Society for the Study of Economic Inequality

Working Paper Series

Tax evasion and the optimal non-linear labour income taxation

Salvador Balle
Lucia Mangiavacchi
Luca Piccoli
Amedeo Spadaro

ECINEQ WP 2015 - 373

ECINEQ 2015 - 373
August 2015

www.ecineq.org

Tax evasion and the optimal non-linear labour income taxation*

Salvador Balle
Department of Physics, University of the Balearic Islands, Spain.

Lucia Mangiavacchi
Luca Piccoli†
Amedeo Spadaro
Department of Applied Economics, University of the Balearic Islands, Spain.

Abstract

The present work studies optimal taxation of labour income when taxpayers are allowed to evade taxes. The analysis is conducted within a general non-linear tax framework, providing a characterisation of the solution for risk-neutral and risk-averse agents. For risk-neutral agents the optimal government choice is to enforce no evasion and to apply the original Mirrlees' rule for the optimal tax schedule. The no evasion condition is precisely determined by a combination of a sufficiently large penalty and a constant auditing probability. Similar results hold for risk-averse agents. Our findings imply that a government aiming at maximizing social welfare should always enforce no evasion and provide simple rules to pursue this objective.

Keywords: Tax evasion, optimal taxation, social welfare.

JEL Classification: H21, H26, H31.

*The authors thank Pierre Boyer, Andreas Peichl, Eugenio Peluso, Sander Renes and Claudio Zoli for their precious comments and suggestions. We also thank the participants to the 30th Annual Congress of the European Economic Association, 6th Meeting of the Society for the Study of Economic Inequality, and to a seminar at the Economics Department, University of Verona. They are not to be held in any way responsible for the contents of the paper.

†Corresponding author e-mail: luca.piccoli@uib.es.

1 Introduction

The modern approach to modelling tax evasion within the optimal tax theory basically starts with Allingham and Sandmo (1972). Almost at the same time a seminal work by Mirrlees (1971) revolutionized previous research on the optimal taxation theory by introducing a framework for the optimal design of non-linear income taxation schedules. However, it was up to relatively recent times that the two approaches 'got married' (Cremer and Gahvari, 1995). Since then, however, very few works have further investigated this topic, shifting the attention towards tax avoidance, with a special interest on the optimal marginal tax rate of top tail of the income distribution.[1] Chetty (2009) generalizes Feldstein's 1999 formula for computing the deadweight loss in presence of avoidance by separating taxable income and total earned income elasticities, and show how the efficiency cost of taxing high income individuals may not be large. Piketty et al. (2014) develop a model where the top incomes respond to marginal tax rates through the labour supply elasticity, the avoidance elasticity and the bargaining elasticity. They argue that the third elasticity is the main source of response. Labour supply elasticity is generally small and is the sole real factor limiting the top tax rate. Finally, tax avoidance elasticity is the result of a poorly designed tax system, and can be confined to be close to zero[2] mostly by costless tax design reforms.

The issue of tax evasion is different in nature and has still to be explored within a fairly general formulation of the optimal taxation problem. Would results similar to tax avoidance still hold under a model of non-linear optimal labour income taxation *a la* Mirrlees with tax evasion? While the original goal of the present study was to completely characterize the non-linear optimal taxation schedule of labour income taxation when the consumer had the opportunity to evade taxes, our main result makes it an easy task. Indeed, we find that the optimal behaviour of the social planner is to enforce no evasion by setting the expected penalty incurred when concealing income larger than the expected benefit. As a consequence, the optimal tax system is the same as in Mirrlees' original problem.

The behavioural modelling of evasion started with Allingham and Sandmo (1972),[3] in which individuals are deterred from evasion by a fixed probability of auditing and a proportional penalty to be applied over and above the payment of the true liability. Within a linear tax system, taxpayers are risk averse and taxable income is exogenous. Yitzhaki (1987) extended the original model by assuming the probability of auditing to be an increasing function of evaded income, but with risk-neutral agents. Several authors will relax the risk-neutrality assumption and introduced expenditure on concealment (see Cowell, 1990; Kaplow, 1990; Cremer and Gahvari, 1994, among others).

A common weakness of these models is that they consider taxable income as exogenous, while it ap-

[1] Tax evasion is conceptually different from avoidance since it always imply an illegal action.
[2] Their empirical results confirm a small value of the avoidance elasticity for the US.
[3] For a more comprehensive review of tax evasion literature see Sandmo (2012), Sandmo (2005) and Slemrod and Yitzhaki (2002).

pears more reasonable to assume that income is generated by labour supply decisions given an exogenous earning capacity. This is the direction taken by Sandmo (1981), who argues that although results are more complex due to the increased number of choice variable in the model, the effects of the penalty and probability of detection are similar to previous models. Cremer and Gahvari (1994) introduced labour supply decision within a model of evasion, but remained within the linear tax model. It is only with Cremer and Gahvari (1995) and Schroyen (1997) that the endogenous labour supply decision has been embedded within a non-linear optimal income tax framework *à la* Mirrlees with evasion. The two articles share a two-skills economy, characterized by high- and low-wage individuals, but differ on the penalty mechanism assumptions. Although the two-skills assumption greatly simplifies the mathematical tractability of the problem, it also implicitly generates some the most relevant results of the papers, such as the optimality of the zero marginal tax rate for high-skill workers. As clearly explained by Piketty and Saez (2013) and Piketty et al. (2014), this famous result, already present in the original Mirrlees paper, is valid only for the unique richest individual in the economy, while progressive marginal tax rates at the top of the income distribution can still be optimal. This kind of reasoning can also be applied to Cremer and Gahvari (1995) and Schroyen (1997), since their high-wage individuals are also the richest individuals in their setting.

This weakness motivated us to build a more general non-linear optimal income tax model with evasion. We relax the two-skills economy assumption, and adopt a labour supply model for both risk-neutral and risk-averse agents. Respect to the previous literature, our study suggest that ethically desirable governmental behaviours –enforcing no evasion and choosing audited individual randomly– are optimal, while optimal taxation rules are the same of the original Mirrlees' model, and thus unaffected by tax evasion.

The remainder of the paper is organized as follows. Section 2 describes the main model and its assumptions. Section 3 discusses the incentive-compatibility constraints and the no evasion conditions. Section 4 presents the social planner's problem and its solutions for risk-neutral and risk-averse agents. Section 5 concludes the work.

2 General model and assumptions

We consider a society composed of rational individuals (agents) and a social planner (principal). The social planner seeks to reduce inequality. In order to do so, it is endowed with the right of collecting taxes from the individuals and redistributing a part of the tax collected. In addition, it has the power to establish an auditing system for controlling that individuals do not evade, and a penalty for evaders.

All individuals have the same utility function $U(C, L)$ that determines individual welfare as a function of the individual's consumption, C, and labour supplied, L. L and C can have any non-negative value, while U is continuous and sufficiently differentiable, increasing in C, decreasing in L and concave in C and L simultaneously.

Every individual is endowed with earning ability $w > 0$, which allows him to earn an income $Y = wL$, where L is the labour supplied by the individual. The earning ability w of each individual is private information.

Individuals consume all their net income and they behave rationally, i. e., in every instance they choose the consumption level and labour supply that maximize their utility function.

The social planner seeks to maximize social welfare by redistributing individuals' income. To this end, the social planner establishes an income tax system, $T(Y)$.

However, the social planner has no information about the real income Y earned by each individual. The real income Y of an individual can be observed by the social planner only through an auditing mechanism, which has an exogenous fixed cost κ.

In these conditions, individuals can evade taxes by reporting an income R different from their real income Y. In order to avoid it, the social planner also establishes a random auditing process, with auditing probability $p(R)\epsilon[0, 1]$. An individual does not know in advance whether or not he will be audited.

Audited individuals will pay the tax corresponding to their real income, plus a penalty $F(Y, R)$, that can depend, for example, on the evaded income $(Y - R)$ or the benefit of evasion $(T(Y) - T(R))$. The penalty $F(Y, R)$ expresses the disutility incurred by audited cheaters, and includes e.g. the economic cost of imprisoning, social repudiation, etc. Clearly, when the declared income is equal to the real income the fine is $F(Y, R) = 0$.

The social planner has to determine the tax and penalty systems and the auditing probability that maximize social welfare subject to a budgetary restriction and to the limitations imposed by the behavior of the agents.

3 Incentive-compatibility constraints

Through the tax schedule and the auditing and penalty mechanism, the social planner seeks to assign to each individual —characterized by its earning ability w— a specified income level $Y(w)$ and a reported income level $R(w)$. The tax system, the penalty and the auditing probability fix a functional relationship between income, reported income and consumption for both audited non-audited individuals, $C^A(Y, R)$ and $C^{NA}(Y, R)$, which we assume to be continuous and at least twice differentiable.

The consumption of audited and non-audited individuals are related to the tax schedule and the evasion penalty by

$$\begin{aligned} C^{NA}(Y,R) &= Y - T(R) \\ C^{A}(Y,R) &= Y - T(Y) - F(Y,R) \end{aligned}$$

Individuals behave rationally and they pick the income and reported income levels that maximize their utility. However, individuals do not know whether or not they are going to be audited by the social planner, hence they must decide their behavior based upon their expected utility. When individual w earns an income Y and reports an income R, his expected utility reads

$$\mathcal{U}(Y,R,w) = [1-p(R)]\, U\left(Y - T(R), \frac{Y}{w}\right) + p(R) U\left(Y - T(Y) - F(Y,R), \frac{Y}{w}\right) . \tag{1}$$

The social planner assigns to each individual w a given income $Y(w)$ and a reported income $R(w)$, hereby assigning to each an expected utility

$$V(w) = \mathcal{U}(Y(w), R(w), w) . \tag{2}$$

However, the agent —whose earning ability is private information and whose real income is unknown if not audited— may decide to earn the income of another type (ρ, say) while reporting the income of a third type, σ, if such a choice reports him a larger expected utility. Thus, the income and the reported income levels that the social planner assigns to individual w ought to verify the incentive-compatibility constraint (ICC)

$$V(w) \equiv \mathcal{U}\left(Y(w), R(w), w\right) \geq \mathcal{U}\left(Y(\rho), R(\sigma), w\right) \quad \forall w, \rho, \sigma . \tag{3}$$

In order to analyze the implications of (3), it is convenient to define the utility difference function $D(\rho,\sigma,w) = V(w) - \mathcal{U}(Y(\rho), R(\sigma), w)$, which must be non-negative everywhere while being zero along the diagonal $w = \rho = \sigma$. Hence, when one of the variables is fixed, D has to have a minimum on the diagonal with respect to the other two variables. This imposes the First-Order Incentive-Compatibility Conditions (FO-ICC)

$$[D_\rho] = 0 = [D_\sigma] = [D_w] \ , \tag{4}$$

where subscripts indicate partial derivation (i.e., $D_w \equiv \partial D/\partial w$) and the square brackets indicate that the magnitude is to be evaluated on the diagonal $\rho = \sigma = w$. By using the chain rule, eqs.(4) can be recast as

$$[\mathcal{U}_Y]\, Y'(w) \;=\; 0 \;\Rightarrow [\mathcal{U}_Y] = 0 \tag{5}$$

$$[\mathcal{U}_R]\, R'(w) \;=\; 0 \;\Rightarrow [\mathcal{U}_R] = 0 \tag{6}$$

$$V'(w) \;=\; \mathcal{U}_w \tag{7}$$

The former conditions define the extrema of $D(\bullet)$, among which maxima need to be selected for ensuring the ICC. Hence one must also impose the Second-Order Incentive-Compatibility Conditions (SO-ICC) that the Hessian of D, $H(D)$ is positive semi-definite on the diagonal, where

$$H(D) = \begin{pmatrix} [D_{\rho\rho}] & [D_{\rho\sigma}] & [D_{\rho w}] \\ [D_{\sigma\rho}] & [D_{\sigma\sigma}] & [D_{\sigma w}] \\ [D_{w\rho}] & [D_{w\sigma}] & [D_{ww}] \end{pmatrix} , \tag{8}$$

which implies that

$$[D_{\rho\rho}] \geq 0 \;, \left[D_{\rho\rho}D_{\sigma\sigma} - D^2_{\rho\sigma}\right] \geq 0 \;, \det H(D) \geq 0 \;. \tag{9}$$

The third SO-ICC is identically fulfilled, as can be easily demonstrated by deriving the FO-ICC in (4) with respect to w, which imposes that $[D_{w\sigma}] = \frac{1}{2}[D_{\rho\rho} - D_{ww} - D_{\sigma\sigma}]$, $[D_{\rho\sigma}] = \frac{1}{2}[D_{ww} - D_{\rho\rho} - D_{\sigma\sigma}]$, $[D_{w\rho}] = \frac{1}{2}[D_{\sigma\sigma} - D_{\rho\rho} - D_{ww}]$. Substitution into (8) gives that $detH(D) \equiv 0$, as should be expected from the constant zero value of $[D]$: the null curvature of $[D]$ must correspond to a zero eigenvalue.

3.1 No evasion condition

In the former discussion about the incentive-compatibility conditions, the income R that each agent reports is not necessarily equal to her/his real income Y. If the social planner wants the agents reporting their true income, i. e., $R = Y$ for all agents independently of their earning ability w, this option must be incentive-compatible. In this case, we must have that

$$\mathcal{U}\left(Y(w), Y(w), w\right) \geq \mathcal{U}\left(Y(w), R(\sigma), w\right) \quad \forall w, \sigma \;. \tag{10}$$

From eq.(1), we have that

$$\begin{aligned} \mathcal{U}(Y,R,w) \;&=\; [1 - p(R)]\, U\left(Y - T(R), \frac{Y}{w}\right) + p(R)U\left(Y - T(Y) - F(Y,R), \frac{Y}{w}\right) \\ &\leq\; U\left(Y - p(R)T(Y) - [1 - p(R)]\, T(R) - p(R)F(Y,R), \frac{Y}{w}\right) \end{aligned} \tag{11}$$

due to the concavity of $U(C, L)$.[4]

In addition, recalling that $F(Y, R) = 0$ if $Y = R$, we have that

$$\mathcal{U}\left(Y(w), Y(w), w\right) \quad = \quad U\left(Y(w) - T(Y(w)), \frac{Y(w)}{w}\right) \; . \tag{12}$$

Thus, since U is increasing on consumption, a sufficient condition for the no evasion condition (10) to be fulfilled is that

$$\begin{aligned} Y - T(Y) \quad &\geq \quad Y - p(R)T(Y) - [1 - p(R)]\, T(R) - p(R)F(Y, R) \\ &\Rightarrow \quad p(R)F(Y, R) \geq [1 - p(R)]\, [T(Y) - T(R)] \; , \end{aligned} \tag{13}$$

which has a clear economic interpretation: the expected penalty incurred by reporting a fake income has to be larger than the expected benefit. In these conditions, agents will report their true income regardless of their earning ability.

Equation (13) highlights two important features: first, the penalty is inversely proportional to the auditing probability; second the penalty is directly proportional to the avoided tax payments ($T(Y) - T(R)$) rather than to the sheltered income ($Y - R$).

3.2 Separable utility function

The simpler case of separable utility functions is widely used in the literature in order to describe the preferences of agents. In this case,

$$U(C, L) = K(C) - A(L) \; , \tag{14}$$

where $K(C)$ expresses the utility associated with consumption C, while $A(L)$ describes the disutility associated with labour. The general assumptions on U imply that in the present case, $K' > 0$, $K'' \leq 0$, $A' \geq 0$ and $A'' \geq 0$.

In this case, the expected utility function reads

$$\begin{aligned} \mathcal{U}(Y, R, w) \quad &= \quad [1 - p(R)]\, K(Y - T(R)) + p(R)K(Y - T(Y) - F(Y, R)) - A\left(\frac{Y}{w}\right) \\ &\equiv \quad Q(Y, R) - A\left(\frac{Y}{w}\right) \end{aligned} \tag{15}$$

[4]Equality in (11) holds when the utility function is lineal in consumption, and thus the higher bound for any concave utility function.

and the FO-ICC, eqs. (5-7), read

$$V'(w) = A'\left(\frac{Y(w)}{w}\right)\frac{Y(w)}{w^2} , \tag{16}$$

$$[\mathcal{U}_Y] = 0 \Rightarrow Q_Y(Y(w), R(w)) = \frac{1}{w}A'\left(\frac{Y(w)}{w}\right) , \tag{17}$$

$$[\mathcal{U}_R] = 0 \Rightarrow Q_R(Y(w), R(w)) = 0 . \tag{18}$$

A very important consequence of the separability of the utility function is that $D_{w\sigma} \equiv 0$, because D_w does not depend on R hence on σ. This implies, by deriving the FO-ICC with respect to w along the diagonal, that

$$[D_{\rho\rho} + D_{\sigma\rho} + D_{w\rho}] = 0 , \tag{19}$$

$$[D_{\rho\sigma} + D_{\sigma\sigma}] = 0 \Rightarrow [D_{\rho\sigma}] = [-D_{\sigma\sigma}] , \tag{20}$$

$$[D_{\rho w} + D_{ww}] = 0 \Rightarrow [D_{\rho w}] = -[D_{ww}] , \tag{21}$$

whence

$$[D_{\rho\rho}] = [D_{\sigma\sigma} + D_{ww}] . \tag{22}$$

As in the general case, $\det H(D) \equiv 0$ along the diagonal. In addition, using (20)-(22) into (9), we see that the other SO-ICC can be recast as

$$[D_{ww}] \geq 0 , \ [D_{\sigma\sigma}] \geq 0 . \tag{23}$$

Explicitly,

$$[D_{ww}] = V''(w) + \frac{Y(w)}{w^3}\left[2A'\left(\frac{Y(w)}{w}\right) + \frac{Y(w)}{w}A''\left(\frac{Y(w)}{w}\right)\right] \geq 0 , \tag{24}$$

$$[D_{\sigma\sigma}] = -Q_{RR}(Y(w), R(w))R'(w) \geq 0$$

$$\Rightarrow Q_{RR}(Y(w), R(w)) \leq 0 . \tag{25}$$

Conditions (18) and (25) reveal the behavior of the agents. In a first step, agents choose —through (18)— the reported income level R that maximizes their expected utility as a function of their (still unspecified) income level Y. For separable utility functions, this is independent of the agent's learning ability, w, thus the dependence of R on w arises only through the dependence of Y on w. In a second step, the income level Y of each agent is determined by (17).

Proposition 1 : The Spence-Mirrlees condition $Y'(w) \geq 0$ is a necessary and sufficient condition for (24).

Proof. Deriving (16) with respect to w, we have that

$$V''(w) = \frac{Y'(w)}{w^2}\left[A'\left(\frac{Y}{w}\right) + \frac{Y(w)}{w}A''(\frac{Y}{w})\right] - \frac{Y(w)}{w^3}\left[2A'(\frac{Y}{w}) + \frac{Y(w)}{w}A''(\frac{Y}{w})\right] ,$$

whence condition (24) can be recast as

$$\frac{Y'(w)}{w^2}\left[A'\left(\frac{Y}{w}\right) + \frac{Y(w)}{w}A''(\frac{Y}{w})\right] \geq 0 ,$$

which is identically satisfied if $Y'(w) \geq 0$. □

Conversely, if (24) is satisfied, it implies that $Y'(w) \geq 0$.

3.3 Quasi-linear utility function

A further simplification often studied in the literature is that of utility functions which are linear on consumption, i. e., $K(C) = C$. This implies risk neutral consumers, in the sense that the utility generated by an uncertain expected consumption level is the same of its certain equivalent. In this case,

$$Q(Y,R) = Y - [1-p(R)]\,T(R) - p(R)\,[T(Y) + F(Y,R)] \equiv \mathcal{C}(Y,R) , \tag{26}$$

which corresponds to the expected consumption, and

$$\begin{aligned} \mathcal{U}(Y,R,w) &= Y - [1-p(R)]\,T(R) - p(R)\,[T(Y) + F(Y,R)] - A\left(\frac{Y}{w}\right) \\ &\equiv \mathcal{C}(Y,R) - A\left(\frac{Y}{w}\right) . \end{aligned} \tag{27}$$

The FO-ICC then read

$$V'(w) = A'\left(\frac{Y(w)}{w}\right)\frac{Y(w)}{w^2} , \tag{28}$$

$$\mathcal{C}_Y(Y(w),R(w)) = \frac{1}{w}A'\left(\frac{Y(w)}{w}\right) , \tag{29}$$

$$\mathcal{C}_R(Y(w),R(w)) = 0 , \tag{30}$$

while the SO-ICC reduce to

$$Y'(w) \geq 0 \text{ and } \mathcal{C}_{RR}(Y(w), R(w)) \leq 0 \; . \tag{31}$$

4 Social planner's problem

The social planner is aware of the rational behaviour of individuals, and it knows the distribution of earning abilities, $f(w)$. The social planner seeks to establish the tax schedule, $T(R)$, the auditing probability, $p(R)\epsilon[0,1]$ and the penalty system, $F(Y,R)$ that lead to an income and reported income structures that maximize the social welfare

$$S = \int_0^\infty f(w)G(V(w))dw \; , \tag{32}$$

where G is the weight function that the social planner gives to individuals' welfare. A social planner who is adverse to inequality is characterized by $G' > 0$ and $G'' < 0$. The maximization is subject to a budgetary restriction that has to be verified by the taxes collected and includes the cost of the auditing,

$$\mathcal{B} = \int_0^\infty f(w) \left\{ [1 - p(R)] \, T(R) + p(R) \left[T(Y) + F(Y,R) - \kappa \right] \right\} dw \geq T_0 \; , \tag{33}$$

where T_0 is an exogenous revenue requirement for the government and κ is the (exogenous) cost of auditing an individual which we assume to be constant.

The cost of auditing can be assumed to be endogenous[5] and the analysis would be basically unchanged. To ensure the concavity of the Hamiltonian the endogenous cost of auditing $\kappa(Y)$ must be increasing and convex in income and this would lead to the same conclusions of the constant case, except that the optimal marginal tax rate should be increased by the expected marginal cost of auditing $p_0\kappa'(Y)$. Given that the additional contribution of having an endogenous cost of auditing is small and that the assumption of a constant cost of auditing is safe –at least for the analysis of personal labour income–, in what follows we maintain the hypothesis of constant cost of auditing.

4.1 Risk-neutral agents

The choice of $T(R)$, $p(R)$ and $F(Y,R)$ is constrained by the behavior of the individuals, which —under the assumption of separable, quasi-linear preferences— is described by the incentive-compatibility conditions

[5]For example Schroyen (1997) assumes κ to be an increasing and convex function of the probability of being audited conditional on income level $p(Y)$. In the two-skills economy this assumption ensures that auditing an entire income class is prohibitively costly. Adopting a similar assumption for an economy with a continuous income distribution makes no sense.

(28)-(31). Using (26) and (2) into (33), the problem can be recast as

$$\max_{Y,R} S = \int_0^\infty f(w)G(V(w))dw \tag{34}$$

subject to

$$\begin{aligned} \mathcal{B} &= \int_0^\infty f(w)\left[Y(w) - V(w) - A\left(\frac{Y(w)}{w}\right) - \kappa p(R(w))\right]dw \geq T_0 && (35) \\ V' &= \frac{Y}{w^2}A'\left(\frac{Y}{w}\right) && (36) \\ Y' &\geq 0\ , \quad \mathcal{C}_{RR} \leq 0\ . && (37) \end{aligned}$$

Introducing the (constant) Lagrange multiplier ν associated to the budgetary restriction and $\mu(w)$ as the costate variable associated to V, the Hamiltonian (assuming non-binding SO-ICC) reads

$$H = f(w)G(V) + \mu\frac{Y}{w^2}A'\left(\frac{Y}{w}\right) + \nu f(w)\left[Y - V - A\left(\frac{Y}{w}\right) - \kappa p(R)\right]\ . \tag{38}$$

Clearly, ν can be set to one without loss of generality by simply scaling $G(V)$ and μ.

Maximization with respect to R immediately yields that $p(R) = p_0$, constant. Maximization with respect to Y yields

$$\frac{\partial H}{\partial Y} = 0 \Rightarrow \frac{\mu}{w^2}\left[A'\left(\frac{Y}{w}) + \frac{Y}{w}A''(\frac{Y}{w}\right)\right] = -f(w)\left[1 - \frac{1}{w}A'\left(\frac{Y}{w}\right)\right]\ , \tag{39}$$

with the second order condition

$$\frac{\partial^2 H}{\partial Y^2} = \frac{\mu}{w^3}\left[2A''\left(\frac{Y}{w}\right) + \frac{Y}{w}A'''\left(\frac{Y}{w}\right)\right] - \frac{f(w)}{w^2}A''\left(\frac{Y}{w}\right) \leq 0\ . \tag{40}$$

In order to verify (40) it is sufficient that $\mu \leq 0$. Then, from (39) we have that

$$1 - \frac{1}{w}A'\left(\frac{Y}{w}\right) \geq 0$$

guarantees $\mu \leq 0$.

On the other hand, we have that

$$\mu' = -\frac{\partial H}{\partial V} = -f(w)\left[G'(V) - 1\right] \text{ with } \mu(\infty) = 0\ ,$$

hence

$$\mu = -\int_w^\infty f(x)\left[1 - G'(V(x))\right]dx\ . \tag{41}$$

Therefore, the requirement $\mu \leq 0$ implies that $G'(V) \leq 1$.

Thus, the income and expected utility of the agents are finally determined by equations (36) and (39), and read

$$\begin{aligned} V' &= \frac{Y}{w^2} A'(\frac{Y}{w}) \text{ with } V(0) = V_0 \text{ , } V(\infty) \text{ free ,} \\ 1 &= \frac{1}{w} A'(\frac{Y}{w}) + \frac{A'(\frac{Y}{w}) + \frac{Y}{w} A''(\frac{Y}{w})}{w^2 f(w)} \int_w^\infty f(x) \left[1 - G'(V(x))\right] dx \; . \end{aligned} \tag{42}$$

We thus see that the income assigned to every individual is given by exactly the same rule as in Mirrlees' original problem.

The penalty and optimal tax can then be determined from (29) and (30).

Proposition 2 : For a government aiming at maximizing social welfare of risk-neutral agents it is optimal to enforce no evasion. The penalty to apply should be proportional to the tax evasion $T(Y) - T(R)$ and inversely proportional to the auditing probability p_0.

Proof. From (30), we have that

$$\mathcal{C}_R = 0 \; \Rightarrow F_R(Y, R) = -\frac{1 - p_0}{p_0} T'(R) \; ,$$

hence one immediately finds that

$$F(Y, R) = \gamma(Y) - \frac{1 - p_0}{p_0} T(R) \; .$$

The requirement that $F(Y, Y) \equiv 0 \; \forall Y$ imposes that

$$F(Y, R) = \frac{1 - p_0}{p_0} \left[T(Y) - T(R)\right] \; , \tag{43}$$

□

Proposition 2 implies that $\mathcal{C} = Y - T(Y)$, hence $\mathcal{C}_{RR} \equiv 0$ and in addition it ensures that agents will report their true income (see eq. 13). Therefore, the penalty is proportional to the tax evasion, and the proportionality constant is $p_0^{-1} - 1$; this means that the penalty has to be substantially high if the fraction p_0 of audited individuals is low. It should also be noted that the constancy of $p(R)$ implies randomness in the auditing system. This conveys a sense of equality for all agents, which in democratic societies can help to enforce the tax and penalty schedule proposed by the social planner. This is in contrast with the previous literature, which under the simplifying assumption of a two-skills economy found that the richest should never be audited (Cremer and Gahvari, 1994; Schroyen, 1997) –an ethically controversial criterion.

Given the optimality of the no evasion condition defined by equation (43), the optimal marginal tax rate can be determined from equation (29).

Proposition 3 : Given Proposition 2, the optimal marginal tax rate in the case of separable quasi-linear utility functions is determined by the same rule of Mirrlees (1971).

Proof. Under enforcement of no evasion, as implied by equation (43), $\mathcal{C}_Y = 1 - T'(Y)$, thus by equation (29), the optimal marginal tax rate is defined by

$$1 - T'(Y(w)) = \frac{1}{w} A'\left(\frac{Y(w)}{w}\right) .$$

By equation (42) we obtain

$$\frac{1}{w} A'\left(\frac{Y(w)}{w}\right) = 1 - \frac{A'\left(\frac{Y(w)}{w}\right) + \frac{Y}{w} A''\left(\frac{Y(w)}{w}\right)}{w^2 f(w)} \int_w^\infty f(x)\left[1 - G'(V(x))\right] dx ,$$

and the optimal marginal tax rate reads

$$\begin{aligned} T'(Y(w)) &= 1 - \frac{1}{w} A'\left(\frac{Y(w)}{w}\right) \\ &= \frac{A'\left(\frac{Y(w)}{w}\right) + \frac{Y(w)}{w} A''\left(\frac{Y(w)}{w}\right)}{w^2 f(w)} \int_w^\infty f(x)\left[1 - G'(V(x))\right] dx . \end{aligned}$$

□

Thus, the possibility of agents reporting fake income levels will not change the Mirrlees' optimal marginal taxation rule in the case of separable quasi-linear utility functions.

Propositions 2 and 3 have important implications respect to the previous literature on income tax evasion, suggesting that avoiding the auditing of the richest individuals is not optimal, nor it is to apply lower marginal tax rates to the top of the distribution because of the evasion opportunity.

4.2 Risk-averse agents

Although the mathematical analysis of this case is more involved than that of risk-neutral agents, the same general results found in the previous section (optimality of enforcing no evasion and Mirrlees' tax structure) can also be derived in the general case. The underlying economic intuition is relatively straightforward.

Risk averse agents have a utility function which is increasing sublinearly in consumption and will always prefer a certain outcome over its uncertain equivalent, *ceteris paribus*. Assuming income tax evasion, the utility derived from the certain equivalent of the expected income is always higher, as formalized by equation (11).

The possibility of tax evasion by reporting fake income levels makes individuals who decide to evade wealthier. Hence, if the government 'permits' some level of evasion, this implicitly implies that the government has some budget margin to reduce tax revenues, which could be equally done by directly reducing tax rates. The latter case, however would imply a larger welfare increase because of the absence of uncertainty and the concavity of the utility function. As a consequence, enforcing no evasion is an optimal strategy for the government, and, since nobody would evade, the optimal marginal tax rate will be the classical Mirrlees' one.

Proposition 4 : For risk-averse agents enforcing no evasion is optimal and the rule that determines the optimal marginal tax rate is the same given in Mirrlees (1971).

Proof. Given a tax structure, auditing probability and penalty, the agent maximizes $\mathcal{U}(Y, R, w)$ respect to Y and R with the first order conditions (5) and (6). The optimal choice of Y and R will depend on the given scenario, and can be written as $Y^*(w; p(R), T(Y), T(R), F(Y, R), ...)$ and $R^*(w; p(R), T(Y), T(R), F(Y, R), ...)$.

On the other hand, the social planner seeks to choose $p(R)$, $T(R)$ and $F(Y, R)$ in order to maximize social welfare

$$S = \int_0^\infty f(w)G(\mathcal{U}(Y^*, R^*, w))dw$$

subject to the budgetary restriction

$$B = \int_0^\infty f(w)\{[1-p(R^*)]T(R^*) + p(R^*)[T(Y^*) + F(Y^*, R^*) - \kappa\}dw \ .$$

It is convenient to reformulate the problem as an optimal control problem, where the social planner maximizes

$$S = \int_0^\infty f(w)G(\mathcal{U}(Y, R, w))dw \ ,$$

with the expected (separable) utility function of the risk-averse agents (1)rewritten as

$$\mathcal{U}(Y, R, w) = [1-p(R)]\, K(n(Y, R)) + p(R)K(a(Y, R)) - A\left(\frac{Y}{w}\right) \ , \tag{44}$$

with $n(Y, R) = Y - T(R)$ being consumption if not audited and $a(Y, R) = Y - T(Y) - F(Y, R)$ being consumption if audited. The above maximization is subject to the restrictions imposed by the agents' behavior and the budgetary requirements:

$$\mathcal{U}_Y = 0 \tag{45}$$

$$\mathcal{U}_R = 0 \tag{46}$$

$$B = \int_0^\infty f(w)\,\{Y - [1-p(R)]n(Y, R) - p(R)[a(Y, R) + \kappa]\}\, dw \ .$$

To solve the problem, the social planner has to choose Y, R, n, a and $p(R)$, taking into account that (45) and (46) involve $n(Y, R)$ y $a(Y, R)$, which in turn determine the optimal tax schedule and the penalty.

Defining the expected utility as

$$V(w) = \mathcal{U}(Y(w), R(w), w) \ ,$$

and using equation (44) one can then express $a(Y,R)$ as

$$a(Y,R) = K^{-1}\left(\frac{V(w) + A\left(\frac{Y}{w}\right) - (1-p(R))K(n(Y,R))}{p(R)}\right) . \tag{47}$$

Moreover,

$$V'(w) = \frac{Y}{w^2}A'\left(\frac{Y}{w}\right) .$$

With this change equations (45) and (46) are identically verified, and the problem can be written as

$$S = \int_0^\infty f(w)G(V(w))dw$$

subject to the restrictions

$$\begin{aligned} V'(w) &= \frac{Y}{w^2}A'\left(\frac{Y}{w}\right) \\ B &= \int_0^\infty f(w)\left\{Y - [1-p(R)]n(Y,R) - p(R)[a(Y,R)+\kappa]\right\}dw , \end{aligned}$$

with $a(Y,R)$ given by (47).

Therefore, the Hamiltonian for this problem reads

$$H = f(w)G(V) + \mu\frac{Y}{w^2}A'\left(\frac{Y}{w}\right) + \lambda f(w)\left\{Y - [1-p(R)]n(Y,R) - p(R)[a+\kappa]\right\} ,$$

which needs being maximized over the control variables.

Remarkably, the dependency of the Hamiltonian on R comes only through a particular combination of $p(R)$ and $n(Y,R)$ that corresponds to the expected consumption $C = (1-p(R))n(Y,R) + p(R)a(Y,R)$, which can be written as a function of $n(Y,R)$

$$\Phi(n) = (1-p(R))n(Y,R) + p(R)K^{-1}\left(\frac{V + A\left(\frac{Y}{w}\right) - (1-p(R))K(n(Y,R))}{p(R)}\right)$$

which has a single minimum when $a(Y,R) = n(Y,R)$. Indeed,

$$\begin{aligned}\frac{d\Phi(n)}{dn} &= (1-p(R))\left[1 - \frac{K'(n(Y,R))}{K'\left(K^{-1}\left(\frac{V+A\left(\frac{Y}{w}\right)-(1-p(R))K(n(Y,R))}{p(R)}\right)\right)}\right] \\ &= (1-p(R))\left[1-\frac{K'(n(Y,R))}{K'(a(Y,R))}\right] \ ,\end{aligned}$$

which is zero if and only if $a(Y,R) = n(Y,R)$. Moreover this stationary point is a minimum since

$$\frac{d^2\Phi(n)}{dn^2} = -(1-p(R))\frac{K''(n(Y,R))}{K'(a(Y,R))} \geq 0 \ .$$

This implies that the maximization of H can be done in two steps: first, minimize $\Phi(n(Y,R))$ for any given Y, R and w, which implies $a(Y,R) = n(Y,R)$, and then maximize H for Y and R.

Note that setting $a(Y,R) = n(Y,R)$ corresponds to the no evasion condition (13), and then the Hamiltonian becomes

$$H = f(w)G(V) + \mu\frac{Y}{w^2}A'\left(\frac{Y}{w}\right) + \lambda f(w)\left[Y - n(Y,R) - p(R)\kappa\right] \ ,$$

with

$$V = K(n(Y,R)) - A\left(\frac{Y}{w}\right) \ .$$

Clearly, the problem is the non-linear utility equivalent of equation (38),

$$H = f(w)G(V) + \mu\frac{Y}{w^2}A'\left(\frac{Y}{w}\right) + \lambda f(w)\left[Y - K^{-1}\left(V + A\left(\frac{Y}{w}\right)\right) - p(R)\kappa\right] \ ,$$

and it corresponds to the classic Mirrlees' case except for the auditing cost $p(R)\kappa$.

Maximizing H immediately leads to $p(R) = p_0$, constant, and the optimal marginal tax rate follows Mirrlees' rule, provided that the utility function can be inverted. $\square$

This proves that the social welfare is maximized when there is no evasion, i.e. $Y = R$, and that this condition can be enforced *ex ante* by a random auditing process where the auditing probability is independent of the declared income. Under no evasion the optimal tax schedule can be determined and the penalty that satisfies the no evasion condition (13) can be chosen *ex post* according to the auditing probability p_0.

5 Discussion

5.1 Policy implications

Our results depart substantially from the previous literature on tax evasion within an optimal non-linear labour income taxation *á la* Mirrlees. Driven by the simplifying assumption of a "two-skills economy", Cremer and Gahvari (1995) and Schroyen (1997) found a zero optimal marginal tax rate for the rich to be optimal, who in addition should never be audited. Although this result was justified by the need of minimizing incentives to evade, they obtained a controversial result for a social welfare maximizing social planner, who typically accomplishes this task through redistribution. While the zero top income marginal tax rate was already present in the original Mirrlees (1971) work, it has been recently remarked (for example by Piketty and Saez, 2013) that this result holds only for the single richest individual in the income distribution.

By generalizing the non-linear optimal income taxation model to the whole earning ability distribution, we find that it is optimal for the social planner to enforce no evasion by applying a constant auditing probability, which means that the audited individuals are chosen randomly, independently on their income level. The penalty to be applied is proportional to the amount of taxes evaded and inversely proportional to the auditing probability. These results could have significant implications for the social planner, since an anti-evasion policy that is not targeted to a specific income group may be perceived as fair by the citizens and thus it may be easier to implement and to enforce.

Given the optimality of enforcing no evasion, the optimal marginal tax rule follows closely that of Mirrlees (1971), including the zero marginal tax rate for the single richest individual in the distribution, but allowing for progressivity of taxation at the top of the distribution. A similar result have been found by Piketty et al. (2014) analysing tax avoidance. Respect to the previous literature on tax evasion, that discouraged tax progressivity at the top of the income distribution, this is a desirable feature of our results, as far as social preferences support income redistribution.

5.2 Disproportionally large penalties

One of the implications of a penalty inversely proportional to the auditing probability is that small errors in the tax report might have dramatic consequences for agents who make errors in the declaration when the auditing probability chosen by the social planner is small. The implication of a penalty inversely proportional to the auditing probability are known at least since Becker (1968) and are relevant for the optimal choice of the government. A lower auditing probability reduces monitoring costs and thus the overall governmental budget constraint. On the other hand, imposing disproportionately large penalties implies that relatively small errors in the revenues declaration could imply dramatic consequences on the taxpayer. It is worth noting that from a theoretical perspective this class of models work with rational agents under perfect information, hence errors in the revenues declaration are ruled out by assumption. Nevertheless, in what follows we analyse the implication of a small error in the revenues declaration within our framework.

Let us consider the case of a tax scheme that has been designed according to the rules proposed in Section 4, and a honest agent that makes a mistake ξ when reporting his income, i. e., $R = Y + \xi, \xi < 0$. The expected utility of the agent is

$$\begin{aligned} V(w,\xi) &= (1-p_0)K\left(Y - T(Y+\xi)\right) \\ &+ p_0 K\left(Y - T(Y) - \frac{T(Y) - T(Y+\xi)}{p_0}\right) - A\left(\frac{Y}{w}\right) . \end{aligned}$$

For small enough mistakes, one has that

$$V(w,\xi) = V(w,0) + \xi p_0 K'\left(Y - T(Y)\right) T'(Y) + \ldots ,$$

which makes clear that the loss of expected utility incurred by unintentionally evading one income unit is given by $p_0 K'\left(Y - T(Y)\right) T'(Y)$. This corresponds to a loss $K'\left(Y - T(Y)\right) T'(Y)$ with probability p_0. Thus if the auditing probability is small the expected loss will be relatively small. Indeed, for small errors the penalty to be paid would be

$$F(Y, Y+\xi) = \frac{T(Y) - T(Y+\xi)}{p_0} = -\xi \frac{T'(Y)}{p0} ,$$

which can be large, but is paid only with probability p_0. Thus the expected penalty will be $E[F(Y, Y + \xi) = p_0 F(Y, Y+\xi) = -\xi T'(Y)$.

Although this does not rule out the issue, it makes clear that it will be relevant only for the possibly small fraction of agents who are audited and found not to declare all of their earned income. It should thus be possible and not excessively costly for the policy maker to implement a mechanism for distinguishing errors and frauds. While our model does not support a stochastic probability of incurring in some error in the revenue declaration, such a mechanism could in part be embodied through an increased cost of auditing κ, since to the cost of evaluating the true income of the taxpayer one would add the cost of establishing whether evasion was intentional or not.

Another concern with disproportionally large penalties is the consideration that an agent may not have sufficient means to pay the penalty if caught evading. This is equivalent to say that an agent would never consider a penalty larger than Y when forming his/her expected utility. Although the expected gain from evasion may be larger than the expected penalty in this case, in case of auditing this agent would not be able to consume anything and likewise would choose not to evade.

5.3 A comparison with the US framework

An example can be useful to understand how our results would compare to the real world. In the US tax frauds are punished by a fine equal to 75% of the underpayment, plus properties requisition, jail, and the cost of persecution if appropriate (IRS, 2015). Although the probability of auditing is not constant overall (see IRS, 2014, Table 9b), it is almost constant at about $p_0 = 0.7\%$ for a relatively wide range of gross declared income ($1\$ \leq R < 200,000\$$), almost 95% of all declarations. We could imagine the example of an inattentive self-employed earning about 50,000$ (Y) who could make an error in the revenue declaration and forget 1,000$, so he declares 49,000$ (R). Knowing that the marginal income tax rate ($T'(Y)$) for this bracket is 25%, the penalty to be paid if audited according to equation (43) would be almost 35,500$, which indeed is quite large. The expected penalty, on the other hand, would be 248.5$. Both of them would be large respect to 188$, the actual penalty that would be applied (excluding costs of prosecution) and to 1$, the expected penalty. This corresponds to a rather low incentive for declaring the true income respect to the almost 250$ of expected penalty –exactly the same as the underpayment– that would result from the no evasion condition.

Since the penalty rule (43) refers to a sufficient condition for the no-evasion condition (11) to be satisfied, one would expect the penalty needed to satisfy (11) to be smaller for risk averse agents. Indeed, when agents are risk-adverse, and thus endowed with a concave utility function, a somewhat smaller penalty may be sufficient to comply the no-evasion condition. Eventually, the penalty could be equal to that imposed by the US system if risk aversion is large. In this case, however, independently of the class of strictly concave utility function chosen to model agents' preferences, it is not generally possible to explicitly solve equation (11) for the penalty. Numerical simulations suggest that the concavity of the utility function produce a significantly different penalty only when the sheltered income is a large proportion of earned income. In our example, the difference in the penalty to be applied is negligible even for extremely risk averse agents.

6 Conclusions

Traditionally, the issue of optimal labour income taxation in presence of evasion has been studied within a linear tax framework and only two previous articles (Cremer and Gahvari, 1995; Schroyen, 1997) tried to address the non-linear income taxation case, with the simplifying assumption that individuals may be either low- or high-skill workers. Driven by this simplifying assumption, this stream of literature has confirmed the zero top income marginal tax rate result obtained by Mirrlees (1971). Indeed, in a two-skills economy, it is natural that the whole share of population with high skill is the richest by design. Only the rich have incentive to evade, and thus the zero marginal tax rate also minimizes the incentive to evade for this class of individuals.

Recent studies, however, have highlighted that the original result have been misinterpreted (i.e. Piketty and Saez, 2013), in the sense that it refers to the single richest individual in the income distribution, and have shown how progressive taxation can be optimal at the right tail of the income distribution as well. Moreover, analysing tax avoidance for the top income distribution Piketty et al. (2014) found that avoidance had basically no impact on the optimal taxation design.

In the present study we develop a theoretical model of optimal non-linear labour income taxation where agents are allowed to evade taxes by sheltering part of their earned income. The analysis is conducted within a general framework, providing a characterisation of the solution for risk-neutral and risk-averse agents. Our results give clear indications to a social planner aiming at maximizing social welfare when agents are allowed to evade income: to enforce no evasion through economic incentives (penalty) and to use of an income independent auditing probability for monitoring income declarations.

Respect to the previous literature, our study suggest that ethically desirable behaviours of a welfare maximizing social planner –enforcing no evasion and choosing audited agents randomly– are optimal. Moreover, under a general implementation of the optimal non-linear income tax framework with evasion, the optimal marginal tax schedule is the same as in the original Mirrlees problem and the government should enforce no evasion by imposing a sufficiently large penalty under a constant auditing probability. This implies that the results obtained using non-linear optimal taxation models, such as those presented in Piketty and Saez (2013) are valid also in presence of tax evasion.

References

Allingham, Michael G, and Agnar Sandmo (1972) 'Income tax evasion: A theoretical analysis.' *Journal of Public Economics* 1(3), 323–338

Becker, Gary S (1968) 'Crime and punishment: An economic approach.' *The Journal of Political Economy* 76(2), 169–217

Chetty, Raj (2009) 'Is the taxable income elasticity sufficient to calculate deadweight loss? the implications of evasion and avoidance.' *American Economic Journal: Economic Policy* pp. 31–52

Cowell, Frank Alan (1990) 'Tax sheltering and the cost of evasion.' *Oxford Economic Papers* pp. 231–243

Cremer, Helmuth, and Firouz Gahvari (1994) 'Tax evasion, concealment and the optimal linear income tax.' *The Scandinavian Journal of Economics* pp. 219–239

_ (1995) 'Tax evasion and the optimum general income tax.' *Journal of Public Economics* 60(2), 235–249

Feldstein, Martin (1999) 'Tax avoidance and the deadweight loss of the income tax.' *Review of Economics and Statistics* 81(4), 674–680

IRS (2014) 'Internal revenue service data book, 2014.' Internal Revenue System, Department of the Treasury of the US

_ (2015) 'Tax guide 2014. for individuals.' Internal Revenue System, Department of the Treasury of the US

Kaplow, Louis (1990) 'Optimal taxation with costly enforcement and evasion.' *Journal of Public Economics* 43(2), 221–236

Mirrlees, James A (1971) 'An exploration in the theory of optimum income taxation.' *The Review of Economic Studies* pp. 175–208

Piketty, Thomas, and Emmanuel Saez (2013) 'Optimal labor income taxation.' *Handbook of Public Economics* 5, 391–474

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva (2014) 'Optimal taxation of top labor incomes: A tale of three elasticities.' *American Economic Journal: Economic Policy* 6(1), 230–71

Sandmo, Agnar (1981) 'Income tax evasion, labour supply, and the equityefficiency tradeoff.' *Journal of Public Economics* 16(3), 265–288

_ (2005) 'The theory of tax evasion: A retrospective view.' *National Tax Journal* pp. 643–663

_ (2012) 'An evasive topic: theorizing about the hidden economy.' *International Tax and Public Finance* 19(1), 5–24

Schroyen, Fred (1997) 'Pareto efficient income taxation under costly monitoring.' *Journal of Public Economics* 65(3), 343–366

Slemrod, Joel, and Shlomo Yitzhaki (2002) 'Tax avoidance, evasion, and administration.' *Handbook of public economics* 3, 1423–1470

Yitzhaki, Shlomo (1987) 'On the excess burden of tax evasion.' *Public Finance Review* 15(2), 123–137