i}{\partial \pi_1} \frac{\partial x}{\partial \pi_1} - \frac{\partial \beta_i}{\partial \pi_1} \frac{\partial x}{\partial \pi_1} \right)
$$

(D.6)

The second round brackets term is the spillover effect, and because of symmetry, it is identical to expression (D.5). The first round brackets term measures the extent to which decreasing returns are present in the use of the audit policy. Because expected utility is linear in probabilities, the direct effect is zero. The indirect effect is always negative: using (4.4),

$$
\frac{\partial \beta_i}{\partial X} \frac{\partial x}{\partial \pi_1} = \Delta \left( \frac{\partial x}{\partial \pi_1} \right)^2 < 0.
$$

(D.7)

Gathering results, it follows that under the specified assumptions the SOC is satisfied.
Figure 1a Sequence of events in Mookherjee & Png (1989).

Figure 1b Sequence of events in this model.

Figure 2 Tax evasion opportunities leading to incentive incompatibility of a second-best tax system.

Figure 3 Pareto efficient and welfaristic responses to a lower penalty rate.
Figure 4. First-, second-, and third-best utility frontiers.

Figure 5. Inefficiency of case I.2 when evasion does not occur.

Figure 6a. Inefficiency of case I.2 when evasion does occur.

Figure 6b. Inefficiency of case I.2 when evasion does occur.
LIJST VAN RECENTE SESO-RAPPORTEN


KONINGS J., Do small plants really create most jobs? Testing convergence of establishments in the U.K. March 1994, 29 blz. (94/298)

CORTENS I., Herwaardering van het leerkrachtenambt, maart 1994, 56 blz. (94/299)

JANSSENS I., Macro-economische indicatoren van het RIZIV binnen het stelsel van de Sociale Zekerheid, april 1994, 40 blz. (94/300)


DE GRAEVE D. & W. NONNEMAN, Pharmacoeconomic studies: pitfalls and problems, June 1994, 21 blz. (94/303)

DE BORGER B. & K. KERSTENS, Produktiviteit en efficiëntie in de Belgische publieke sector: situering en resultaten van recent onderzoek, juni 1994, 40 blz. (94/304)

SCHROYEN F., Informational feasibility, decentralization and public finance mechanisms. July 1994, 35 blz. (94/305)


VANDENBUSSCHE H., Twin cases: Cartels and European antidumping policy. August 1994, 40 blz. (94/307)

VANHOUDT P., Wat denken economen over lange-termijn economische groei? Een overzicht, augustus 1994, 32 p. (94/308)

KESENNE S., Win maximization and the distribution of playing talent in professional team sports. September 1994, 22 p. (94/309)

KONINGS J., Hoe verschillen opvattingen over wat "fair" is tussen economen en niet-economen ? september 1994, 27 p. (94/310)