

Blomquist, Sören; Micheletto, Luca

Working Paper

Age Related Optimal Income Taxation

Working Paper, No. 2003:7

Provided in Cooperation with:

Department of Economics, Uppsala University

Suggested Citation: Blomquist, Sören; Micheletto, Luca (2003) : Age Related Optimal Income Taxation, Working Paper, No. 2003:7, Uppsala University, Department of Economics, Uppsala, <https://nbn-resolving.de/urn:nbn:se:uu:diva-4461>

This Version is available at:

<https://hdl.handle.net/10419/82774>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Age Related Optimal Income Taxation*

Sören Blomquist[†] and Luca Micheletto[‡]

February 7, 2003

Abstract

The focus of the present paper is on the intragenerational effects of nonlinear income taxation in a multiperiod framework. We investigate whether it is possible to achieve redistribution at smaller efficiency costs by enlarging the message space adopted in standard tax system (which only includes reported income) to consider also the age of taxpayers. Since it would be awkward to analyze an age related tax without taking into account the time-dimension, we use an intertemporal extension of the Stiglitz-Stern (1982, 1982) discrete adaptation of the Mirrlees (1971) optimal income taxation model. In the simplest version of the model we neglect the possibility of savings. This case can be interpreted as a situation with extreme liquidity constraints. It is shown that switching to an age related tax system opens the way for a Pareto improving tax reform entailing a cut in marginal tax rates for young agents. In a second version of the model we retain the possibility of savings and, assuming that the policy maker can tax interest incomes on a linear scale, we also analyze the optimal values of the interest income tax rate for the age dependent and the age independent tax systems.

Keywords: Optimal taxation; Age specific taxes; Tagging.
JEL Classification: H21; H23; H24.

*We would like to thank Vidar Christiansen and seminar participants at University of Uppsala, University of Copenhagen (EPRU), PET 2002 in Paris and IIPF 2002 in Helsinki for their valuable comments. Usual disclaimer applies.

[†]Nationalekonomiska institutionen, Ekonomikum, Uppsala University, Kyrkogårdsgatan 10, P.O. Box 513, SE-751 20, Uppsala, Sweden. E-mail address: soren.blomquist@nek.uu.se; tel.: ++46 18 4711102; fax: ++46 18 4711478.

[‡]Istituto di Economia Politica, Università “L. Bocconi”, via U. Gobbi 5, 20136 Milano, Italy. E-mail address: luca.micheletto@uni-bocconi.it; tel. ++39 02 58365324; fax ++39 02 58365318.

1 Introduction

Many countries have ambitious redistributive goals implying high marginal income tax rates and inefficiencies. There is therefore a continuous ongoing search for taxes that can achieve redistribution with smaller efficiency losses. Akerlof (1978) pointed out that if the tax system can be differentiated between individuals according to some characteristic correlated with ability, then there is a potential for reducing the conflict between redistribution and efficiency.¹ Since in most countries average income varies systematically with age, in this paper we will investigate if it is possible to achieve redistribution at smaller efficiency costs by relating the tax payments to the age of persons.²

The design of the tax system will depend on the objective of the social planner. There is a strong case for age dependent taxes if the social planner is concerned with annual utilities. If skill level and age covary perfectly, an age tax could in this case achieve the first best. However, if the social planner is concerned with lifetime utilities, then, if age and skill covary perfectly, individuals have identical income paths and no redistribution is needed. If there is no covariation, there will be no gains either. In the intermediate case, with some covariation, there might be a role for an age related tax. In this paper we will be concerned with this latter case.

When designing a model to analyze the potential benefits of an age dependent income tax there are several important modelling choices. Perhaps the most fundamental is whether to use an atemporal or an intertemporal model. In our view it would be awkward to analyze an age dependent tax without taking into account that age has a time-dimension. Individuals use intertemporal reallocations of consumption possibilities and, as we will see, this savings behavior has important implications for how the tax system should be designed. We therefore adopt an intertemporal model where individuals can save.

As a vehicle for our analysis we will use an extension of the Stiglitz-Stern (1982, 1982) simplified “two-types” version of the Mirrlees (1971) optimal income taxation model. We construct an overlapping generations model with individuals living for two periods and facing a consumption-leisure choice in

¹The argument is straightforward: we know that, equity and feasibility aside, the best tax is a lump-sum tax; therefore, if the government is engaged in redistribution but cannot observe ability directly, efficiency is maximized by taxing objects which the individual cannot affect (or activities that go on the same regardless of the tax) and which are correlated with the skill level.

²This in turn raises the question of horizontal equity; Picard (2001) provides a recent example where it is claimed that in the income taxation literature the choice of the control variables is restricted by horizontal equity and that the design of income tax based on age is considered not ethical. On the other way, the relevance of the concept of horizontal equity for normative purposes has recently been questioned in a series of articles by Kaplow (1989, 1995, 2000) and Kaplow and Shavell (2000, 2001).

each period. All individuals are low skilled in the first period of life. In the second period the proportion π^l stays low skilled and the proportion π^{lh} becomes high skilled. This means that individuals have different income paths and we would like to redistribute from the lh to the ll individuals. Ideally in this situation we would like to tax some index of lifetime income. However, real-life tax systems exclusively use annual income as tax base. We therefore impose the restriction in our analysis that annual income is the tax base. Hence, the social planner is concerned with lifetime utilities but can only use taxes imposed on annual income.

We will consider the design of the tax system under various observational assumptions. The first tax system that we consider is such that the social planner knows the joint age-skill distribution but is restricted to offer the same (labor income, disposable income)-bundle to all agents of a given skill level, irrespective of their age. Two income points are designed; one intended for the low skilled and another intended for the high skilled. The second tax system is designed under the assumption that the social planner, knowing the mechanism that some low skilled persons become high skilled in the second period, attempts to design three income points: one for young low skilled persons, another for old low skilled persons and a third for (old) high skilled persons. However, it is assumed that the planner cannot observe the age of individuals. Depending on the assumptions we make with regard to the individual preferences, the tax system will in one case collapse to a two points system, while in another case three points will be used. Finally, in the third tax system the assumption of non-observability of age is removed and the planner designs three income points.

There are many studies using an OLG model with a homogeneous population and no intragenerational redistribution. These models often focus on growth related issues. In this paper we focus on the intragenerational redistributive effects of income taxation in a multiperiod framework. This complicates the model. To keep the analysis manageable we simplify and completely abstract from all growth related issues. To illustrate the basic workings of our model, in its simplest version we also abstract from savings. Since in our model a likely case is that where individuals would like to borrow in the first period of life, the no savings case can also be interpreted as a situation with extreme liquidity constraints.

There are also some studies using an OLG model with heterogeneous skill levels (Brett (1998), Pirttilä and Tuomala (1998)). These articles differ from our paper as they focus on the division of life into a working period and a retirement period. In our model we focus on the fact that individuals have different income paths and that inequality widens with the age of cohorts.

The possibility of lump-sum taxation with some redistributive power has earlier been discussed by, for example, Hahn (1973), Akerlof (1978) and Viard (2001). In his concluding remarks Hahn (1973) argues that lump sum taxes are available. However, the examples he gives of lump sum taxes

used in the past seem of little relevance for taxation today. Akerlof (1978) discusses how the tax-transfer system can be made more efficient if a “needy group” can be identified, tagged, and given a tax-transfer system of its own. This idea is also pursued in Immonen et al. (1998). Kremer (2002) investigates the conditions under which an age dependent income tax might be beneficial. However, he considers the case where annual, and not lifetime, utilities enter the social welfare function. Kremer presents empirical data that support the idea that an age dependent income tax would be of value. The idea of making the income tax age dependent has also been suggested in a recent article in *Fortune* by Mankiw (1998).

The paper is organized as follows. In Section 2 we present the model in its general form. Section 3 deals with the simplified case where there are no opportunities to save. We show how observability of age allows a Pareto improvement upon an optimal income tax system that does not condition the tax on age. In Section 4 we assume access to the international capital market, so that savings are possible and the problem of interest income taxation is investigated. Section 5 provides some additional comments and Section 6 concludes the paper.

2 The Model

The economy is described by an OLG model where there is no population growth, each cohort consists of a large number of individuals and its size is normalized to one. All agents are low skilled (earning a unitary wage w_l) when “young” (in the first period of life) and each agent faces an exogenous probability π^{lh} to become high skilled (earning a unitary wage w_h) in the second period of his/her life; therefore, by the assumption of a large number of households, the proportions of low- and high skilled individuals among “old” people are given by $\pi^{ll} = 1 - \pi^{lh}$ and π^{lh} respectively.

All agents are *ex ante* identical and derive utility from consumption when young (c_y) and consumption when old (c_o). Moreover, they get disutility from labor supplied when young (L_y) and when old (L_o). Lifetime utility is represented by the additive separable quasi-concave utility function $U = u(c_y, L_y) + \frac{1}{1+\rho}u(c_o, L_o)$, which is assumed to be identical across households and where ρ is a rate of time preference.³ At the rate of interest r prevailing in the credit market, agents are free to save (borrow) in the first period of their life in order to finance future (present) consumption. Labor income ($I = wL$) is assumed to be taxed on a nonlinear scale through a general income tax function $T(I)$ and interest incomes (rs) on a linear scale through a residence-based taxation at a proportional rate t (with full offsets for net interest paid). Production is linear and uses labor as the only factor.⁴

³Notice that preferences are constrained to be age-independent.

⁴Since interest income taxation can be interpreted as a special case of commodity

The government's problem is to maximize the *ex post* utility of those who remain low skilled subject to the constraint of a minimum level of utility to those who become high skilled, a set of incentive-compatibility constraints, a balanced budget constraint and the resource constraint of the economy.⁵ Notice that this implies that the objective function of the government does not coincide with what young people actually maximize (i.e. *ex ante* expected utility).⁶ This would not have been the case if we had assumed the view that the policy maker was concerned with maximizing *ex ante* expected utility (as for instance done by Cremer and Gahvari (1995)) or if we had assumed that people perfectly knew in the first period of their life what would have been their skill level in the second period.

Looking at the consumer's behavior and denoting by $B = I - T(I)$ the after tax labor income, we have that the level of consumption in the second period of life will be $c_o^{lh} = s(1 + r(1 - t)) + B_o^{lh}$ for those who turn out to be high skilled when old, and $c_o^{ll} = s(1 + r(1 - t)) + B_o^{ll}$ for those who turn out to remain low skilled. In the first period of life the conditional demand for savings of an expected utility maximizing agent can be written with obvious notation as $s = B_y - c_y = s(\pi^{ll}, \rho, r(1 - t), B_y, B_o^{ll}, B_o^{lh}, I_y, I_o^{ll}, I_o^{lh})$.

The households' problem is

$$\begin{aligned} \max_{c_y, I_y} u\left(c_y, \frac{I_y}{w_l}\right) &+ \frac{\pi^{ll}}{1 + \rho} u\left((1 + r(1 - t))[I_y - T(I_y) - c_y] + I_o^{ll} - T(I_o^{ll}), \frac{I_o^{ll}}{w_l}\right) + \\ &+ \frac{1 - \pi^{ll}}{1 + \rho} u\left((1 + r(1 - t))[I_y - T(I_y) - c_y] + I_o^{lh} - T(I_o^{lh}), \frac{I_o^{lh}}{w_h}\right). \end{aligned}$$

The first order conditions are the following:

$$\frac{\partial u(\bullet)}{\partial c_y} = \frac{1 + r(1 - t)}{1 + \rho} E \frac{\partial u(\bullet)}{\partial c_o}, \quad (1)$$

$$\frac{\partial u(\bullet)}{\partial I_y} = -\frac{1 + r(1 - t)}{1 + \rho} (1 - T_y') E \frac{\partial u(\bullet)}{\partial c_o} \quad (2)$$

Combining (1) and (2), we have that the marginal income tax rate faced by young people is implicitly given by

taxation, proportional interest income taxation can be justified by referring to the attempt to limit the scope for arbitrage opportunities among agents (see Hammond (1987) for the general theory of the desirability of linear pricing when commodities are exchangeable on side markets, and Lindencrona (1993) for an application to the topic of taxation of capital income).

⁵The pre-set level of utility for agents of type *lh* will be hereafter selected in such a way that redistribution goes from the high- to the low-wage households, what in literature is referred to as the "normal" case.

⁶This circumstance explains also why at some points the standard way of addressing the optimal taxation problem is not applicable and the analysis becomes more complex.

$$\frac{\partial u(\bullet)}{\partial I_y} = - (1 - T'_y) \frac{\partial u(\bullet)}{\partial c_y} \implies 1 + \frac{\frac{\partial u(\bullet)}{\partial I_y}}{\frac{\partial u(\bullet)}{\partial c_y}} = T'_y. \quad (3)$$

Since savings are given for old individuals, the marginal income tax rates faced by old low skilled and old high skilled agents are implicitly given by respectively:

$$T'_{o(low)} = 1 + \frac{\frac{\partial u_o^l}{\partial I_o^l}}{w^l \frac{\partial u_o^l}{\partial c_o^l}} = 1 + \frac{\frac{\partial u_o^l}{\partial I_o^l}}{\frac{\partial u_o^l}{\partial c_o^l}}, \quad (4)$$

$$T'_{o(high)} = 1 + \frac{\frac{\partial u_o^h}{\partial I_o^h}}{w^h \frac{\partial u_o^h}{\partial c_o^h}} = 1 + \frac{\frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}}. \quad (5)$$

It will be convenient to look by now in more details at the optimal level of savings chosen by young agents. Having denoted by q the net rate of return on savings ($q = r(1 - t)$), the second order condition that must be satisfied for an optimal level of consumption is $D = \frac{\partial u_y}{\partial c_y \partial c_y} + \frac{(1+q)^2}{1+\rho} E \frac{\partial u_o}{\partial c_o \partial c_o} < 0$.

Implicit differentiation of (1) gives the following comparative statics results which we will use in the next sections:

$$\frac{dc_y}{dB_y} = \frac{\frac{(1+q)^2}{1+\rho} E \frac{\partial^2 u_o}{\partial c_o \partial c_o}}{D} > 0; \quad (6)$$

$$\frac{dc_y}{dB_o^l} = \frac{\frac{1+q}{1+\rho} \pi^{ll} \frac{\partial^2 u_o^l}{\partial c_o^l \partial c_o^l}}{D} > 0; \quad (7)$$

$$\frac{dc_y}{dB_o^h} = \frac{\frac{1+q}{1+\rho} \pi^{lh} \frac{\partial^2 u_o^h}{\partial c_o^h \partial c_o^h}}{D} > 0; \quad (8)$$

$$\frac{dc_y}{dI_y} = - \frac{\frac{\partial^2 u_y}{\partial c_y \partial I_y}}{D}; \quad (9)$$

$$\frac{dc_y}{dI_o^l} = \frac{\frac{1+q}{1+\rho} \pi^{ll} \frac{\partial^2 u_o^l}{\partial c_o^l \partial I_o^l}}{D}; \quad (10)$$

$$\frac{dc_y}{dI_o^h} = \frac{\frac{1+q}{1+\rho} \pi^{lh} \frac{\partial^2 u_o^h}{\partial c_o^h \partial I_o^h}}{D}; \quad (11)$$

$$\frac{dc_y}{dq} = \frac{\frac{1}{1+\rho} E \frac{\partial u_o}{\partial c_o} + s \frac{1+q}{1+\rho} E \frac{\partial^2 u_o}{\partial c_o \partial c_o}}{D}. \quad (12)$$

The first three inequalities follow from the assumption that c_y is a normal good. As regards the sign of (9), (10), (11) and (12), we have that

$\frac{dc_y}{dI_y} > (<) 0$, $\frac{dc_y}{dI_o^l} < (>) 0$, $\frac{dc_y}{dI_o^h} < (>) 0$ if consumption and leisure are Edgeworth substitutes (complements) in u . Finally, notice that the sign of (12) will be unambiguously negative for a borrower, since for such an individual income and substitution effects push in the same direction, while it becomes ambiguous for a lender, depending on the relative magnitudes of the substitution effect (negative) and the income effect (positive).

3 The Model without Savings

In this Section we start the analysis with a simple framework where individuals cannot save. This proves to be a useful starting point since the possibility to neglect the problem of savings and the related problem of the optimal interest income tax simplify matters remarkably and allows us to get sharp results. We also assume the productive technology is linear and uses effective labor as the only productive factor. Thus, the production function can be described as:

$$Q = w_l (L_y + \pi^{ll} L_o^l) + w_h \pi^{lh} L_o^h.$$

3.1 Case 1: The Government Does not Try to Set up a Three Points System

The first case we explore is when the government knows the joint distribution of skill and age but it doesn't try to set up a three income points system, i.e. it offers the same (labor income, disposable income)-bundle to young agents and old low skilled ones.

Denoting by V indirect utilities and by a “hat” a variable when referred to a mimicker, the government's problem is the following:

$$\max_{B^l, B^h, I^l, I^h} V^l(B^l, I^l) + \frac{1}{1+\rho} V^l(B^l, I^l)$$

subject to

$$V^l(B^l, I^l) + \frac{1}{1+\rho} V^h(B^h, I^h) \geq \bar{V}, \quad (\lambda)$$

$$V^h(B^h, I^h) \geq \widehat{V}^h(B^l, I^l), \quad (\mu)$$

$$(1 + \pi^{ll}) (I^l - B^l) + \pi^{lh} (I^h - B^h) \geq 0, \quad (\theta)$$

where $\bar{V}$ is a pre-set utility level and Lagrange multipliers are within parentheses.⁷

The first order conditions are the following:

$$B^l : \left(\frac{2+\rho}{1+\rho} + \lambda \right) V_B^l = \mu \widehat{V_B^h} + \theta (1 + \pi^l) \quad (13)$$

$$B^h : \left(\frac{\lambda}{1+\rho} + \mu \right) V_B^h = \theta \pi^{lh} \quad (14)$$

$$I^l : \left(\frac{2+\rho}{1+\rho} + \lambda \right) V_I^l = \mu \widehat{V_I^h} - \theta (1 + \pi^l) \quad (15)$$

$$I^h : \left(\frac{\lambda}{1+\rho} + \mu \right) V_I^h = -\theta \pi^{lh} \quad (16)$$

Dividing (16) by (14) and using condition (5), we get the usual result of “no distortion at the top”:

$$\frac{\partial T(I^h)}{\partial I^h} = 0, \quad (17)$$

which here means that the labor/leisure choice of those who turn out to be high skilled in the second period of their life should not be distorted at the margin.

On the other hand, dividing (15) by (13), we have that young people and those who remain low skilled in the second period of their life face a positive (due to the standard assumption of single-crossing) marginal income tax rate given by

$$\frac{\partial T(I^l)}{\partial I^l} = \frac{\mu \widehat{V_B^h}}{\theta (1 + \pi^l)} \left(\frac{\widehat{V_I^h}}{\widehat{V_B^h}} - \frac{V_I^l}{V_B^l} \right), \quad (18)$$

where the term inside brackets represents the difference between the marginal valuation of leisure in terms of consumption for a low skilled agent and a mimicker. Denoting by $MRS_{I,B} = -\frac{V_I}{V_B}$ the marginal rate of substitution between labor and consumption, we can also write condition (18) as:

$$\frac{\partial T(I^l)}{\partial I^l} = \frac{\mu \widehat{V_B^h}}{\theta (1 + \pi^l)} \left(MRS_{I,B}^l - \widehat{MRS}_{I,B} \right).$$

⁷Notice that the fact that the government knows the joint distribution of skill and age affects the way the set of participation constraints is shaped by determining the cardinality of this set.

3.2 Case 2: Age is not Observable but the Government Tries to Set up a Three Points System

The second case we consider is the one where, even if age is not directly observable, the government tries to use the information on how skills are distributed across age groups to set up a three income points system, i.e. it tries to offer three different points in the (I, B) -space. In this case the government's problem would be as follows:

$$\max_{B_y, B_o^l, B_o^h, I_y, I_o^l, I_o^h} V^l(B_y, I_y) + \frac{1}{1+\rho} V^l(B_o^l, I_o^l)$$

subject to

$$V^l(B_y, I_y) + \frac{1}{1+\rho} V^h(B_o^h, I_o^h) \geq \bar{V}, \quad (\lambda)$$

$$V^h(B_o^h, I_o^h) \geq \widehat{V}^h(B_o^l, I_o^l), \quad (\mu)$$

$$V^h(B_o^h, I_o^h) \geq \widehat{V}^h(B_y, I_y), \quad (\eta)$$

$$V^l(B_y, I_y) \geq \widehat{V}^l(B_o^l, I_o^l), \quad (\phi)$$

$$V^l(B_o^l, I_o^l) \geq \widehat{V}^l(B_y, I_y), \quad (\varphi)$$

$$(I_y - B_y) + \pi^{ll} (I_o^l - B_o^l) + \pi^{lh} (I_o^h - B_o^h) \geq 0. \quad (\theta)$$

Together, constraints (ϕ) and (φ) imply that the utility that agents get in the first period of life must be equal to the utility obtained in the second period of life by those who remain low skilled. Notice that this alone is not sufficient to conclude that the level of consumption and labor supply of a young agent is the same as the one of an old low skilled person. Moreover, different consumption-leisure bundles that are equally preferred by a low skilled agent will be in general not indifferent when evaluated by a high skilled agent acting as a mimicker. However, it can be easily proved that an optimum is only compatible with both the constraints (μ) and (η) binding at the same time, which in turn means that not only the consumption-leisure bundles for a young household and for an old low skilled one should lie on

the same indifference curve of a low skilled agent, but that those bundles should also lie on the same indifference curve of a high skilled agent: by single-crossing this can happen only if the two bundles are actually the same bundle. Otherwise, a tax reform can be implemented that leaves each agent at the same utility level and that at the same time generates additional revenue to the government (see fig. 2-5 in Appendix).

This means that the policy maker is actually offering only two points in the (I, B) -space and therefore we are back to case 1. Notice also that such a result is due to the fact that, having assumed age-independent preferences and ruling out the possibility of savings, at any given point in the (I, B) -space the indifference curve for a young agent has the same slope as the one for an old low skilled agent. In a more general context this property would not hold and a policy maker would actually do better by trying to set up a three income points system even if age were not observable.

3.3 Case 3: Age is Observable and the Government Tries to Set up a Three Points System

Before presenting the analysis of the optimal income tax system, we show how the observability of age allows to Pareto improve upon the optimal tax system where age is not observable. Fig. 1 gives an example of a Pareto-improving tax reform that could be implemented by conditioning the income tax schedule to the age of individuals.

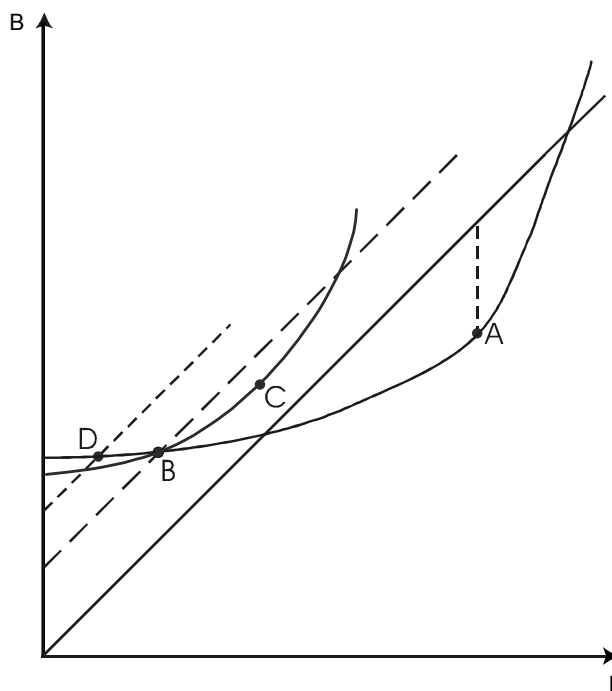


Figure 1: A Pareto improving tax reform

The reform can be illustrated as follows. The two income points system consists of points A and B . High skilled people are located at point A and young and old low skilled are bunched at point B . The indifference curve going through point A indicates the utility obtained for a high skilled person in second period of life. The indifference curve going through points B and C indicates the utility obtained by a low skilled person in first period of life. However, since the utility function for the second period is just the first period utility function multiplied by a positive number, it also represents the indifference curve indicating the utility level obtained for a low skilled person in second period of life (thus, there are two separate utility levels represented by the same indifference curve). Given that the tax can be conditioned on age a strict Pareto improvement can be obtained in the following way. Offer the young low skilled the point C , where they have the same utility as at B . However, at point C their leisure-consumption choice is undistorted. This implies that resources are released so that old low skilled people can be located at point D , where they obtain a higher utility than in the two income points system. In terms of lifetime utilities the expected lifetime utility of individuals has gone up. The actual lifetime utility of people being low skilled in both periods has increased whereas the lifetime utility of those who are high skilled in the second period is unchanged. The changes in consumption and work are as follows. The old high skilled would perform as before. The young low skilled would work more, have higher labor income and consume more. The old low skilled would work less and also have less consumption.

We next consider the optimal tax. When the policy maker can observe age and uses the information on the correlation between skill and age in order to optimally shape the income tax schedule, the government's problem becomes

$$\max_{B_y, B_o^l, B_o^h, I_y, I_o^l, I_o^h} V^l(B_y, I_y) + \frac{1}{1+\rho} V^l(B_o^l, I_o^l)$$

subject to

$$V^l(B_y, I_y) + \frac{1}{1+\rho} V^h(B_o^h, I_o^h) \geq \bar{V}, \quad (\lambda)$$

$$V^h(B_o^h, I_o^h) \geq \widehat{V}^h(B_o^l, I_o^l), \quad (\mu)$$

$$(I_y - B_y) + \pi^{ll} (I_o^l - B_o^l) + \pi^{lh} (I_o^h - B_o^h) \geq 0, \quad (\theta)$$

where in writing the self-selection constraints we have used the property that mimicking cannot occur between agents at different points in their lifetime.

The first order conditions are the following:

$$B_y : (1 + \lambda) V_B^{y(l)} = \theta \quad (19)$$

$$B_o^l : \frac{1}{1 + \rho} V_B^{o(l)} = \mu \widehat{V}_B^h + \theta \pi^{ll} \quad (20)$$

$$B_o^h : \left(\frac{\lambda}{1 + \rho} + \mu \right) V_B^{o(h)} = \theta \pi^{lh} \quad (21)$$

$$I_y : (1 + \lambda) V_I^{y(l)} = -\theta \quad (22)$$

$$I_o^l : \frac{1}{1 + \rho} V_I^{o(l)} = \mu \widehat{V}_I^h - \theta \pi^{ll} \quad (23)$$

$$I_o^h : \left(\frac{\lambda}{1 + \rho} + \mu \right) V_I^{o(h)} = -\theta \pi^{lh} \quad (24)$$

Again, dividing (24) by (21) and using condition (5), we have that the labor/leisure choice of old high skilled households is not distorted at the margin:

$$\frac{\partial T(I^h)}{\partial I^h} = 0.$$

In this case, however, also the income/tax point intended for the young agents is not mimicked by anyone else. This implies that also the labor/leisure choice of young households will not be distorted at the margin; dividing (22) by (19) and using (3), we find:

$$\frac{\partial T(I_y)}{\partial I_y} = 0.$$

Finally, dividing (23) by (20) and using (4), we get the result that the old low skilled agents face a non-zero marginal income tax rate:

$$\frac{\partial T(I_o^l)}{\partial I_o^l} = \frac{\mu \widehat{V}_B^h}{\theta \pi^{ll}} \left(\frac{\widehat{V}_I^h}{\widehat{V}_B^h} - \frac{V_I^{o(l)}}{V_B^{o(l)}} \right) = \frac{\mu \widehat{V}_B^h}{\theta \pi^{ll}} \left(MRS_{I,B}^{o(l)} - M\widehat{RS}_{I,B} \right) > 0.$$

3.4 Comments

Using the information on the correlation between skill and age is Pareto improving only if age is directly observable. In the model without savings and with age not directly observable, the information on the joint distribution of skill and age would have been Pareto improving if we had assumed an age-dependent utility function (assuming for instance that old people appreciate leisure relatively more than young people).

4 The Model with Access to the International Capital Market

The productive technology is represented by the same function as before but, since we allow for both borrowing and lending in the international capital market, the resource constraint of the economy takes the form:

$$Q = w_l \left(L_y + \pi^{ll} L_o^l \right) + w_h \pi^{lh} L_o^h + (1 + r) K,$$

where r denotes the marginal productivity of capital K (gross rate of return on savings).^{8,9}

Combining the households' budget constraints

$$\begin{aligned} c_y &= B_y - s \\ c_o^l &= B_o^l + s(1 + r) - trs \\ c_o^h &= B_o^h + s(1 + r) - trs \end{aligned}$$

and the resource constraint

$$I_y + \pi^{ll} I_o^l + \pi^{lh} I_o^h + s(1 + r) = c_y + \pi^{ll} c_o^l + \pi^{lh} c_o^h + s,$$

we get the government's budget constraint:

$$(I_y - B_y) + \pi^{ll} (I_o^l - B_o^l) + \pi^{lh} (I_o^h - B_o^h) + rst = 0.$$

This implies that in the government's problem we need to take into account only one from the resource constraint and government budget constraint.

Before turning to the analysis of Pareto efficient tax policies when the government aims at maximizing actual lifetime utilities, it turns out to be useful to make an intermediate step and deal with the case when the government maximizes the expected utility of agents subject to a self-selection and a budget constraint. On one hand, since all individuals are identical *ex ante*, this might appear as the natural concept of optimality one should employ; on the other hand, as compared to the case when the government looks at *ex post* lifetime utilities and engages in Pareto efficient taxation, we will see that things become simpler and neater results are obtained.

⁸Time indexes are suppressed since we are focusing on steady-state solutions.

⁹The model could also be interpreted as one of a closed economy where, besides labor, it is also used another productive factor, namely capital. However, in this case, given the model we set up, savings couldn't be negative.

4.1 The Expected Utility Case with a “Two Points” System

The government’s problem is the following:

$$\begin{aligned} & \max_{I^l, I_o^h, B^l, B_o^h, q} u \left(B^l - s \left(I^l, I_o^h, B^l, B_o^h, q, \pi^{lh} \right), I^l \right) + \\ & + \frac{\pi^{ll}}{1 + \rho} u \left((1 + q) s(\bullet) + B^l, I^l \right) + \\ & + \frac{\pi^{lh}}{1 + \rho} u \left((1 + q) s(\bullet) + B_o^h, I_o^h \right) \end{aligned}$$

subject to

$$\begin{aligned} & u \left(B^l - s(\bullet), I^l \right) + \frac{\pi^{ll}}{1 + \rho} u \left((1 + q) s(\bullet) + B^l, I^l \right) + \\ & + \frac{\pi^{lh}}{1 + \rho} u \left((1 + q) s(\bullet) + B_o^h, I_o^h \right) \\ \geq & u \left(B^l - s^m \left(I^l, B^l, I^l, B^l, q, \pi^{lh} \right), I^l \right) + \frac{\pi^{ll}}{1 + \rho} u \left((1 + q) s^m(\bullet) + B^l, I^l \right) + \\ & + \frac{\pi^{lh}}{1 + \rho} \widehat{u} \left((1 + q) s^m(\bullet) + B^l, I^l \right), \end{aligned} \tag{\mu}$$

$$u \left((1 + q) s(\bullet) + B_o^h, I_o^h \right) \geq \widehat{u} \left((1 + q) s(\bullet) + B^l, I^l \right), \tag{\xi}$$

$$\left(1 + \pi^{ll} \right) \left(I^l - B^l \right) + \pi^{lh} \left(I_o^h - B_o^h \right) + (r - q) s(\bullet) = 0. \tag{\theta}$$

Notice that in this problem, as compared to what happened in the case without savings, the set of self-selection constraints is larger. In particular, we have an additional self-selection constraint (the μ -constraint) requiring that the lifetime expected utility of a mimicker must be lower or equal to the one of a non-mimicker. Faced with the redistributive policy of the government and anticipating that he/she might be high skilled in the second period, a young agent could in fact be tempted to misrepresent his/her type in the second period and therefore to adjust his/her savings behavior in the first period in order to maximize the gain achievable picking the point intended for the low skilled agents. It is straightforward to show that the μ -constraint will be the only binding self-selection constraint at an optimum and therefore that the ξ -constraint can be neglected. Assume for this purpose that the μ -constraint is satisfied. Then, since the level of savings s^m has been chosen optimally by the potential mimicker, it follows

that no other level of savings (call it s^*) can guarantee him/her a higher expected lifetime utility, which means:

$$\begin{aligned}
& u\left(B^l - s^m(\bullet), I^l\right) + \frac{\pi^{ll}}{1+\rho} u\left((1+q)s^m(\bullet) + B^l, I^l\right) + \\
& + \frac{\pi^{lh}}{1+\rho} \widehat{u}\left((1+q)s^m(\bullet) + B^l, I^l\right) \\
& \geq u\left(B^l - s^*(\bullet), I^l\right) + \frac{\pi^{ll}}{1+\rho} u\left((1+q)s^*(\bullet) + B^l, I^l\right) + \\
& + \frac{\pi^{lh}}{1+\rho} \widehat{u}\left((1+q)s^*(\bullet) + B^l, I^l\right).
\end{aligned}$$

Substituting for s^* in the above inequality the value s chosen by a “fair” young agent and using the μ -constraint, we get

$$\begin{aligned}
& u\left(B^l - s(\bullet), I^l\right) + \frac{\pi^{ll}}{1+\rho} u\left((1+q)s(\bullet) + B^l, I^l\right) + \\
& + \frac{\pi^{lh}}{1+\rho} u\left((1+q)s(\bullet) + B_o^h, I_o^h\right) \\
& \geq u\left(B^l - s(\bullet), I^l\right) + \frac{\pi^{ll}}{1+\rho} u\left((1+q)s(\bullet) + B^l, I^l\right) + \\
& + \frac{\pi^{lh}}{1+\rho} \widehat{u}\left((1+q)s(\bullet) + B^l, I^l\right);
\end{aligned}$$

simplifying terms gives the ξ -constraint.

Denoting by a double “hat” the variables referred to the potential mimicker who turns out to be low skilled in the second period (i.e. the old low skilled agent who saved in the first period the amount s^m), the f.o.c. are the following:¹⁰

$$\begin{aligned}
\frac{\partial u_y}{\partial I^l} + \theta \left[1 + \pi^{ll} - (r - q) \frac{\partial c_y}{\partial I^l} \right] &= -\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} - \mu \left(\frac{\partial u_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} \right) + \\
&+ \mu \frac{\partial \widehat{u}_y}{\partial I^l} + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} \right); \tag{25}
\end{aligned}$$

$$\frac{\partial u_y}{\partial c_y} - \theta \left[1 + \pi^{ll} - (r - q) \left(1 - \frac{\partial c_y}{\partial B^l} \right) \right] = -\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) +$$

¹⁰Notice that in the f.o.c. for the “two points system”, since the government is constrained to choose $B_y = B_o^l = B^l$ as well as $I_y = I_o^l = I^l$, the analytical expressions for $\frac{\partial c_y}{\partial B^l}$ and $\frac{\partial c_y}{\partial I^l}$ would correspond to the sum of respectively (6) and (7), and (9) and (10).

$$-\mu \frac{\partial u_y}{\partial c_y} + \mu \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right); \quad (26)$$

$$\theta \left[\pi^{lh} - (r-q) \frac{\partial c_y}{\partial I_o^h} \right] = -\frac{\pi^{lh}}{1+\rho} (1+\mu) \frac{\partial u_o^h}{\partial I_o^h}; \quad (27)$$

$$\theta \left[\pi^{lh} + (r-q) \frac{\partial c_y}{\partial B_o^h} \right] = \frac{\pi^{lh}}{1+\rho} (1+\mu) \frac{\partial u_o^h}{\partial c_o^h}; \quad (28)$$

$$\begin{aligned} & \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h} \right) s = \theta \left[s + (r-q) \frac{\partial c_y}{\partial q} \right] + \\ & + \mu s^m \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) - \mu s \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h} \right). \end{aligned} \quad (29)$$

Hereafter, we will denote the marginal rates of substitution between labor and consumption for the different agents populating our economy by $MRS_{I,c}^y = -(\frac{\partial u_y}{\partial I_y})/(\frac{\partial u_y}{\partial c_y})$; $MRS_{I,c}^{o(l)} = -(\frac{\partial u_o^l}{\partial I_o^l})/(\frac{\partial u_o^l}{\partial c_o^l})$; $MRS_{I,c}^{o(h)} = -(\frac{\partial u_o^h}{\partial I_o^h})/(\frac{\partial u_o^h}{\partial c_o^h})$; $MRS_{I,c}^m = -(\frac{\partial \widehat{u}_y}{\partial I_y})/(\frac{\partial \widehat{u}_y}{\partial c_y})$; $\widehat{MRS}_{I,c} = -(\frac{\partial \widehat{u}}{\partial I_o^l})/(\frac{\partial \widehat{u}}{\partial c_o^l})$; $\widehat{\widehat{MRS}}_{I,c} = -(\frac{\partial \widehat{u}}{\partial I_o^h})/(\frac{\partial \widehat{u}}{\partial c_o^h})$. Notice that, besides three groups of “fair” agents (young, old low skilled and old high skilled), we have young (low skilled) mimickers choosing in the first period the level of savings s^m that maximizes the expected lifetime gain from mimicking in the second period, old high skilled mimickers and finally old low skilled mimickers which are those who chose strategically s^m in the first period but turned out to be low skilled also in the second period.

Starting the analysis with the characterization of the efficient interest income tax rate and denoting by a “tilde” compensated demands, we get:

Proposition 1 *When the government maximizes expected utility in the two income points system the optimal interest income tax rate is implicitly given by the following condition:*

$$\begin{aligned} (r-q) \frac{\partial \widetilde{c}_y}{\partial q} &= \underbrace{\frac{\mu}{\theta} \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) (s - s^m)}_C + \\ &+ s \left[\underbrace{1 - (r-q) \frac{\partial s}{\partial B_y} \Big|_{B_y=B^l}}_D - \underbrace{\frac{1}{\theta} \frac{\partial u_y}{\partial c_y}}_E - \underbrace{\frac{\mu}{\theta} \left(\frac{\partial u_y}{\partial c_y} - \frac{\partial \widehat{u}_y}{\partial c_y} \right)}_F \right]. \end{aligned} \quad (30)$$

Proof. See Appendix. ■

To interpret (30), notice that the left-hand side of the equation could also have been written (remember that $q = r(1 - t)$) as $t \frac{\partial \tilde{s}}{\partial t}$, a quantity which should look familiar since it closely parallels the *index of discouragement* originally defined by Mirrlees (1976).

As regards the right-hand side of (30), term labelled C is reminiscent of the standard self-selection terms appearing in the formulas for optimal linear commodity taxation when an optimal nonlinear income tax schedule is in place (see e.g. Edwards, Keen and Tuomala, 1994). This kind of rules prescribe that commodity taxation must be handled as an instrument to weaken the binding self-selection constraints; moreover, the bigger the screening power of commodity taxation the larger the scope for its use. The same holds here for term C since it depends on the difference between the level of savings chosen by a “fair” agent and the one chosen by a mimicker. We recover the standard prescription to tax relatively heavier those commodities especially appreciated by a mimicker: if the level of savings of a mimicker exceeds the one of a “fair” agent, then tax (at a positive rate) returns to savings; otherwise, subsidize them. In this case, since in order to maximize his/her expected lifetime utility the mimicker will carry over in the second period a higher amount of savings, the difference $s - s^m$ will be negative and term C calls for a positive tax rate on the returns to savings. According to eq. (30) this is however only part of the story since other factors, which we are about to analyze, must be taken into account.

To get an intuition for terms labelled D , E and F , remember that in the two income points system we are forced to offer the same bundle to young people and old low skilled ones: this means that we cannot move freely B_y . The restriction imposed on B_y (i.e. on the labor income tax schedule) affects how the other tax instruments which can be levied (in our case the interest income tax) are shaped. The term in (30) labelled D captures the total effect on tax revenues that would have followed a marginal increase in B_y (starting from $B_y = B^l$) if we actually could have done it (without being forced to move at the same time B_o^l). This total effect is made up by a direct negative effect on labor income taxes collected (1) and by an indirect effect on interest income taxes collected coming from the adjustment in the level of savings $(- (r - q) \frac{\partial s}{\partial B_y} |_{B_y=B^l})$. Term labelled E captures the private welfare gain (normalized by the marginal cost of public funds) potentially achievable through the increase in B_y , whereas the last term (labelled F) evaluates the same increase in terms of the welfare gain (if positive), or loss (if negative), associated with the effect on the self-selection constraint.

Thus, we have that the second line of (30) will be positive if it is negative the net social welfare effect descending from the *hypothetical* marginal increase of B_y starting from $B_y = B^l$.

Neglecting for a moment term C , we would therefore have that the dis-

tortion imposed on the demand for savings should be greater the greater the net social welfare effect potentially achievable from a *hypothetical* marginal increase of B_y starting from $B_y = B^l$; moreover, we would like to discourage (encourage) savings if this net social welfare effect is positive (negative). Notice that since savings are a “commodity” that in principle can both be demanded and supplied by young agents, discouraging savings would require $t > 0$ if savings were positive but $t < 0$ if savings were negative. The reverse holds for the case when we would like to encourage savings.

The appearance of terms labelled D , E and F in (30) can be viewed as another instance of the general principle that, whenever there are restrictions on the set of feasible taxes, those taxes which can be levied are adjusted to serve as partial substitutes for the taxes which cannot be levied.

Turning to the marginal income tax rate faced by the old high skilled agents is given, we get the following result:

Proposition 2 *When the government maximizes expected utility in the two income points system the marginal (labor) income tax rate $T'_{o(h)}$ faced by old high skilled agents is given by:*

$$T'_{o(h)} = \frac{r - q}{\pi^{lh}} \left(\frac{dc_y}{dI_o^h} + \frac{dc_y}{dB_o^h} MRS_{I,c}^{o(h)} \right). \quad (31)$$

Proof. Dividing (27) by (28) and multiplying by $\pi^{lh} + (r - q) \frac{\partial c_y}{\partial B_o^h}$ gives:

$$\pi^{lh} - (r - q) \frac{\partial c_y}{\partial I_o^h} = - \frac{\frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} \left[\pi^{lh} + (r - q) \frac{\partial c_y}{\partial B_o^h} \right]. \quad (32)$$

The result provided by (31) is obtained using the definition of marginal income tax rate given by (5) to collect terms in (32). ■

Looking at (27) and (28), we can observe that the result that $T'_{o(h)}$ is in general different from zero is a consequence of the fact that a marginal change in I_o^h and a marginal change in B_o^h induce adjustment effects in the level of savings by young agents which are of different scale and which in turn imply for the government budget effects of different magnitude. If $\frac{\partial c_y}{\partial I_o^h} = -\frac{\partial c_y}{\partial B_o^h}$, then dividing (27) by (28) we would have got $T'_{o(h)} = 0$.

Optimal tax policy imposes a distortion in the labor-leisure choice of old high skilled agents. Whether they undersupply ($T'_{o(h)} > 0$) or oversupply ($T'_{o(h)} < 0$) labor depends on the sign of the budget effect on interest income tax receipts coming from the adjustment in the level of savings which follows when old high skilled agents are induced to marginally increase their labor supply. The total effect on savings is provided by the quantity inside brackets (multiplied by -1) on the right-hand side of (31). It is given by the sum of the direct effect coming from a marginal increase in labor supply ($-\frac{dc_y}{dI_o^h}$) and

the indirect effect coming from the increase in disposable income which is required to make old high skilled agents willing to marginally increase their labor supply $(-\frac{dc_y}{dB_o^h} MRS_{I,c}^{o(h)})$.

According to condition (31), if the total effect on savings is positive ($dc_y < 0$) and interest incomes are taxed at a positive rate ($t = \frac{r-q}{r} > 0 \implies r - q > 0$), then revenue collected by taxing the returns to savings are increased and we should marginally subsidize old high skilled agents to make them overprovide labor.

Denoting by $ds|_{du_o^h=0}$ the total effect on savings $(-\frac{dc_y}{dB_o^h} MRS_{I,c}^{o(h)} - \frac{dc_y}{dI_o^h})$ we would therefore have:

$$\begin{aligned} T'_{o(h)} &< 0 \quad \text{if} \quad ds|_{du_o^h=0} > (<) 0 \quad \text{and} \quad t > (<) 0; \\ T'_{o(h)} &> 0 \quad \text{if} \quad ds|_{du_o^h=0} > (<) 0 \quad \text{and} \quad t < (>) 0. \end{aligned}$$

Let's look now at the total amount of taxes paid by an old high skilled agent and introduce the concept of marginal effective tax rate (hereafter METR) as the change in his/her total tax payment that would occur if he/she were to earn a little more. Denoting the total amount of taxes paid by a high skilled agent in the second period by $\tau(I_o^h) = T(I_o^h) + (r - q)s(I^l, I_o^h, B^l, I_o^h - T(I_o^h), q, \pi^{lh})$ and differentiating w.r.t. I_o^h to get the METR which we will denote by $\tau'_{o(h)}$, it is:

$$\tau'_{o(h)} = T'_{o(h)} + (r - q) \left[\frac{\partial s}{\partial I_o^h} + (1 - T'_{o(h)}) \frac{\partial s}{\partial B_o^h} \right] = \quad (33)$$

$$= 1 + (r - q) \frac{\partial s}{\partial I_o^h} - \frac{\frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} \left[(r - q) \frac{\partial s}{\partial B_o^h} - 1 \right]. \quad (34)$$

Proposition 3 states the main result.

Proposition 3 *When the government maximizes expected utility in the two income points system the METR faced by old high skilled agents is*

$$\tau_o^{lh} = \pi^{ll} T'_{o(h)}. \quad (35)$$

Proof. The METR can also be written as

$$\tau_o^{lh} = 1 - (r - q) \frac{\partial c_y}{\partial I_o^h} + \frac{\frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} \left[1 + (r - q) \frac{\partial c_y}{\partial B_o^h} \right]. \quad (36)$$

The result is obtained substituting for $1 - (r - q) \frac{\partial c_y}{\partial I_o^h}$ in (36) the corresponding expression derived adding $1 - \pi^{lh}$ on both sides of (32). ■

The marginal distortions imposed by labor income taxation and interest income taxation push in opposite directions as it happened in the atemporal model for the effects of income and commodity taxation.¹¹ Here, however, they don't "average out" to zero. Instead the METR has the same sign as the marginal labor income tax rate. In some sense, therefore, the distortion imposed at the margin by the labor income tax is "too high", or at least too high to get the usual "no distortion at the top result". Going back to eq. (31) allows us to trace the source of such a discrepancy. The basic reason is that changing the bundle offered to old high skilled agents does not only affect the total amount of taxes paid by this sub-set of the population but, through the savings function, it also affects the amount of interest income taxes collected from the old low skilled individuals.¹² This is a direct consequence of the fact that in our model savings decisions of young agents take place under conditions of uncertainty about their future skill level and that all young agents share the same uncertainty, so that savings will be homogeneous and everybody pays the same amount of interest income taxes in the second period. Observing that the smaller the relative size of the high skilled group among old agents the greater the extent of this "external" effect, we can then understand why in (35) the value of the distortion provided by the METR is an increasing function of the proportion π^l .

Finally, notice that there is actually a quantity that at the optimum would be unaffected if the old high skilled agents were to earn a little more: the total amount of taxes collected by the government. To show this, define the quantity $\tau_{tot} = (1 + \pi^l) T(I^l) + \pi^{lh} T(I_o^h) + (r - q) s(\bullet)$ and observe that differentiation w.r.t. I_o^h gives

$$\begin{aligned} \frac{d\tau_{tot}}{dI_o^h} &= \pi^{lh} T'_{o(h)} + (r - q) \left[\frac{\partial s}{\partial I_o^h} + \left(1 - T'_{o(h)}\right) \frac{\partial s}{\partial B_o^h} \right] = \\ &= \pi^{lh} - (r - q) \frac{\partial c_y}{\partial I_o^h} + \frac{\frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} \left[\pi^{lh} + (r - q) \frac{\partial c_y}{\partial B_o^h} \right]; \end{aligned} \quad (37)$$

the result is straightforward substituting for $\pi^{lh} - (r - q) \frac{\partial c_y}{\partial I_o^h}$ in (37) the right-hand side of (32).

Let's look now at the low skilled agents. In the two points system young agents and old low skilled ones are bunched together at the same income point. At this common point the labor income tax schedule is kinked, but it is possible to show that there always exists an implementing tax structure whose left (right)-hand derivative at I^l is equal to $1 - MRS_{I,B}$ for those with the steepest (flattest) indifference curves among those who are bunched

¹¹ See again Edwards, Keen and Tuomala (1994).

¹² If this didn't happen, then in (31) π^{lh} at the denominator would disappear and we would recover the standard "end point" result.

together. In the following Proposition we characterize the implementing tax structure assuming that savings of young agents are negative. This is in our model a quite reasonable assumption provided that the value of the gross rate of return r does not exceed by a too large amount the value of ρ . Afterwards, we will comment on how results change if the reverse assumption holds.

Proposition 4 *When the government maximizes expected utility and savings of young agents are negative, the optimal allocation in the two income points system can be implemented through a labor income tax schedule whose left-hand- and right-hand derivatives at the common point in the (I, B) -space for young workers and old low skilled workers are respectively given by:*

$$\begin{aligned}
T'_{(left)}(I^l) = & \frac{\mu\pi^{lh}\frac{\partial\widehat{u}}{\partial c_o^l}}{\theta(1+\pi^{ll})(1+\rho)} \left(MRS_{I,c}^y - \widehat{MRS}_{I,c} \right) + \\
& + \frac{\mu\pi^{ll}\frac{\partial\widehat{u}}{\partial c_o^l}}{\theta(1+\pi^{ll})(1+\rho)} \left(MRS_{I,c}^y - \widehat{\widehat{MRS}}_{I,c} \right) + \\
& + \frac{\mu\frac{\partial\widehat{u}_y}{\partial c_y}}{\theta(1+\pi^{ll})} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) + \\
& + \frac{(1+\mu)\pi^{ll}}{\theta(1+\pi^{ll})} \frac{\frac{\partial u_o^l}{\partial c_o^l}}{1+\rho} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^y \right) + \\
& + \frac{r-q}{1+\pi^{ll}} \left[\frac{dc_y}{dI^l} - \left(1 - \frac{dc_y}{dB^l} \right) MRS_{I,c}^y \right]; \quad (38)
\end{aligned}$$

$$\begin{aligned}
T'_{(right)}(I^l) = & \frac{\mu\pi^{lh}\frac{\partial\widehat{u}}{\partial c_o^l}}{\theta(1+\pi^{ll})(1+\rho)} \left(MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c} \right) + \\
& + \frac{\mu\pi^{ll}\frac{\partial\widehat{u}}{\partial c_o^l}}{\theta(1+\pi^{ll})(1+\rho)} \left(MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} \right) + \\
& + \frac{\mu\frac{\partial\widehat{u}_y}{\partial c_y}}{\theta(1+\pi^{ll})} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^m \right) + \\
& + \frac{(1+\mu)\frac{\partial u_y}{\partial c_y}}{\theta(1+\pi^{ll})} \left(MRS_{I,c}^y - MRS_{I,c}^{o(l)} \right) + \\
& + \frac{r-q}{1+\pi^{ll}} \left[\frac{dc_y}{dI^l} - \left(1 - \frac{dc_y}{dB^l} \right) MRS_{I,c}^{o(l)} \right]. \quad (39)
\end{aligned}$$

Proof. See Appendix. ■

We will come back later, in the analysis of the METRs, to the terms appearing in (38) and (39). For the moment, just notice that if savings of young agents had been positive, then we would have had that the left-hand

derivative of the implementing tax structure at the bunching point was given by (39) and the right-hand derivative by (38).

Now consider the METRs. Whereas adapting expression (33) provides a natural way to define the METR faced by old low skilled agents, it is not obvious which definition to use when looking at young agents. Since interest income taxes are paid when old, the change in the total tax payment of young agents that would occur if they were to earn a little more depends on the temporal horizon that we choose. If we focus on the first period, then the change in the total tax payment is simply given by the marginal labor income tax rate and we already got an expression for it. If instead we take a lifetime perspective and include also the change in interest income taxes paid in the second period (which is certain because it does not depend on the individual's skill level when old), then something similar to (33) should be considered. In this case, however, changes in future tax payments should be discounted by r . Using the implicit definition for the marginal labor income tax rate provided by (3), the METR faced by young agents would then be:

$$\begin{aligned}\tau'_y &= 1 + \frac{\frac{\partial u_y}{\partial I^l}}{\frac{\partial u_y}{\partial c_y}} + \frac{r-q}{1+r} \left[\frac{\partial s}{\partial I^l} - \frac{\frac{\partial u_y}{\partial I^l}}{\frac{\partial u_y}{\partial c_y}} \frac{\partial s}{\partial B^l} \right] = \\ &= 1 - \frac{r-q}{1+r} \frac{\partial c_y}{\partial I^l} + \frac{\frac{\partial u_y}{\partial I^l}}{\frac{\partial u_y}{\partial c_y}} \left[1 - \frac{r-q}{1+r} \left(1 - \frac{\partial c_y}{\partial B^l} \right) \right].\end{aligned}$$

Proposition 5 *When the government maximizes expected utility in the two income points system the METRs faced by young agents and old low skilled agents are respectively given by*

$$\begin{aligned}\tau'_y &= \frac{\mu}{\theta} \frac{\pi^{lh}}{1+\pi^{ll}} \frac{1}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^y - \widehat{MRS}_{I,c} \right) + \\ &+ \frac{\mu}{\theta} \frac{\pi^{ll}}{1+\pi^{ll}} \frac{1}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^y - \widehat{\widehat{MRS}}_{I,c} \right) + \\ &+ \frac{\mu}{\theta} \frac{1}{1+\pi^{ll}} \frac{\partial \widehat{u}_y}{\partial c_y} \underbrace{\left(MRS_{I,c}^y - MRS_{I,c}^m \right)}_{+} + \\ &- \frac{1+\mu}{\theta} \frac{1}{1+\rho} \frac{1}{1+\pi^{ll}} \frac{\partial u_o^l}{\partial c_o^l} \underbrace{\left(MRS_{I,c}^y - MRS_{I,c}^{o(l)} \right)}_{+} + \\ &+ \frac{r-q}{1+\pi^{ll}} \frac{r-\pi^{ll}}{1+r} \left[\frac{dc_y}{dI^l} - \underbrace{\left(1 - \frac{dc_y}{dB^l} \right)}_{+} MRS_{I,c}^y \right] \quad (40)\end{aligned}$$

and

$$\begin{aligned}
\tau_o^l = & \frac{\mu}{\theta} \frac{\pi^{lh}}{1 + \pi^{ll}} \frac{1}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c} \right) + \\
& + \frac{\mu}{\theta} \frac{\pi^{ll}}{1 + \pi^{ll}} \frac{1}{1 + \rho} \frac{\partial \widehat{\widehat{u}}}{\partial c_o^l} \underbrace{\left(MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} \right)}_{-} + \\
& + \frac{\mu}{\theta} \frac{1}{1 + \pi^{ll}} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^m \right) + \\
& - \frac{1 + \mu}{\theta (1 + \pi^{ll})} \frac{\partial u_y}{\partial c_y} \underbrace{\left(MRS_{I,c}^{o(l)} - MRS_{I,c}^y \right)}_{-} + \\
& - \frac{\pi^{ll}}{1 + \pi^{ll}} (r - q) \left[\frac{dc_y}{dI^l} - \underbrace{\left(1 - \frac{dc_y}{dB^l} \right)}_{+} MRS_{I,c}^{o(l)} \right]. \quad (41)
\end{aligned}$$

Proof. See Appendix. ■

Eq. (40) and (41) provide analytically complex expressions for the global distortion faced by the two groups of low skilled agents. Moreover, their signs remain *a priori* ambiguous as it appears looking at the sign of each term for which it was possible to decide on it. Nevertheless, the structure of the formulas can be easily interpreted. Remember that in the standard (timeless) two types model of optimal taxation the METR faced by the low skilled agents depends on the difference between the marginal rate of substitution for a mimicker and for a mimicked.¹³ The same happens here with the first three terms on the r.h.s. of (40) and (41). The difference is that, whilst in the timeless version of the model the potential mimicker is unambiguously identified, here there are more than one potential mimickers: besides the old high skilled agents, one has also to consider those young choosing strategically the level of savings (in order to fully exploit the possibility to mimic in the second period) and those who are low skilled in the second stage of life but chose strategically the level of savings in the first one. This difference accounts for the fact that a mimicker and a mimicked are not necessarily people sharing the same disposable income but enjoying different amounts of leisure. Depending on which case we consider, we can alternatively have people with different disposable income and different leisure (when the mimicker is a high skilled agent), or people with same leisure but different amounts of disposable income (when the mimicker is a low skilled agent, either young or old).

In eq. (40), the first term involves the difference between $MRS_{I,c}^y$ and $\widehat{MRS}_{I,c}$. The high skilled mimicker certainly has a lower labor supply but,

¹³ Once again, see Edwards, Keen and Tuomala (1994).

what about consumption? We know that a young agent will consume $B^l - s$ whereas the high skilled mimicker will consume $B^l + s^m(1 + q)$. We also know that $s^m > s$ and that, under reasonable assumption, $s < 0$. This means that, as long as $s^m < -\frac{s}{1+q}$, the consumption of a high skilled mimicker will be lower than the one of a young agent. If this is the case, then $MRS_{I,c}^y - \widehat{MRS}_{I,c} > 0$. However, if it is $s^m > -\frac{s}{1+q}$, then the sign of $MRS_{I,c}^y - \widehat{MRS}_{I,c}$ becomes ambiguous.

The second term in (40) involves the difference between $MRS_{I,c}^y$ and $\widehat{\widehat{MRS}}_{I,c}$. Now the mimicker is an old low skilled and therefore mimicker and mimicked will have the same labor supply. Once again, the consumption of the mimicker is given by $B^l + s^m(1 + q)$ and we can conclude that $MRS_{I,c}^y - \widehat{\widehat{MRS}}_{I,c} > (<) 0$ when $s^m < (>) -\frac{s}{1+q}$.

The third term in (40) is related to the difference between $MRS_{I,c}^y$ and $MRS_{I,c}^m$. The sign of this term is unambiguous since the agents considered have the same labor supply and two different levels of consumption which can be ranked unambiguously. The consumption of a young mimicker is in fact $B^l - s^m$ which is always smaller than $B^l - s$ since $s^m > s$. For this reason, it is $MRS_{I,c}^y - MRS_{I,c}^m > 0$.

Turning to eq. (41), the sign of the first term depends on that of $MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c}$. The agents that are compared are an old low skilled and a high skilled mimicker. The labor supply of the high skilled mimicker is lower than the one of the old low skilled. As regards consumption, the high skilled mimicker will consume $B^l + s^m(1 + q)$ while the old low skilled will consume $B^l + s(1 + q)$. Therefore, being $s^m > s$, the consumption of the high skilled mimicker will exceed the one of the old low skilled. A mimicker will in this case enjoy both a higher level of consumption and a higher level of leisure and the sign of $MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c}$ remains ambiguous.

The second term in (41) takes into account the difference between $MRS_{I,c}^{o(l)}$ and $\widehat{\widehat{MRS}}_{I,c}$, i.e. between an old low skilled who chose s in the first period and an old low skilled who chose s^m . They have the same labor supply but the old low skilled who chose s consumes less, and therefore $MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} < 0$.

The third term in (41) looks at $MRS_{I,c}^{o(l)}$ as compared with $MRS_{I,c}^m$. Both are low skilled and have the same labor supply but the old low skilled consumes $B^l + s(1 + q)$ while the young who chooses to save s^m consumes $B^l - s^m$. Since the consumption of the latter will be higher as long as $s^m < -s(1 + q)$, we can conclude that $MRS_{I,c}^{o(l)} - MRS_{I,c}^m < (>) 0$ when $s^m < (>) -s(1 + q)$.

As usual, in shaping the tax instruments the policy maker tries to weaken the self-selection constraint by making the mimicking option less appealing

and in order to attain incentive-compatibility a distortion is imposed on those towards whom redistribution is directed. However, since the decision whether to become a mimicker or not is taken in the first period of life, the expected lifetime pay off for a mimicker can alternatively be affected by acting on his/her utility when young, on his/her utility as an old low skilled, or finally on his/her utility as an old high skilled. Thus, the first three terms on the r.h.s. of (40) and (41) provide a value for the distortion which optimally balances these different mimicking-discouraging effects.

The fourth terms on the r.h.s. of (40) and (41) are something peculiar to the “two points” system. They are a direct consequence of the fact that under this regime people with different slopes of the indifference curves in the (I, B) -space are offered the same bundle (I^l, B^l) . If at a given allocation the slopes are different, then it will also be different the minimal compensation required to marginally increase labor supply. To give an insight on how this property can be usefully exploited by the policy maker, let's take for instance the expression for the METR faced by young agents. If, as we previously assumed, savings of young agents are negative, then normality of both leisure and consumption ensures that $MRS_{I,c}^y - MRS_{I,c}^{o(l)}$ is positive.

$MRS_{I,c}^y > MRS_{I,c}^{o(l)}$ means that the compensation required by a young agent to marginally increase labor supply exceeds the compensation required for the same purpose by an old low skilled. If this is the case, when young agents are induced to increase labor supply by moving them along their indifference curve (i.e. increasing B^l by $MRS_{I,c}^y$), the increase in disposable income makes old low skilled agents better off (because they would have required an increase by just $MRS_{I,c}^{o(l)}$). This increase in utility for old low skilled agents is both valuable because it increases the value of the maximand of the government's problem and because it relaxes the binding self-selection constraint. The value of the difference $MRS_{I,c}^y - MRS_{I,c}^{o(l)}$ reflects the magnitude of this social gain related to the overprovision of labor by young agents. Other things being equal, the greater this value and the lower the METR faced by young agents. A symmetric argument holds for the METR faced by old low skilled agents. In this case, if $MRS_{I,c}^y > MRS_{I,c}^{o(l)}$, then, other things being equal, it would pay to increase the METR faced by this group of people in order to make them underprovide labor. The reason is that, if we make them marginally increase labor supply holding their utility constant, the increase in disposable income is not sufficient to prevent young agents to be worse off. This has a double negative effect, since it both decreases the value of the maximand of the government's problem and it tightens the binding self-selection constraint.

The last terms appearing in (40) and (41) are instead related to (budget) effects on interest tax receipts of the same kind of those, previously described, determining the METR faced by old high skilled agents.

4.2 The Expected Utility Case with a “Three Points” System

When age is observable and the policy maker can use the information on the correlation between skill and age in order to optimally shape the income tax schedule, the problem looks as follows:

$$\begin{aligned} & \max_{I_y, I_o^l, I_o^h, B_y, B_o^l, B_o^h, q} u \left(B_y - s \left(I_y, I_o^l, I_o^h, B_y, B_o^l, B_o^h, q, \pi^{lh} \right), I_y \right) + \\ & + \frac{\pi^{ll}}{1 + \rho} u \left((1 + q) s(\bullet) + B_o^l, I_o^l \right) + \\ & + \frac{\pi^{lh}}{1 + \rho} u \left((1 + q) s(\bullet) + B_o^h, I_o^h \right) \end{aligned}$$

subject to

$$\begin{aligned} & u \left(B_y - s(\bullet), I_y \right) + \frac{\pi^{ll}}{1 + \rho} u \left((1 + q) s(\bullet) + B_o^l, I_o^l \right) + \\ & + \frac{\pi^{lh}}{1 + \rho} u \left((1 + q) s(\bullet) + B_o^h, I_o^h \right) \\ \geq & u \left(B_y - s^m \left(I_y, I_o^l, B_y, B_o^l, q, \pi^{lh} \right), I_y \right) + \frac{\pi^{ll}}{1 + \rho} u \left((1 + q) s^m(\bullet) + B_o^l, I_o^l \right) + \\ & + \frac{\pi^{lh}}{1 + \rho} \widehat{u} \left((1 + q) s^m(\bullet) + B_o^h, I_o^h \right), \end{aligned} \quad (\mu)$$

$$u \left((1 + q) s(\bullet) + B_o^h, I_o^h \right) \geq \widehat{u} \left((1 + q) s(\bullet) + B_o^l, I_o^l \right), \quad (\xi)$$

$$(I_y - B_y) + \pi^{ll} \left(I_o^l - B_o^l \right) + \pi^{lh} \left(I_o^h - B_o^h \right) + (r - q) s(\bullet) = 0. \quad (\theta)$$

By the same argument put forward in the previous case, we can avoid taking into account the ξ -constraint since we know that at an optimum the only binding self-selection constraint is the one associated with the Lagrange multiplier μ .

In the three points system the first order conditions referred to the pre-tax labor income and after-tax labor income for old high skilled and to the net rate of return on savings formally do not change when compared to those derived for the two points system, whereas the first order conditions of the government's problem with respect to the pre-tax labor income and after-tax labor income for young agents are respectively given by:

$$\theta \left[1 - (r - q) \frac{\partial c_y}{\partial I_y} \right] = -\frac{\partial u_y}{\partial I_y} + \mu \left(\frac{\partial \widehat{u}_y}{\partial I_y} - \frac{\partial u_y}{\partial I_y} \right); \quad (42)$$

$$\theta \left[-1 + (r - q) \left(1 - \frac{\partial c_y}{\partial B_y} \right) \right] = -\frac{\partial u_y}{\partial c_y} + \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} - \frac{\partial u_y}{\partial c_y} \right); \quad (43)$$

and those for old low skilled agents are:

$$\begin{aligned} \theta \left[\pi^{ll} - (r - q) \frac{\partial c_y}{\partial I_o^l} \right] &= -\frac{\pi^{ll}}{1 + \rho} \frac{\partial u_o^l}{\partial I_o^l} - \mu \frac{\pi^{ll}}{1 + \rho} \frac{\partial u_o^l}{\partial I_o^l} + \\ &+ \mu \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial \widehat{u}}{\partial I_o^l} + \frac{\pi^{lh}}{1 + \rho} \frac{\partial \widehat{u}}{\partial I_o^l} \right); \end{aligned} \quad (44)$$

$$\begin{aligned} \theta \left[\pi^{ll} + (r - q) \frac{\partial c_y}{\partial B_o^l} \right] &= \frac{\pi^{ll}}{1 + \rho} \frac{\partial u_o^l}{\partial c_o^l} + \mu \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial u_o^l}{\partial c_o^l} \right) + \\ &- \mu \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right). \end{aligned} \quad (45)$$

Proposition 6 *When the government maximizes expected utility in the three income points system interest income taxation must satisfy the following condition:*

$$(r - q) \frac{\partial \widetilde{c}_y}{\partial q} = \frac{\mu}{\theta} \left(\frac{\pi^{lh}}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{ll}}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) (s - s^m). \quad (46)$$

Proof. See Appendix. ■

Comparing Proposition 6 with Proposition 1, we see that, once we remove the constraint requiring young and old low skilled agents to be stuck at the same income point along the labor income tax schedule, the optimal interest income tax rule exactly mirrors the commodity tax rule of the standard atemporal version of the two-types mixed taxation model. As long as the savings behavior of a mimicker differs from the one of a “fair” agent and starting from a situation where interest incomes are tax-free, a Pareto-improving reform can be implemented taxing (heavier) the good relatively more demanded by a mimicker, which in our case happens to be consumption in the second period, i.e. savings. This implies $t = \frac{r-q}{r} > 0$.

The difference between rules (46) and (30) highlights the fact that, even if the analytical expressions for the marginal labor income tax rate and METR faced by old high skilled agents formally do not change when we switch to a three income points system (it is easy to show that (31) and (35) are still valid), the actual values of those marginal tax rates are in fact quite different. The magnitude of the global distortion imposed at the

margin on old high skilled agents is bigger the greater the wedge between r and q . Thus, other things being equal, the fact whether old high skilled are more distorted in the two- or in the three income points system depends on whether the effects of terms C and $D - (E + F)$ reinforce each other or push in opposite directions.

Let's look now at the young agents and the old low skilled ones. Propositions 7 and 8 give the main results.

Proposition 7 *When the government maximizes expected utility in the three income points system the marginal (labor) income tax rate T'_y and the marginal effective tax rate τ'_y faced by young agents are respectively given by:*

$$T'_y = \frac{\mu}{\theta} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) + (r - q) \left[\frac{\partial c_y}{\partial I_y} - \left(1 - \frac{\partial c_y}{\partial B_y} \right) MRS_{I,c}^y \right]; \quad (47)$$

$$\begin{aligned} \tau'_y &= \frac{r}{1+r} T'_y + \frac{\mu}{\theta} \frac{1}{1+r} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) = \\ &= \frac{\mu}{\theta} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) + \\ &\quad + \frac{r}{1+r} (r - q) \left[\frac{\partial c_y}{\partial I_y} - \left(1 - \frac{\partial c_y}{\partial B_y} \right) MRS_{I,c}^y \right]. \end{aligned} \quad (48)$$

Proof. See Appendix. ■

In the three income points system with observability of age, young agents cannot be mimicked by old agents. However, since the decision to become eventually a mimicker in the second period of life is taken in the first period and affects the savings behavior, a mimicker will be characterized by a level of savings which deviates from the level intended by the government for the young agents. In a certain sense we could also say that there are actually mimickers also among young agents: those who choose s^m instead of s . Not surprisingly, to weaken the binding self-selection constraint the government should therefore try to make costlier to deviate from s . Since $MRS_{I,c}^y > MRS_{I,c}^m$, this can be accomplished by imposing a positive distortion on young agents. This accounts for the presence of the term $\frac{\mu}{\theta} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right)$ both in (47) and in (48).

The term $(r - q) \left[\frac{\partial c_y}{\partial I_y} - \left(1 - \frac{\partial c_y}{\partial B_y} \right) MRS_{I,c}^y \right]$ in (47) can be explained referring to the same argument provided in subsection 4.1 to interpret the marginal income tax rate faced by high skilled agents. It is a budget term, reflecting the net effect on receipts from interest income taxation of a marginal increase in labor supply by young agents.

The term $\frac{r}{1+r} (r - q) \left[\frac{\partial c_y}{\partial I_y} - \left(1 - \frac{\partial c_y}{\partial B_y} \right) MRS_{I,c}^y \right]$ in (48) is instead a consequence of the fact that in the definition of τ'_y we discounted by r the change

in the amount of interest income taxes the young agents are going to pay when old. It is easy to check that, if we didn't discount it, we would have got $\tau'_y = \frac{\mu}{\theta} \frac{\partial \widehat{u}_y}{\partial c_y} (MRS_{I,c}^y - MRS_{I,c}^m) > 0$.

As regards the old low skilled, we have:

Proposition 8 *When the government maximizes expected utility in the three income points system the marginal (labor) income tax rate T_o^l and the marginal effective tax rate τ_o^l faced by old low skilled agents are respectively given by:*

$$\begin{aligned} T_o^l &= \frac{\mu}{\theta} \frac{1}{1+\rho} \left[\frac{\pi^{lh}}{\pi^{ll}} \frac{\partial \widehat{u}}{\partial c_o^l} (MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c}) + \frac{\partial \widehat{u}}{\partial c_o^l} (MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c}) \right] + \\ &\quad + \frac{r-q}{\pi^{ll}} \left[\frac{\partial c_y}{\partial I_o^l} + MRS_{I,c}^{o(l)} \frac{\partial c_y}{\partial B_o^l} \right]; \end{aligned} \quad (49)$$

$$\begin{aligned} \tau_o^l &= \pi^{lh} T_o^l + \\ &\quad + \frac{\mu}{\theta} \frac{1}{1+\rho} \left[\pi^{lh} \frac{\partial \widehat{u}}{\partial c_o^l} (MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c}) + \pi^{ll} \frac{\partial \widehat{u}}{\partial c_o^l} (MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c}) \right] \\ &= \frac{\mu}{\theta} \frac{1}{1+\rho} \left[\frac{\pi^{lh}}{\pi^{ll}} \frac{\partial \widehat{u}}{\partial c_o^l} (MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c}) + \frac{\partial \widehat{u}}{\partial c_o^l} (MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c}) \right] + \\ &\quad + \frac{\pi^{lh}}{\pi^{ll}} (r-q) \left[\frac{\partial c_y}{\partial I_o^l} + MRS_{I,c}^{o(l)} \frac{\partial c_y}{\partial B_o^l} \right]. \end{aligned} \quad (50)$$

Proof. See Appendix. ■

To interpret the first term appearing both in (49) and (50), remember that among the old agents there are two types of mimickers: those who are high skilled but pick the point intended for the old low skilled and those who are low skilled but chose in the first period s^m instead of s . The government imposes a distortion on the old low skilled agents trying to lower the expected pay-off in the second period for a mimicker and in this way to relax the binding self-selection constraint. Since $MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} < 0$, the condition of old low skilled mimicker is worsened by imposing a subsidy at the margin. On the other hand, the kind of distortion required to hurt a high skilled mimicker depends on the sign of the difference $MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c}$. As we previously noticed, this sign is ambiguous since in this case a mimicker enjoys both a higher level of consumption and a higher level of leisure. Eventually the required distortions could push in opposite directions and then, to determine which one prevails, it becomes important the relative magnitude of the weights π^{ll} and π^{lh} .

The term $\frac{r-q}{\pi^{ll}} \left[\frac{\partial c_y}{\partial I_o^l} + MRS_{I,c}^{o(l)} \frac{\partial c_y}{\partial B_o^l} \right]$ in (49) is a budget term, reflecting the net effect on receipts from interest income taxation of a marginal increase

in labor supply by old low skilled agents. It can be interpreted referring to the same argument put forward in subsection 4.1 when we analyzed the marginal income tax rate faced by high skilled agents. Other things being equal, the value of T_o^l will be higher (lower) if an increase in I_o^l , accomplished with a compensating increase in B_o^l by $MRS_{I,c}^{o(l)} \frac{\partial c_y}{\partial B_o^l}$, induces adjustments in the level of savings which decrease (increase) the amount of interest income taxes collected.

A budget term is also the last in (50). A rationale for this term can be provided observing that a change in the bundle offered to old low skilled agents affects also, through the savings function, the amount of interest income taxes paid by the old high skilled agents. It is only the extent of this “external” effect that matters for the METR faced by old low skilled agents. This “external” effect will be stronger the higher the proportion of high skilled in the old population. That’s why, as π^{lh} approaches zero, the value of the last term in (50) will also approach zero. The argument is the exact counterpart of the one exposed in the analysis of the METR faced by old high skilled agents.

Let’s look now at how the problem changes and the results are affected if the government aims at maximizing the actual lifetime utility of those who remain low skilled during the entire life subject to a given value for the lifetime utility of those who become high skilled, a self-selection constraint and a budget constraint.

4.3 Case 1: The Actual Lifetime Utility Case with a “Two Points” System

When in the design of the fiscal policy the government is constrained to offer the same bundle in the (I, B) -space for both the young and the old low skilled agents, its problem is the following:¹⁴

$$\begin{aligned}
& \max_{I^l, I_o^h, B^l, B_o^h, q} u \left(c_y \left(I^l, I_o^h, B^l, B_o^h, q, \pi^{lh} \right), I^l \right) + \\
& + \frac{1}{1+\rho} u \left(\left(B^l - c_y(\bullet) \right) (1+q) + B^l, I^l \right) \\
& \text{subject to} \\
& u \left(c_y \left(I^l, I_o^h, B^l, B_o^h, q, \pi^{lh} \right), I^l \right) + \\
& + \frac{1}{1+\rho} u \left(\left(B^l - c_y(\bullet) \right) (1+q) + B_o^h, I_o^h \right) \geq \bar{V}, \tag{\lambda}
\end{aligned}$$

¹⁴By the same argument used in the sections dealing with the expected utility case we set up the problem considering the only self-selection constraint which is binding at an optimum.

$$\begin{aligned}
& u \left(B^l - s(\bullet), I^l \right) + \frac{\pi^{ll}}{1+\rho} u \left((1+q) s(\bullet) + B^l, I^l \right) + \\
& + \frac{\pi^{lh}}{1+\rho} u \left((1+q) s(\bullet) + B_o^h, I_o^h \right) \\
\geq & u \left(B^l - s^m \left(I^l, B^l, I^l, B^l, q, \pi^{lh} \right), I^l \right) + \frac{\pi^{ll}}{1+\rho} u \left((1+q) s^m(\bullet) + B^l, I^l \right) + \\
& + \frac{\pi^{lh}}{1+\rho} \widehat{u} \left((1+q) s^m(\bullet) + B^l, I^l \right), \tag{\mu}
\end{aligned}$$

$$(1 + \pi^{ll}) \left(I^l - B^l \right) + \pi^{lh} \left(I_o^h - B_o^h \right) + (r - q) \left(B^l - c_y(\bullet) \right) = 0. \tag{\theta}$$

The f.o.c. are the following:¹⁵

$$\begin{aligned}
& (1 + \lambda) \left[\frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial I^l} + \frac{\partial u_y}{\partial I^l} \right] - \frac{1+q}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \frac{\partial c_y}{\partial I^l} + \frac{1}{1+\rho} \frac{\partial u_o^l}{\partial I^l} \\
= & -\theta \left[1 + \pi^{ll} - (r - q) \frac{\partial c_y}{\partial I^l} \right] + \mu \left(\frac{\partial \widehat{u}_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} \right) + \\
& -\mu \left(\frac{\partial u_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} \right); \tag{51}
\end{aligned}$$

$$\begin{aligned}
& (1 + \lambda) \frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial B^l} + \frac{1}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \left(1 - \frac{\partial c_y}{\partial B^l} \right) \frac{1+q}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \\
= & \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) + \theta \left[1 + \pi^{ll} - (r - q) \left(1 - \frac{\partial c_y}{\partial B^l} \right) \right] + \\
& + \mu \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} - \frac{\partial u_y}{\partial c_y} - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right); \tag{52}
\end{aligned}$$

$$\begin{aligned}
& (1 + \lambda) \frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial I_o^h} + \frac{\lambda}{1+\rho} \frac{\partial u_o^h}{\partial I_o^h} - \frac{1+q}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \frac{\partial c_y}{\partial I_o^h} \\
= & -\mu \frac{\pi^{lh}}{1+\rho} \frac{\partial u_o^h}{\partial I_o^h} - \theta \left[\pi^{lh} - (r - q) \frac{\partial c_y}{\partial I_o^h} \right]; \tag{53}
\end{aligned}$$

$$\begin{aligned}
& (1 + \lambda) \frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial B_o^h} + \frac{\lambda}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h} - \frac{1+q}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \frac{\partial c_y}{\partial B_o^h} \\
= & \theta \left[\pi^{lh} + (r - q) \frac{\partial c_y}{\partial B_o^h} \right] - \mu \frac{\pi^{lh}}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h}; \tag{54}
\end{aligned}$$

¹⁵See footnote 10.

$$\begin{aligned}
& (1 + \lambda) \frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial q} + \frac{1}{1 + \rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \left[s(\bullet) - (1 + q) \frac{\partial c_y}{\partial q} \right] \\
= & \mu s^m \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) - \mu s \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial u_o^l}{\partial c_o^l} + \frac{\pi^{lh}}{1 + \rho} \frac{\partial u_o^h}{\partial c_o^h} \right) + \\
& + \theta \left[s + (r - q) \frac{\partial c_y}{\partial q} \right]. \tag{55}
\end{aligned}$$

Denoting by Π the quantity $(1 + \lambda) \frac{\partial u_y}{\partial c_y} - \frac{1+q}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) - \theta(r - q)$ we can state the following Proposition:

Proposition 9 *When the government is maximizing actual lifetime utilities, Pareto efficient interest income taxation in the two income points system requires that the following condition holds:*

$$\begin{aligned}
\Pi \frac{\partial \widetilde{c}_y}{\partial q} = & s(1 + \lambda) \frac{\partial u_y}{\partial c_y} - \theta s - s \Pi \frac{ds}{dB_y} \big|_{B_y=B^l} + \\
& + \mu(s^m - s) \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) - \mu s \left(\frac{\partial \widehat{u}_y}{\partial c_y} - \frac{\partial u_y}{\partial c_y} \right). \tag{56}
\end{aligned}$$

Proof. See Appendix. ■

The term Π appearing in (56) measures the effect on the Lagrangian of the government's problem coming from a marginal reallocation of consumption of agents across periods, and in particular from a marginal reduction in the level of savings of young agents. When the objective function of the government entailed maximization of expected utility, we could use the envelope theorem to conclude that this effect reduced to the budget term $-\theta(r - q)$. This is no longer the case when the government, engaged in Pareto efficient taxation, aims at maximizing the actual lifetime utility of a given group of agents. In this setting, the discrepancy between what government maximizes in selecting the fiscal variables and what young agents maximize in taking their decisions implies that we cannot invoke the envelope theorem to reduce the total effect to a pure budget effect. Instead, this is also made up by the welfare terms $\frac{\partial u_y}{\partial c_y} - \frac{1+q}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_y}{\partial c_y} - \lambda \frac{1+q}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h}$. Apart from this, eq. (56) resembles closely the corresponding one for the expected utility cases, eq. (30), to which it actually reduces when substituting $-\theta(r - q)$ for Π and taking into account that $\lambda = 0$ in the expected utility case (since we do not have a “minimum utility level” constraint).

The distortions imposed on high skilled agents are described in Proposition 10.

Proposition 10 *When the government is maximizing actual lifetime utilities, Pareto efficient taxation in the two income points system implies that*

the marginal (labor) income tax rate and the METR faced by old high skilled agents are respectively:

$$T'_{o(h)} = -\frac{\Pi}{\theta\pi^{lh}} \left(\frac{dc_y}{dI_o^h} + \frac{dc_y}{dB_o^h} MRS_{I,c}^{o(h)} \right), \quad (57)$$

$$\tau_o'^h = \left(1 + \frac{\theta\pi^{lh}(r-q)}{\Pi} \right) T'_{o(h)}. \quad (58)$$

Proof. See Appendix. ■

Even in this case substituting $-\theta(r-q)$ for Π is sufficient to recover the corresponding equations for the expected utility case, eq. (31) and (35). Since the basic mechanisms at work in the two cases are the same, with the only complication brought about by the limited possibility to exploit the envelope theorem, considerations close to the ones exposed in the comment of (31) and (35) can be provided here.

Now consider the point along the income tax schedule where young agents and old low skilled ones are pooled. As in the expected utility case we will provide in the following Proposition a result for the implementing tax structure which holds for the case when savings of young agents are negative. The same argument put forward in that case applies here to derive the corresponding result for the case when savings of young agents are positive.

Proposition 11 *When the government is maximizing actual lifetime utilities and savings of young agents are negative, the optimal allocation in the two income points system can be implemented through a labor income tax schedule whose left-hand- and right-hand derivatives at the common point in the (I, B) -space for young workers and old low skilled workers are respectively given by:*

$$\begin{aligned} T'_{(left)}(I^l) &= \frac{\mu}{\theta} \frac{\pi^{lh}}{1+\pi^{ll}} \frac{1}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^y - \widehat{MRS}_{I,c} \right) + \\ &+ \frac{\mu}{\theta} \frac{\pi^{ll}}{1+\pi^{ll}} \frac{1}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^y - \widehat{\widehat{MRS}}_{I,c} \right) + \\ &+ \frac{\mu}{\theta(1+\pi^{ll})} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) + \\ &+ \underbrace{\frac{1+\mu\pi^{ll}}{\theta(1+\pi^{ll})}}_G \frac{1}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^y \right) + \\ &- \frac{\Pi}{\theta(1+\pi^{ll})} \left[\frac{\partial c_y}{\partial I^l} - \left(1 - \frac{\partial c_y}{\partial B^l} \right) MRS_{I,c}^y \right]; \quad (59) \end{aligned}$$

$$\begin{aligned}
T'_{(right)}(I^l) &= \frac{\mu}{\theta} \frac{\pi^{lh}}{1 + \pi^{ll}} \frac{1}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c} \right) + \\
&+ \frac{\mu}{\theta} \frac{\pi^{ll}}{1 + \pi^{ll}} \frac{1}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} \right) + \\
&+ \frac{\mu}{\theta (1 + \pi^{ll})} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^m \right) + \\
&+ \frac{1 + \lambda + \mu}{\theta (1 + \pi^{ll})} \frac{\partial u_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^{o(l)} \right) + \\
&- \frac{\Pi}{\theta (1 + \pi^{ll})} \left[\frac{\partial c_y}{\partial I^l} - \left(1 - \frac{\partial c_y}{\partial B^l} \right) MRS_{I,c}^{o(l)} \right]. \quad (60)
\end{aligned}$$

Proof. See Appendix. ■

Notice again that, substituting $-\theta(r - q)$ for Π in (60), and taking into account that $\lambda = 0$ in the expected utility case, allows us to recover expression (39). This is instead not sufficient in order to obtain (38) from (59). The formal reason is that at the numerator of term labelled G in (59) we have $1 + \mu\pi^{ll}$ whereas in the corresponding term of (38) it is $(1 + \mu)\pi^{ll} = \pi^{ll} + \mu\pi^{ll}$. Intuitively, this happens because in the actual lifetime utility case the objective function of the government assigns unitary weight to the utility of old low skilled agents, whereas in the expected utility case the corresponding weight is π^{ll} .

Looking at the METRs, the following result holds:

Proposition 12 *When the government is maximizing actual lifetime utilities, Pareto efficient taxation in the two income points system implies that the METRs faced by young agents and old low skilled ones are respectively:*

$$\begin{aligned}
\tau'_y &= \frac{\mu}{\theta} \frac{\pi^{lh}}{1 + \pi^{ll}} \frac{1}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^y - \widehat{MRS}_{I,c} \right) + \\
&+ \frac{\mu}{\theta} \frac{\pi^{ll}}{1 + \pi^{ll}} \frac{1}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^y - \widehat{\widehat{MRS}}_{I,c} \right) + \\
&+ \frac{\mu}{\theta (1 + \pi^{ll})} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) + \\
&+ \underbrace{\frac{1 + \mu\pi^{ll}}{\theta (1 + \pi^{ll})}}_H \frac{1}{1 + \rho} \frac{\partial u_o^l}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^y \right) + \\
&- \left(\frac{r - q}{1 + r} + \frac{\Pi}{\theta (1 + \pi^{ll})} \right) \left[\frac{\partial c_y}{\partial I^l} - \left(1 - \frac{\partial c_y}{\partial B^l} \right) MRS_{I,c}^y \right]; \quad (61)
\end{aligned}$$

$$\begin{aligned}
\tau_o^n = & \frac{\mu}{\theta} \frac{\pi^{lh}}{1 + \pi^{ll}} \frac{1}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c} \right) + \\
& + \frac{\mu}{\theta} \frac{\pi^{ll}}{1 + \pi^{ll}} \frac{1}{1 + \rho} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} \right) + \\
& + \frac{\mu}{\theta (1 + \pi^{ll})} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^m \right) + \\
& + \frac{1 + \lambda + \mu}{\theta (1 + \pi^{ll})} \frac{\partial u_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^{o(l)} \right) + \\
& - \left(\frac{\Pi}{\theta (1 + \pi^{ll})} + r - q \right) \left[\frac{\partial c_y}{\partial I^l} - \left(1 - \frac{\partial c_y}{\partial B^l} \right) MRS_{I,c}^{o(l)} \right]. \quad (62)
\end{aligned}$$

Proof. See Appendix. ■

Once again, substituting $-\theta(r - q)$ for Π (and $\lambda = 0$) in (62) we are back to the corresponding expression for the expected utility case, eq. (41). Performing this substitutions is not sufficient to recover (40) from (61) since at the numerator of term labelled H in (61) we have $1 + \mu\pi^{ll}$ whereas in the corresponding term of (40) it was $(1 + \mu)\pi^{ll}$.

4.4 Case 2: The Actual Lifetime Utility Case with a “Three Points” System

When the government is free to make the nonlinear income tax schedule dependent on the observable characteristic “age”, its problem becomes the following:

$$\begin{aligned}
& \max_{I_y, I_o^l, I_o^h, B_y, B_o^l, B_o^h, q} u \left(c_y \left(I_y, I_o^l, I_o^h, B_y, B_o^l, B_o^h, q, \pi^{lh} \right), I_y \right) + \\
& + \frac{1}{1 + \rho} u \left((B_y - c_y(\bullet)) (1 + q) + B_o^l, I_o^l \right)
\end{aligned}$$

subject to

$$\begin{aligned}
& u \left(c_y \left(I_y, I_o^l, I_o^h, B_y, B_o^l, B_o^h, q, \pi^{lh} \right), I_y \right) + \\
& + \frac{1}{1 + \rho} u \left((B_y - c_y(\bullet)) (1 + q) + B_o^h, I_o^h \right) \geq \overline{V}, \quad (\lambda)
\end{aligned}$$

$$\begin{aligned}
& u(B_y - s(\bullet), I_y) + \frac{\pi^{ll}}{1 + \rho} u\left((1 + q)s(\bullet) + B_o^l, I_o^l\right) + \\
& + \frac{\pi^{lh}}{1 + \rho} u\left((1 + q)s(\bullet) + B_o^h, I_o^h\right) \\
\geq & u\left(B_y - s^m\left(I_y, I_o^l, B_y, B_o^l, q, \pi^{lh}\right), I_y\right) + \frac{\pi^{ll}}{1 + \rho} u\left((1 + q)s^m(\bullet) + B_o^l, I_o^l\right) + \\
& + \frac{\pi^{lh}}{1 + \rho} \widehat{u}\left((1 + q)s^m(\bullet) + B_o^l, I_o^l\right), \tag{\mu}
\end{aligned}$$

$$(I_y - B_y) + \pi^{ll}(I_o^l - B_o^l) + \pi^{lh}(I_o^h - B_o^h) + (r - q)(B_y - c_y(\bullet)) = 0. \tag{\theta}$$

We have already noticed that switching from the two points system to the three points system the first order conditions referred to the pre-tax labor income and after-tax labor income for old high skilled and to the net rate of return on savings do not change formally. As regards the first order conditions of the government's problem with respect to the pre-tax labor income and after-tax labor income for young agents and old low skilled agents, they are respectively given by:

$$\begin{aligned}
& (1 + \lambda) \left[\frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial I_y} + \frac{\partial u_y}{\partial I_y} \right] - \frac{1 + q}{1 + \rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \frac{\partial c_y}{\partial I_y} \\
= & \mu \left(\frac{\partial \widehat{u}_y}{\partial I_y} - \frac{\partial u_y}{\partial I_y} \right) - \theta \left[1 - (r - q) \frac{\partial c_y}{\partial I_y} \right]; \tag{63}
\end{aligned}$$

$$\begin{aligned}
& (1 + \lambda) \frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial B_y} + \frac{1 + q}{1 + \rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \left(1 - \frac{\partial c_y}{\partial B_y} \right) \\
= & \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} - \frac{\partial u_y}{\partial c_y} \right) - \theta \left[-1 + (r - q) \left(1 - \frac{\partial c_y}{\partial B_y} \right) \right]; \tag{64}
\end{aligned}$$

$$\begin{aligned}
& (1 + \lambda) \frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial I_o^l} - \frac{1 + q}{1 + \rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \frac{\partial c_y}{\partial I_o^l} + \frac{1}{1 + \rho} \frac{\partial u_o^l}{\partial I_o^l} \\
= & \mu \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial \widehat{u}}{\partial I_o^l} + \frac{\pi^{lh}}{1 + \rho} \frac{\partial \widehat{u}}{\partial I_o^l} \right) - \mu \frac{\pi^{ll}}{1 + \rho} \frac{\partial u_o^l}{\partial I_o^l} + \\
& - \theta \left[\pi^{ll} - (r - q) \frac{\partial c_y}{\partial I_o^l} \right]; \tag{65}
\end{aligned}$$

$$\begin{aligned}
& (1 + \lambda) \frac{\partial u_y}{\partial c_y} \frac{\partial c_y}{\partial B_o^l} + \frac{1}{1 + \rho} \frac{\partial u_o^l}{\partial c_o^l} - \frac{1 + q}{1 + \rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \frac{\partial c_y}{\partial B_o^l} \\
= & \mu \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial \hat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1 + \rho} \frac{\partial \hat{u}}{\partial c_o^l} \right) - \mu \left(\frac{\pi^{ll}}{1 + \rho} \frac{\partial u_o^l}{\partial c_o^l} \right) + \\
& + \theta \left[\pi^{ll} + (r - q) \frac{\partial c_y}{\partial B_o^l} \right]. \tag{66}
\end{aligned}$$

Proposition 13 *When the government is maximizing actual lifetime utilities, Pareto efficient interest income taxation in the three income points system requires that the following condition holds:*

$$-\Pi \frac{\partial \tilde{c}_y}{\partial q} = \mu \left(\frac{\pi^{lh}}{1 + \rho} \frac{\partial \hat{u}}{\partial c_o^l} + \frac{\pi^{ll}}{1 + \rho} \frac{\partial \hat{u}}{\partial c_o^l} \right) (s - s^m). \tag{67}$$

Proof. See Appendix. ■

Eq. (67) closely parallels eq. (46). The only difference refers to the term by which $\frac{\partial \tilde{c}_y}{\partial q}$ is multiplied on the left-hand side. In the expected utility case the fundamental trade-off to be faced in considering the optimal interest income tax rate is that between the desire to discourage mimicking and the effects on tax revenue. The screening virtues of interest income taxation can be exploited to implement a compensated marginal change in q that makes the mimicker worse off. Since the reform is accomplished in a compensated way, the expected utility of non mimickers is not affected. Thus, when the government's objective function entails maximization of expected utility, the only side (cost) effect of such a reform is represented by a budget effect coming from the adjustments in the hicksian demands. Things are instead different when the government aims at maximizing the actual lifetime utility for those who remain low skilled during the entire life. In that case the discrepancy between what the government wants to maximize and what agents actually maximize in taking their consumption/savings decision implies that the adjustments in the hicksian demands are not welfare neutral from the government's perspective. This is the reason why, on the left-hand side of (67), $-\Pi$, a quantity which is made up also by welfare terms, replaces $\theta(r - q)$.

Turning to the old high skilled agents, we have that the formal expressions for the marginal (labor) income tax rate and METR do not change switching from the two income points system to the three income points system. However, as already observed for the expected utility case, this does not mean that when we move from a system to the other the value of those tax rates are unaffected. This is obvious if we think that the optimal wedges between gross- and net rate of return to savings required by conditions (56) and (67) are different.

As regards the young agents and the old low skilled ones, Propositions 14 and 15 provide the relevant results.

Proposition 14 *When the government is maximizing actual lifetime utilities, Pareto efficient taxation in the three income points system implies that the marginal (labor) income tax rate T'_y and the METR τ'_y faced by young agents are respectively:*

$$T'_y = \frac{\mu}{\theta} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) - \frac{\Pi}{\theta} \left[\frac{\partial c_y}{\partial I_y} - \left(1 - \frac{\partial c_y}{\partial B_y} \right) MRS_{I,c}^y \right]; \quad (68)$$

$$\begin{aligned} \tau'_y &= \left(1 + \frac{\theta}{\Pi} \frac{r-q}{1+r} \right) T'_y - \frac{\mu}{\Pi} \frac{r-q}{1+r} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) = \\ &= \frac{\mu}{\theta} \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) + \\ &\quad - \frac{\Pi(1+r) + \theta(r-q)}{\theta(1+r)} \left[\frac{\partial c_y}{\partial I_y} - \left(1 - \frac{\partial c_y}{\partial B_y} \right) MRS_{I,c}^y \right]. \end{aligned} \quad (69)$$

Proof. See Appendix. ■

Proposition 15 *When the government is maximizing actual lifetime utilities, Pareto efficient taxation in the three income points system implies that the marginal (labor) income tax rate T_o^l and the METR τ_o^l faced by old low skilled agents are respectively:*

$$\begin{aligned} T_o^l &= \frac{\mu}{\theta} \frac{1}{1+\rho} \left[\frac{\pi^{lh}}{\pi^{ll}} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c} \right) + \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} \right) \right] + \\ &\quad - \frac{\Pi}{\theta \pi^{ll}} \left(\frac{\partial c_y}{\partial I_o^l} + \frac{\partial c_y}{\partial B_o^l} MRS_{I,c}^{o(l)} \right); \end{aligned} \quad (70)$$

$$\begin{aligned} \tau_o^l &= \left(1 + \frac{\theta \pi^{ll}}{\Pi} (r-q) \right) T_o^l + \\ &\quad - \frac{\mu}{\Pi} \frac{r-q}{1+\rho} \left[\pi^{lh} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c} \right) + \pi^{ll} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} \right) \right] \\ &= \frac{\mu}{\theta} \frac{1}{1+\rho} \left[\frac{\pi^{lh}}{\pi^{ll}} \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c} \right) + \frac{\partial \widehat{u}}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - \widehat{\widehat{MRS}}_{I,c} \right) \right] + \\ &\quad - \frac{\Pi + \theta \pi^{ll} (r-q)}{\theta \pi^{ll}} \left(\frac{\partial c_y}{\partial I_o^l} + \frac{\partial c_y}{\partial B_o^l} MRS_{I,c}^{o(l)} \right). \end{aligned} \quad (71)$$

Proof. See Appendix. ■

Notice that it would be sufficient to substitute $-\theta(r-q)$ for Π in (68), (69), (70) and (71) in order to recover the corresponding equations derived for the expected utility case.

5 A Comparison between the Models with and without Savings

In this Section we would like to call attention to a qualitative difference underlying the switch from a two- to a three income points system in a model without savings and in a model with savings. In a model without savings, ranking people according to the slopes of the indifference curves in the (I, B) -space, the policy maker actually faces two different groups of agents: the low skilled and the high skilled.^{16,17} This is because within the low skilled group, at any given point in the (I, B) -space, the indifference curve for a young agent has the same slope as the one for an old agent. With two different groups in the population the two income points system would in this case perform as well as it does the second-best Pareto efficient allocation of the atemporal Stiglitz (1982) model. In the no-savings case the possibility to observe age and implement a three income points systems allows to achieve a Pareto-improvement as it would have happened in the atemporal model if we had found an exogenous way to identify some of the low skilled agents and be sure that nobody could try to mimic them.¹⁸

Things are instead qualitatively different when we analyze the case with savings. The reason is that, in spite of the assumption of age-independent preferences, when we allow for the possibility to save, it is no longer true that, at any given point in the (I, B) -space, the indifference curves of a young and of an old low skilled agent are equally sloped. Instead, depending on the sign of s , if young agents are borrowers or savers, it is possible to predict which ones will be flatter and which ones steeper. But this in turn means that in this case the policy maker is actually facing three different groups of people and that it should really try to offer three different bundles. Now the natural counterpart of the atemporal Stiglitz model would be the three income points system where age is not observable but the government exploits the information on the correlation between skill and age in order to select three different points in the (I, B) -space. Contrary to what happened in the case without savings, the impossibility to observe age does not prevent the policy maker doing better by offering three different points than by pooling young and old low skilled agents at the same allocation. In fact, suppose that the government has, as it is in our model, the information on the joint distribution of skill and age but it does not use it to try to implement a three

¹⁶We have already observed that under reasonable assumptions about the magnitude of the parameters involved, young agents in our model would like to borrow money. The model without savings can therefore be interpreted as an extreme version of a model involving liquidity constraints.

¹⁷Notice that this happens because we have assumed preferences that are age-independent.

¹⁸Here the term “exogenous” is used to refer to a case where there is no need of self-selection devices in order to induce people to reveal their true type.

income points system. When age is not observable this does not entail any loss if there are no savings; however, if we introduce savings in the model, the same attitude would imply that the policy maker is actually not optimizing the use of the information at its disposal, and designing a two income points system would amount at inefficiently constraining the set of feasible taxes it can impose. Therefore, switching from a two income points system to the three income points that we have analyzed in this paper (where age is observable) entails in the case with savings a double gain.

6 Concluding Remarks

The usual atemporal optimal income taxation model has been thoroughly explored. In this paper we make several important extensions of that model. We believe it is an important characteristic of real economies that individuals' income paths widen with age. Some individuals stay low skill all their life, whereas others are low skill when young but high skill when old. This means that some individuals are low skill part of their life and high skill another part of life. This should have implications for how the tax system is constructed. In this paper we study if it is possible to achieve redistribution in a more efficient way by making the income tax schedule dependent on age.

To pursue our analysis we use a simple OLG model where everybody lives for two periods. In the first period everyone is low skilled and *ex ante* equal. As individuals age and move into the second period of their lives, a proportion π^{lh} of the population becomes high skilled whereas the rest stays low skilled. Individuals maximize expected lifetime utility. The objective of the policy maker (which can only use annual income as tax base) is to maximize actual lifetime utility for those who remain low skilled during the entire life subject to a minimum level of actual lifetime utility for those who become high skilled in the second period. Thus, there is a discrepancy between the maximand of the policy maker and that of an individual. We study two versions of the model. In the simplest we assume that individuals cannot save. Given this set-up we find that if a two points income tax system is used, the usual result with high skilled being undistorted and low skilled, both young and old, distorted, facing a positive marginal income tax rate, is recovered. We show how a strict Pareto improvement can be obtained if an age dependent three points income tax system is used. Since the young low skilled cannot be mimicked, the Pareto improvement is accomplished by offering them a point where they obtain the same utility as before, but where their consumption/leisure choice is undistorted. This implies that resources are released so that old low skilled people can be located at a point where they obtain a higher utility than in the two points income tax system. In terms of lifetime utilities the expected lifetime utility of individuals has gone

up. The actual lifetime utility of people being low skilled in both periods has increased whereas the lifetime utility of those who are high skilled in the second period is unchanged. The changes in consumption and work are as follows: the old high skilled would perform as before; the young low skilled would work more and consume more; finally, the old low skilled would work less and also have less consumption.

In the second version of the model we assume the economy has access to the international capital market, implying that individuals can freely save or dissave at the prevailing rate r . In this case, it turns out to be convenient to look, as intermediate step, at the case when the government maximizes the expected utility of agents. The corresponding formulas for the actual lifetime utility case can then be easily derived (and similarly interpreted) but are analytically more complex due to the limited possibility to use the envelope theorem when the maximand of the policy maker and of an individual differ. In the paper we have characterized both the age independent and the age dependent income tax systems. Since we consider the age dependent income tax system more interesting and since it also gives rise to simpler tax formulas we focus in this summary on the age dependent tax.

There should be a tax/subsidy on savings. This tax serves the purpose of deterring mimicking. We regard it as plausible that the mimicker would save more than a “fair” agent, implying that there should actually be a tax on interest income. (In the age independent tax system there are also other aspects that affect how the interest tax should be set.) In contrast to the case without savings, if individuals have access to the international capital market, the marginal income tax faced by high skill individuals should be non-zero. The reason being that a distortion of their labor supply have budget effects because of the tax on interest income. In contrast also to results in atemporal models, the effective marginal tax should be non-zero. This is because of an effect that only arises in an intertemporal model with savings. Changing the bundle offered to old high skilled agents does not only affect the total amount of taxes paid by this sub-set of the population but, through the savings function, it also affects the amount of interest income taxes collected from the old low skilled individuals.

The result that young people should be undistorted is also modified when savings is a possibility. The reasons are not the same, but similar to the reasons why old high skilled should be distorted. Finally, old low skilled should be distorted. There are several reasons for this. First, there is the reason that “classical” mimicking should be deterred. An old high skill person should not prefer the income/tax point intended for an old low skill. Second, in the expression for the effective marginal tax for old low skill individuals there are also terms due to deterrence of mimicking in terms of savings behavior. Third, there is a budget term which is the counterpart of the budget term appearing in the formula for the effective marginal tax

rate faced by the high skilled agents: altering the bundle offered to old low skilled agents does not only affect the total amount of taxes paid by this sub-set of the population but it also affects the amount of interest income taxes collected from the high skilled individuals.

We have characterized an optimal age dependent tax both under the assumption that individuals cannot save and allowing for the possibility to save. What is the most policy relevant case? This is in the end an empirical question. However, given that the expected income in the second period is higher than the income when young, individuals would probably like to borrow when young and pay back when old. In such a situation individuals would be unable to borrow when young. Hence, the case where individuals cannot save might be the most relevant one.

The model we have studied in this paper is quite abstract, as are many optimal tax models, and the results obtained cannot be directly applied in the design of tax systems. However, we have shown that there is scope for using age-dependent taxes to increase the efficiency of redistributive fiscal policies. To obtain more detailed insights in how these tax systems should be designed we believe simulation studies, like those originally performed by Mirrlees (1971) in his atemporal model, would be useful. Further, an important issue left unexplored in this paper is how educational incentives would be affected by an age dependent income tax. (Kremer (2002) contains an informal discussion of this topic.)

There may be further arguments than that developed here for using an age dependent income tax. For example, preferences for leisure might vary with age, which implies that even if there were no correlation between average skill level and age, there might still be gains from an age dependent tax.

7 Appendix

7.1 Fig. 2-5.

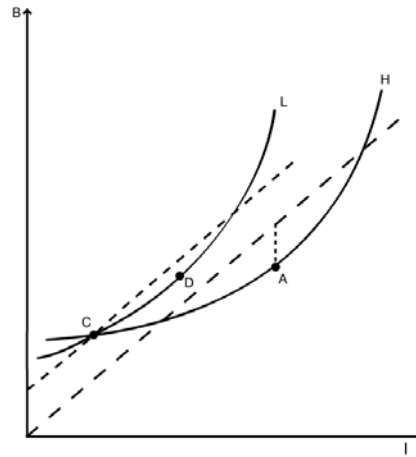


Fig. 2

Suppose that C is the bundle for young agents, A the one for old high skilled and D that for old low skilled: this solution is not implementable since the old high skilled would like to mimic the old low skilled and the self-selection constraint is violated.

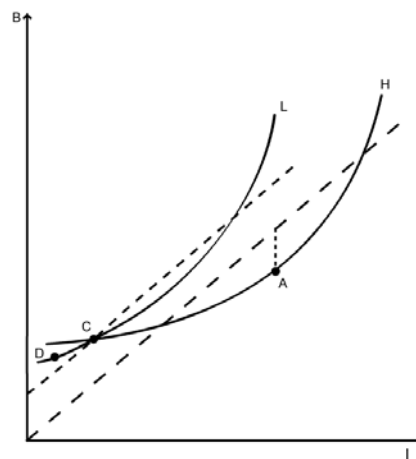


Fig. 3

Suppose that C is the bundle for young agents, A the one for old high skilled and D that for old low skilled: this solution is not optimal since moving the old low skilled from point D to point C increases the revenue collected by the government without violating any self-selection constraints.

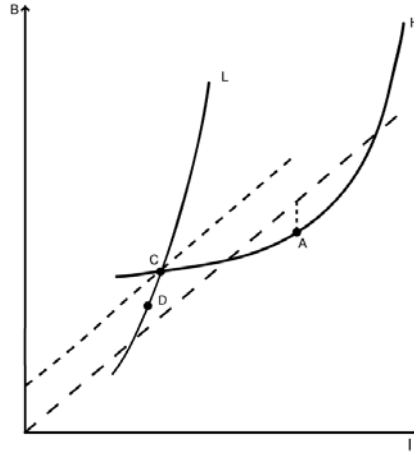


Fig. 4

Suppose that C is the bundle for young agents, A the one for old high skilled and D that for old low skilled: this solution is not optimal since moving the young from point C to point D increases the revenue collected by the government and slackens the binding self-selection constraint preventing old high skilled to mimic young agents.

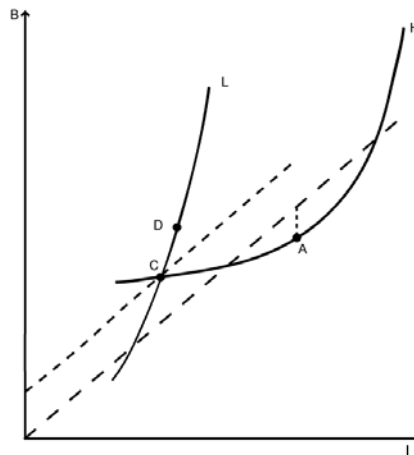


Fig. 5

Suppose that C is the bundle for young agents, A the one for old high skilled and D that for old low skilled: this solution is not implementable since the old high skilled would like to mimic the old low skilled and the self-selection constraint is violated as it happened in fig. 2.

7.2 Proof of Proposition 1

Rearranging f.o.c. (26) and (28), we have:

$$\theta(r-q) \frac{\partial c_y}{\partial B^l} = \theta(r-q) - \theta(1+\pi^{ll}) + \frac{\partial u_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} - \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}_o}{\partial c_o^l} \right) - \mu \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}_o}{\partial c_o^l} - \frac{\partial u_y}{\partial c_y} - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right); \quad (72)$$

$$\theta(r-q) \frac{\partial c_y}{\partial B_o^h} = -\theta\pi^{lh} + \frac{\pi^{lh}}{1+\rho} (1+\mu) \frac{\partial u_o^h}{\partial c_o^h}. \quad (73)$$

Taking into account that in the two income points system $B_o^l = B_y = B^l$ and using the Slutsky-type decomposition $\frac{\partial c_y}{\partial q} = \frac{\partial \widetilde{c}_y}{\partial q} + s \left(\frac{\partial c_y}{\partial B_o^l} + \frac{\partial c_y}{\partial B_o^h} \right)$ we can rewrite f.o.c. (29) as

$$\left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h} \right) s = \theta \left\{ s + (r-q) \left[\frac{\partial \widetilde{c}_y}{\partial q} + s \left(\underbrace{\frac{\partial c_y}{\partial B_o^l}}_L \Big|_{B_o^l=B^l} + \frac{\partial c_y}{\partial B_o^h} \right) \right] \right\} + \mu s^m \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}_o}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}_o}{\partial c_o^h} \right) - \mu s \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h} \right). \quad (74)$$

Summing up (72) and (73), and multiplying all terms by s , we can obtain the following expression for $\theta s(r-q) \left(\frac{\partial c_y}{\partial B^l} + \frac{\partial c_y}{\partial B_o^h} \right)$:

$$\theta(r-q) s \left(\frac{\partial c_y}{\partial B^l} + \frac{\partial c_y}{\partial B_o^h} \right) = \theta s(r-q) - \theta s(1+\pi^{ll}) + \frac{\pi^{ll}}{1+\rho} s \frac{\partial u_o^l}{\partial c_o^l} - \theta s \pi^{lh} + \frac{\pi^{lh}}{1+\rho} s (1+\mu) \frac{\partial u_o^h}{\partial c_o^h} - \mu s \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}_o}{\partial c_o^l} \right) - \mu s \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}_o}{\partial c_o^l} - \frac{\partial u_y}{\partial c_y} - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) + s \frac{\partial u_y}{\partial c_y}. \quad (75)$$

Since $\frac{\partial c_y}{\partial B^l} = \frac{\partial c_y}{\partial B^l} = \frac{\partial c_y}{\partial B_y} \Big|_{B_y=B^l} + \frac{\partial c_y}{\partial B_o^l} \Big|_{B_o^l=B^l}$, we can derive from (75) an expression for $\theta(r-q) s \left(\frac{\partial c_y}{\partial B_o^l} \Big|_{B_o^l=B^l} + \frac{\partial c_y}{\partial B_o^h} \right)$ that can be substituted for the corresponding term in (74). Simplifying and collecting terms gives eq. (30).

7.3 Proof of Proposition 4

Write the f.o.c. (25) and (26) as:

$$\frac{\partial u_y}{\partial I^l} = -\theta \left[1 + \pi^{ll} - (r-q) \frac{\partial c_y}{\partial I^l} \right] - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} - \mu \left(\frac{\partial u_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} \right) + \mu \frac{\partial \widehat{u}_y}{\partial I^l} + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}_o}{\partial I^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}_o}{\partial I^l} \right); \quad (76)$$

$$\begin{aligned} \frac{\partial u_y}{\partial c_y} &= \theta \left[1 + \pi^{ll} - (r - q) \left(1 - \frac{\partial c_y}{\partial B^l} \right) \right] - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) + \\ &\mu \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} - \frac{\partial u_y}{\partial c_y} - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right). \end{aligned} \quad (77)$$

Dividing (76) by (77) and multiplying the result by the expression on the r.h.s. of (77) gives:

$$\begin{aligned} \mu \left(\frac{\partial \widehat{u}_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} \right) &= \theta \left[1 + \pi^{ll} - (r - q) \frac{\partial c_y}{\partial I^l} \right] + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} + \mu \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} + \\ \frac{\partial u_y}{\partial I^l} \theta \left(1 + \pi^{ll} \right) - \frac{\partial u_y}{\partial c_y} \theta (r - q) \left(1 - \frac{\partial c_y}{\partial B^l} \right) - \frac{\partial u_y}{\partial c_y} \left[\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} - \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) \right] + \\ \frac{\partial u_y}{\partial I^l} \mu \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} - \frac{\partial u_y}{\partial c_y} - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) &+ \mu \frac{\partial u_y}{\partial I^l}. \end{aligned} \quad (78)$$

From (78) it is possible to collect terms in order to get an expression for $\theta (1 + \pi^{ll}) \left(1 + \frac{\partial u_y}{\partial I^l} / \frac{\partial u_y}{\partial c_y} \right)$:

$$\begin{aligned} \theta (1 + \pi^{ll}) \left(1 + \frac{\partial u_y}{\partial I^l} / \frac{\partial u_y}{\partial c_y} \right) &= \frac{\mu \pi^{lh} \frac{\partial \widehat{u}}{\partial c_o^l}}{1+\rho} \left(MRS_{I,c}^y - \widehat{MRS}_{I,c} \right) + \frac{\mu \pi^{ll} \frac{\partial \widehat{u}}{\partial c_o^l}}{1+\rho} MRS_{I,c}^y - \\ \frac{\mu \pi^{ll} \frac{\partial \widehat{u}}{\partial c_o^l}}{1+\rho} \widehat{MRS}_{I,c} + \mu \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^m \right) &+ \frac{(1+\mu) \pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^y \right) + \\ \theta (r - q) \left[\frac{dc_y}{dI^l} - \left(1 - \frac{dc_y}{dB^l} \right) MRS_{I,c}^y \right]. \end{aligned} \quad (79)$$

The assumption of normality of all goods in the economy implies that, when savings of young agents are negative, the young agents have indifference curves in the (I, B) -space which are steeper than the ones for old low skilled agents. We also know that at a bunching point there exists an implementing tax structure whose left-hand derivative is equal to $1 - MRS$ of those agents with the highest level of MRS among those who are pooled; since in this particular case the young agents are those with the highest value of the marginal rate of substitution, this means that it will exist an implementing tax structure at the bunching point I^l whose left-hand derivative is equal to $1 + \frac{\partial u_y}{\partial I^l} / \frac{\partial u_y}{\partial c_y}$. Eq. (38) is then obtained from (79) dividing both sides by $\theta (1 + \pi^{ll})$.

For the second part of Proposition 4, write the f.o.c. (25) and (26) as:

$$\begin{aligned} \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} &= -\theta \left[1 + \pi^{ll} - (r - q) \frac{\partial c_y}{\partial I^l} \right] - \frac{\partial u_y}{\partial I^l} - \mu \left(\frac{\partial u_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} \right) + \mu \frac{\partial \widehat{u}_y}{\partial I^l} + \\ \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} \right); \end{aligned} \quad (80)$$

$$\begin{aligned} \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} &= \theta \left[1 + \pi^{ll} - (r - q) \left(1 - \frac{\partial c_y}{\partial B^l} \right) \right] - \frac{\partial u_y}{\partial c_y} + \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) + \\ \mu \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} - \frac{\partial u_y}{\partial c_y} - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right); \end{aligned} \quad (81)$$

Dividing (80) by (81) and multiplying the result by the expression on the r.h.s. of (81) gives:

$$\begin{aligned} \mu \left(\frac{\partial \widehat{u}_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} - \frac{\partial u_y}{\partial I^l} \right) &= \theta \left[1 + \pi^{ll} - (r-q) \frac{\partial c_y}{\partial I^l} \right] + \theta \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l} + \\ \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l} \left\{ \theta \left[\pi^{ll} - (r-q) \left(1 - \frac{\partial c_y}{\partial B^l} \right) \right] - \frac{\partial u_y}{\partial c_y} + \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) + \mu \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) \right\} - \\ \mu \frac{\partial u_y}{\partial c_y} \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l} + \frac{\partial u_y}{\partial I^l}. \end{aligned} \quad (82)$$

Collecting terms in (82) gives:

$$\begin{aligned} \theta (1 + \pi^{ll}) \left(1 + \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l} \right) &= \frac{\mu \pi^{lh} \frac{\partial \widehat{u}}{\partial c_o^l}}{1+\rho} \left(MRS_{I,c}^{o(l)} - \widehat{MRS}_{I,c} \right) + \frac{\mu \pi^{ll} \frac{\partial \widehat{u}}{\partial c_o^l}}{1+\rho} MRS_{I,c}^{o(l)} - \\ \frac{\mu \pi^{ll} \frac{\partial \widehat{u}}{\partial c_o^l}}{1+\rho} \widehat{\widehat{MRS}}_{I,c} + \mu \frac{\partial \widehat{u}_y}{\partial c_y} \left(MRS_{I,c}^{o(l)} - MRS_{I,c}^m \right) &+ (1 + \mu) \frac{\partial u_y}{\partial c_y} \left(MRS_{I,c}^y - MRS_{I,c}^{o(l)} \right) + \\ \theta (r-q) \left[\frac{\partial c_y}{\partial I^l} - \left(1 - \frac{\partial c_y}{\partial B^l} \right) MRS_{I,c}^{o(l)} \right]. \end{aligned} \quad (83)$$

Since we know that at a bunching point there exists an implementing tax structure whose right-hand derivative is equal to $1 - MRS$ of those agents with the lowest level of MRS among those who are pooled, which in our case are the old low skilled agents, there will exist an implementing tax structure whose right-hand derivative at the bunching point I^l is equal to $1 + \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l}$. Eq. (39) is then obtained from (83) dividing both sides by $\theta (1 + \pi^{ll})$.

7.4 Proof of Proposition 5

The METR faced by young agents is given by:

$$\tau'_y = 1 - \frac{r-q}{1+r} \frac{\partial c_y}{\partial I^l} + \frac{\partial u_y}{\partial I^l} \left[1 - \frac{r-q}{1+r} \left(1 - \frac{\partial c_y}{\partial B^l} \right) \right]. \quad (84)$$

Substituting in (84) the value for $1 + \frac{\partial u_y}{\partial I^l} / \frac{\partial u_y}{\partial c_y}$ derived from (79) and collecting terms gives eq. (40).

The METR faced by old low skilled agents is instead given by:

$$\begin{aligned} \tau'_{o(l)} &= 1 + (r-q) \frac{\partial s}{\partial I^l} - \frac{\frac{\partial u_o^l}{\partial I^l}}{\frac{\partial u_o^l}{\partial c_o^l}} \left[(r-q) \frac{\partial s}{\partial B^l} - 1 \right] = 1 - (r-q) \frac{\partial c_y}{\partial I^l} + \frac{\frac{\partial u_o^l}{\partial I^l}}{\frac{\partial u_o^l}{\partial c_o^l}} - \\ \frac{\frac{\partial u_o^l}{\partial I^l}}{\frac{\partial u_o^l}{\partial c_o^l}} (r-q) \left(1 - \frac{\partial c_y}{\partial B^l} \right). \end{aligned} \quad (85)$$

Substituting in (85) the value for $1 + \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l}$ derived from (83) and collecting terms gives eq. (41).

7.5 Proof of Proposition 6

Write f.o.c. (45) as

$$\theta(r-q) \frac{\partial c_y}{\partial B_o^l} = -\theta\pi^{ll} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) - \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} \right). \quad (86)$$

Taking into account that in the three income points system term labelled L in (74) reduces to $\frac{\partial c_y}{\partial B_o^l}$, we can use (86) and (73) to derive an expression for $\theta s(r-q) \left(\frac{\partial c_y}{\partial B_o^l} + \frac{\partial c_y}{\partial B_o^h} \right)$ to be substituted in (74). Simplifying and collecting terms gives eq. (46).

7.6 Proof of Proposition 7

Write f.o.c. (42) and (43) as

$$\frac{\partial u_y}{\partial I_y} = \mu \left(\frac{\partial \hat{u}_y}{\partial I_y} - \frac{\partial u_y}{\partial I_y} \right) - \theta \left[1 - (r-q) \frac{\partial c_y}{\partial I_y} \right]; \quad (87)$$

$$\frac{\partial u_y}{\partial c_y} = \mu \left(\frac{\partial \hat{u}_y}{\partial c_y} - \frac{\partial u_y}{\partial c_y} \right) - \theta \left[-1 + (r-q) \left(1 - \frac{\partial c_y}{\partial B_y} \right) \right]. \quad (88)$$

Dividing (87) by (88) and multiplying the result by the r.h.s. of (88) gives:

$$\frac{\frac{\partial u_y}{\partial I_y}}{\frac{\partial u_y}{\partial c_y}} \left\{ \mu \left(\frac{\partial \hat{u}_y}{\partial c_y} - \frac{\partial u_y}{\partial c_y} \right) - \theta \left[-1 + (r-q) \left(1 - \frac{\partial c_y}{\partial B_y} \right) \right] \right\} = \mu \left(\frac{\partial \hat{u}_y}{\partial I_y} - \frac{\partial u_y}{\partial I_y} \right) - \theta \left[1 - (r-q) \frac{\partial c_y}{\partial I_y} \right].$$

Eq. (47) is obtained from the equation above dividing both sides by θ and then using the implicit definition of the marginal labor income tax rate faced by young agents ($T_y' = 1 + \frac{\partial u_y}{\partial I_y} / \frac{\partial u_y}{\partial c_y}$) to collect terms.

In the three income points system the expression for the METR faced by young agents becomes:

$$\tau_y' = 1 - \frac{r-q}{1+r} \frac{\partial c_y}{\partial I_y} + \frac{\frac{\partial u_y}{\partial I_y}}{\frac{\partial u_y}{\partial c_y}} \left[1 - \frac{r-q}{1+r} \left(1 - \frac{\partial c_y}{\partial B_y} \right) \right]. \quad (89)$$

Substituting in (89) the value of the marginal labor income tax rate $T_y' = 1 + \frac{\partial u_y}{\partial I_y} / \frac{\partial u_y}{\partial c_y}$ given by (47) and collecting terms provides eq. (48).

7.7 Proof of Proposition 8

Write f.o.c. (44) and (45) as

$$\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I_o^l} = -\theta \left[\pi^{ll} - (r-q) \frac{\partial c_y}{\partial I_o^l} \right] - \mu \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I_o^l} + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \hat{u}}{\partial I_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \hat{u}}{\partial I_o^l} \right); \quad (90)$$

$$\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} = \theta \left[\pi^{ll} + (r-q) \frac{\partial c_y}{\partial B_o^l} \right] - \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} \right). \quad (91)$$

Dividing (90) by (91) and multiplying the result by the r.h.s. of (91) gives:

$$\frac{\frac{\partial u_o^l}{\partial I_o^l}}{\frac{\partial u_o^l}{\partial c_o^l}} \left\{ \theta \left[\pi^{ll} + (r-q) \frac{\partial c_y}{\partial B_o^l} \right] - \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} \right) \right\} = -\theta \pi^{ll} + \theta (r-q) \frac{\partial c_y}{\partial I_o^l} - \mu \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I_o^l} + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \hat{u}}{\partial I_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \hat{u}}{\partial I_o^l} \right).$$

Eq. (49) is obtained from the equation above dividing both sides by θ and then using the implicit definition of the marginal labor income tax rate faced by old low skilled agents ($T_o^l = 1 + \frac{\partial u_o^l}{\partial I_o^l} / \frac{\partial u_o^l}{\partial c_o^l}$) to collect terms.

In the three income points system the expression for the METR faced by old low skilled agents becomes:

$$\tau_o^l = 1 - (r-q) \frac{\partial c_y}{\partial I_o^l} + \frac{\frac{\partial u_o^l}{\partial I_o^l}}{\frac{\partial u_o^l}{\partial c_o^l}} \left[1 + (r-q) \frac{\partial c_y}{\partial B_o^l} \right]. \quad (92)$$

Substituting in (92) the value of the marginal labor income tax rate $T_{o(l)}' = 1 + \frac{\partial u_o^l}{\partial I_o^l} / \frac{\partial u_o^l}{\partial c_o^l}$ given by (49) and collecting terms provides eq. (50).

7.8 Proof of Proposition 9

Using the definition of Π ($\Pi \equiv (1+\lambda) \frac{\partial u_y}{\partial c_y} - \frac{1+q}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) - \theta(r-q)$) and rearranging f.o.c. (52) and (54), we have:

$$\Pi \frac{\partial c_y}{\partial B_o^l} \Big|_{B_o^l=B^l} = -\frac{1}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} - \frac{1+q}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) - \Pi \frac{\partial c_y}{\partial B_y} \Big|_{B_y=B^l} + \mu \frac{\partial \hat{u}_y}{\partial c_y} + \mu \frac{\pi^{ll}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} + \mu \left(\frac{\pi^{lh}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} - \frac{\partial u_y}{\partial c_y} - \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) + \theta (1 + \pi^{ll} - r + q); \quad (93)$$

$$\frac{\partial c_y}{\partial B_o^h} \Pi = \theta \pi^{lh} - \left(\mu \frac{\pi^{lh}}{1+\rho} + \frac{\lambda}{1+\rho} \right) \frac{\partial u_o^h}{\partial c_o^h}. \quad (94)$$

In the two income points system and using the Slutsky-type decomposition f.o.c. (55) can be written as

$$\Pi \left[\frac{\partial \tilde{c}_y}{\partial q} + s \left(\frac{\partial c_y}{\partial B_o^l} \Big|_{B_o^l=B^l} + \frac{\partial c_y}{\partial B_o^h} \right) \right] = \theta s - \frac{1}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) s - \mu s \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} - \mu s \frac{\pi^{lh}}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h} + \mu s^m \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \hat{u}}{\partial c_o^h} \right). \quad (95)$$

Summing up (93) and (94), and multiplying the result by s allows to get an expression for $\Pi s \left(\frac{\partial c_y}{\partial B_o^l} \Big|_{B_o^l=B^l} + \frac{\partial c_y}{\partial B_o^h} \right)$ which can be substituted in (95). Simplifying and collecting terms gives condition (56).

7.9 Proof of Proposition 10

F.o.c. (53) and (54) can be rewritten in a more compact way as

$$\frac{\partial c_y}{\partial I_o^h} \Pi + \theta \pi^{lh} = - \left(\mu \frac{\pi^{lh}}{1+\rho} + \frac{\lambda}{1+\rho} \right) \frac{\partial u_o^h}{\partial I_o^h}, \quad (96)$$

$$\frac{\partial c_y}{\partial B_o^h} \Pi - \theta \pi^{lh} = - \left(\mu \frac{\pi^{lh}}{1+\rho} + \frac{\lambda}{1+\rho} \right) \frac{\partial u_o^h}{\partial c_o^h}. \quad (97)$$

Dividing (96) by (97) and multiplying by $\frac{\partial c_y}{\partial B_o^h} \Pi - \theta \pi^{lh}$ we get

$$\frac{\frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} \left(\frac{\partial c_y}{\partial B_o^h} \Pi - \theta \pi^{lh} \right) = \frac{\partial c_y}{\partial I_o^h} \Pi + \theta \pi^{lh}. \quad (98)$$

Rearranging terms gives

$$\theta \pi^{lh} \left(1 + \frac{\frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} \right) = \Pi \left(\frac{\frac{\partial c_y}{\partial B_o^h} \frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} - \frac{\partial c_y}{\partial I_o^h} \right). \quad (99)$$

Eq. (57) is obtained from eq. (99) using the definition of marginal income tax rate given by (5).

To get eq. (58) multiply both sides of (99) by $\frac{r-q}{\Pi}$: this gives an expression for $(r-q) \left(\frac{\frac{\partial c_y}{\partial B_o^h} \frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} - \frac{\partial c_y}{\partial I_o^h} \right) = -(r-q) \left(\frac{\frac{\partial s}{\partial B_o^h} \frac{\partial u_o^h}{\partial I_o^h}}{\frac{\partial u_o^h}{\partial c_o^h}} - \frac{\partial s}{\partial I_o^h} \right)$ which, substituted in the definition of the METR (see (34)), gives the result.

7.10 Proof of Proposition 11

For eq. (59), write f.o.c. (51) and (52) as

$$-(1+\lambda) \frac{\partial u_y}{\partial I^l} = \Pi \frac{\partial c_y}{\partial I^l} + \frac{1}{1+\rho} \frac{\partial u_o^l}{\partial I^l} + \mu \left(\frac{\partial u_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} \right) - \mu \left(\frac{\partial \hat{u}_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \hat{u}}{\partial I^l} \right) - \mu \frac{\pi^{lh}}{1+\rho} \frac{\partial \hat{u}}{\partial I^l} + \theta (1 + \pi^{ll});$$

$$-(1+\lambda) \frac{\partial u_y}{\partial c_y} = -\Pi \left(1 - \frac{\partial c_y}{\partial B^l}\right) + \frac{1}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} - \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) - \mu \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \mu \left(\frac{\partial u_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) - \theta (1 + \pi^{ll}).$$

Then proceed exactly as in the first part of the proof of Proposition 4 to get an expression for $1 + \frac{\partial u_y}{\partial I^l} / \frac{\partial u_y}{\partial c_y}$, which, for the same reasons provided there, gives again the value of the left-hand derivative at the bunching point I^l for one of the possible implementing tax structures.

For eq. (60), f.o.c. (51) and (52) should be rewritten as

$$-\frac{1}{1+\rho} \frac{\partial u_o^l}{\partial I^l} = \Pi \frac{\partial c_y}{\partial I^l} + (1+\lambda) \frac{\partial u_y}{\partial I^l} + \mu \left(\frac{\partial u_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I^l} \right) - \mu \left(\frac{\partial \widehat{u}_y}{\partial I^l} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} \right) - \mu \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial I^l} + \theta (1 + \pi^{ll})$$

and

$$-\frac{1}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} = -\Pi \left(1 - \frac{\partial c_y}{\partial B^l}\right) + (1+\lambda) \frac{\partial u_y}{\partial c_y} - \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) - \mu \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \mu \left(\frac{\partial u_y}{\partial c_y} + \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) - \theta (1 + \pi^{ll}).$$

Then, proceed as in the second part of the proof of Proposition 4 to derive an expression for $1 + \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l}$. The same kind of remarks provided there can be used here to show that there exists an implementing tax structure whose right-hand derivative at the bunching point I^l is equal to $1 + \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l}$.

7.11 Proof of Proposition 12

The definitions of the METRs faced by young and old low skilled agents in the two income points system are always those respectively given by (84) and (85). Eq. (61) and (62) are obtained substituting in (84) and (85) the values for $1 + \frac{\partial u_y}{\partial I^l} / \frac{\partial u_y}{\partial c_y}$ and $1 + \frac{\partial u_o^l}{\partial I^l} / \frac{\partial u_o^l}{\partial c_o^l}$ respectively given by (59) and (60) and then collecting terms.

7.12 Proof of Proposition 13

Take f.o.c. (66) and write it as

$$\frac{\partial c_y}{\partial B_o^l} \Pi = \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) - \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) + \theta \pi^{ll} - \frac{1}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l}. \quad (100)$$

In the three income points system and using the Slutsky-type decomposition f.o.c. (55) can be written as

$$\Pi \left[\frac{\partial \widetilde{c}_y}{\partial q} + s \left(\frac{\partial c_y}{\partial B_o^l} + \frac{\partial c_y}{\partial B_o^h} \right) \right] = s \left\{ \theta - \frac{1}{1+\rho} \left(\frac{\partial u_o^l}{\partial c_o^l} + \lambda \frac{\partial u_o^h}{\partial c_o^h} \right) \right\} + \mu s^m \frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \mu s^m \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} - \mu s \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial u_o^h}{\partial c_o^h} \right). \quad (101)$$

Summing up (100) and (94) and multiplying the result by s allows to get an expression for $\Pi s \left(\frac{\partial c_y}{\partial B_o^l} + \frac{\partial c_y}{\partial B_o^h} \right)$ which can be substituted in (101). Simplifying and collecting terms gives condition (67).

7.13 Proof of Proposition 14

Write f.o.c. (63) and (64) as

$$-(1 + \lambda) \frac{\partial u_y}{\partial I_y} = \Pi \frac{\partial c_y}{\partial I_y} - \mu \left(\frac{\partial \widehat{u}_y}{\partial I_y} - \frac{\partial u_y}{\partial I_y} \right) + \theta; \quad (102)$$

$$-(1 + \lambda) \frac{\partial u_y}{\partial c_y} = -\Pi \left(1 - \frac{\partial c_y}{\partial B_y} \right) - \theta - \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} - \frac{\partial u_y}{\partial c_y} \right). \quad (103)$$

Dividing (102) by (103) and multiplying the result by the r.h.s. of (103) gives:

$$\frac{\frac{\partial u_y}{\partial I_y}}{\frac{\partial u_y}{\partial c_y}} \left[-\Pi \left(1 - \frac{\partial c_y}{\partial B_y} \right) - \theta - \mu \left(\frac{\partial \widehat{u}_y}{\partial c_y} - \frac{\partial u_y}{\partial c_y} \right) \right] = \Pi \frac{\partial c_y}{\partial I_y} - \mu \left(\frac{\partial \widehat{u}_y}{\partial I_y} - \frac{\partial u_y}{\partial I_y} \right) + \theta.$$

Dividing both sides by θ , we can collect terms in order to get an expression for $1 + \frac{\partial u_y}{\partial I_y} / \frac{\partial u_y}{\partial c_y}$. By the implicit definition of the marginal labor income tax rate faced by young agents ($T_y' = 1 + \frac{\partial u_y}{\partial I_y} / \frac{\partial u_y}{\partial c_y}$) this expression will provide the value of T_y' given by (68).

The definition of the METR faced by the young agents in the three income points system is always the one given by (89). Eq. (69) is obtained substituting in (89) the value for $1 + \frac{\partial u_y}{\partial I_y} / \frac{\partial u_y}{\partial c_y}$ given by (68) and then collecting terms.

7.14 Proof of Proposition 15

Write f.o.c. (65) and (66) as

$$-\frac{1}{1+\rho} \frac{\partial u_o^l}{\partial I_o^l} = \Pi \frac{\partial c_y}{\partial I_o^l} + \theta \pi^{ll} - \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial I_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial I_o^l} \right) + \mu \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I_o^l}; \quad (100)$$

$$-\frac{1}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} = \Pi \frac{\partial c_y}{\partial B_o^l} - \theta \pi^{ll} - \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right). \quad (101)$$

Dividing (100) by (101) and multiplying the result by the r.h.s. of (101) gives:

$$\frac{\frac{\partial u_o^l}{\partial I_o^l}}{\frac{\partial u_o^l}{\partial c_o^l}} \left\{ \Pi \frac{\partial c_y}{\partial B_o^l} - \theta \pi^{ll} - \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial c_o^l} \right) + \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial c_o^l} \right) \right\} = \Pi \frac{\partial c_y}{\partial I_o^l} + \theta \pi^{ll} - \mu \left(\frac{\pi^{ll}}{1+\rho} \frac{\partial \widehat{u}}{\partial I_o^l} + \frac{\pi^{lh}}{1+\rho} \frac{\partial \widehat{u}}{\partial I_o^l} \right) + \mu \frac{\pi^{ll}}{1+\rho} \frac{\partial u_o^l}{\partial I_o^l}.$$

Dividing both sides by θ , we can collect terms in order to get an expression for $1 + \frac{\partial u_o^l}{\partial I_o^l} / \frac{\partial u_o^l}{\partial c_o^l}$. By the implicit definition of the marginal labor income tax rate faced by old low skilled agents ($T_o^l = 1 + \frac{\partial u_o^l}{\partial I_o^l} / \frac{\partial u_o^l}{\partial c_o^l}$) this expression will provide the value of T_o^l given by (70).

The definition of the METR faced by the old low skilled agents in the three income points system is always the one given by (88). Eq. (71) is obtained substituting in (88) the value for $1 + \frac{\partial u_o^l}{\partial I_o^l} / \frac{\partial u_o^l}{\partial c_o^l}$ given by (70) and then collecting terms.

References

- [1] Akerlof, G. A. (1978). "The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." *American Economic Review* 68, pp. 8-19.
- [2] Brett, C. (1998). "A Note on Nonlinear Taxation in an Overlapping Generations Model." Mimeo.
- [3] Cremer, H. and Gahvari, F. (1995). "Uncertainty, Optimal Taxation and the Direct versus Indirect Controversy." *Economic Journal* 105, pp. 1165-1179.
- [4] Edwards J., Keen, M. and Tuomala, M. (1994). "Income Tax, Commodity Taxes and Public Good Provision: A Brief Guide." *FinanzArchiv* 51, pp. 472-487.
- [5] Hahn, F. (1973). "On Optimum Taxation." *Journal of Economic Theory* 6, pp. 96-106.
- [6] Hammond, P. J. (1987). "Markets as Constraints: Multilateral Incentive Compatibility in Continuum Economies." *Review of Economic Studies* 54, pp. 399-412.
- [7] Immonen, R., Kanbur, R., Keen, M. and Tuomala, M. (1998). "Tagging and Taxing: The Use of Categorical and Income Information in Designing Tax/Transfer Schemes." *Economica* 65, pp. 179-192.
- [8] Kaplow, L. (1989). "Horizontal Equity: Measures in Search of a Principle." *National Tax Journal* 42, pp. 139-154.
- [9] Kaplow, L. (1995). "A Fundamental Objection to Tax Equity Norms: A Call for Utilitarianism." *National Tax Journal* 48, pp. 497-514.
- [10] Kaplow, L. (2000). "Horizontal Equity: New Measures, Unclear Principles." *NBER Working Paper* 7649.
- [11] Kaplow, L. and Shavell, S. (2000). "Notions of Fairness versus the Pareto Principle: On the Role of Logical Consistency." *Yale Law Journal* 110, pp. 237-249.
- [12] Kaplow, L. and Shavell, S. (2001). "Fairness versus Welfare." *Harvard Law Review* 114, pp. 966-1388.
- [13] Kremer, M. (2002). "Should Taxes Be Independent of Age?" Mimeo.
- [14] Lindencrona, G. (1993). "The Taxation of Financial Capital and the Prevention of Tax Avoidance." In *Tax Reforms in the Nordic Countries, 1973-1993 Jubilee Publication*. Nordic Council for Tax Research, Uppsala.

- [15] Mankiw, N. G. (1998). "Ax Taxes for Xers!" *Fortune*, March 16, 1998.
- [16] Mirrlees, J. A. (1971). "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38, pp. 175-208.
- [17] Mirrlees, J. A. (1976). "Optimal Tax Theory: A Synthesis." *Journal of Public Economics* 6, pp. 327-358.
- [18] Picard, P. M. (2001). "Job Additionality and Deadweight Spending in Perfectly Competitive Industries: The Case for Optimal Employment Subsidies." *Journal of Public Economics* 79, pp. 521-541.
- [19] Pirttilä, J. and Tuomala, M. (2001). "On Optimal Non-Linear Taxation and Public Good Provision in an Overlapping Generations Economy." *Journal of Public Economics* 79, pp. 485-501.
- [20] Stern, N. H. (1982). "Optimal Taxation with Errors in Administration." *Journal of Public Economics* 17, pp. 181-212.
- [21] Stiglitz, J. E. (1982). "Self-Selection and Pareto-Efficient Taxation." *Journal of Public Economics* 17, pp. 213-240.
- [22] Viard, A. D. (2001). "Optimal Categorical Transfer Payments: The Welfare Economics of Limited Lump-Sum Redistribution." *Journal of Public Economic Theory* 3, pp. 483-500.