UPPSALA UNIVERSITET

Uppsala Center for Fiscal Studies

Department of Economics

Working Paper 2010:12

The Welfare Gains of Age Related Optimal Income Taxation

Spencer Bastani, Sören Blomquist and Luca Micheletto

Uppsala Center for Fiscal Studies
Department of Economics
Uppsala University
P.O. Box 513
SE-751 20 Uppsala
Sweden
Fax: +46 18 471 14 78

Working paper 2010:12
October 2010

The Welfare Gains of Age Related Optimal Income Taxation

Spencer Bastani, Sören Blomquist and Luca Micheletto

The Welfare Gains of Age Related Optimal Income Taxation*

Spencer Bastani,
Uppsala Center for Fiscal Studies at the
Department of Economics, Uppsala Sweden

Sören Blomquist
Uppsala Center for Fiscal Studies at the
Department of Economics, Uppsala Sweden
CESifo, Germany

and

Luca Micheletto
Uppsala Center for Fiscal Studies,
Faculty of Law, University of Milan, Italy,
Econpubblica, Bocconi University, Italy and
CESifo, Germany

Abstract

Using a calibrated overlapping generations model we quantify the welfare gains of an age dependent income tax. Agents face uncertainty regarding future abilities and can by saving transfer consumption across periods. The welfare gain of switching from an age-independent to an age-dependent nonlinear tax amounts in our benchmark model to around three percent of GDP. The gains are particularly high when there are restrictions on debt policy. The gains of using a nonlinear- as opposed to a linear tax are even larger. Surprisingly, it is of secondary importance to optimally choose the tax on interest income.

Keywords: labor income taxation, capital income taxation, age-dependent taxes, OLG model

JEL Code: H21, H23, H24.

*We would like to thank Massimo Bordignon, Bas Jacobs and Tomas Sjögren for comments on an earlier draft of the paper.

1. Introduction

In a highly influential paper Akerlof (1978) demonstrated how redistribution can be achieved at lower efficiency costs if different tax schedules apply to different subgroups of the population; “tagging” schemes are always welfare-enhancing as long as the distributions of wage rates (abilities) differ across the subgroups. Over time this idea has gained considerable attention and presently there is large interest in tagging and optimal income taxation. The workings of such schemes have recently been studied by, for example, Immonen et al. (1998), Boadway and Pestieau (2006), Alesina et al. (2010), Mankiw and Weinzierl (2010) and Cremer et al. (2010).

In most countries it is easy to observe individuals’ age and this makes feasible to divide the population into age groups. Furthermore, the distribution of wage rates differs by age, the average wage being higher for older cohorts and the dispersion of wages wider. Hence, one particular form of tagging is to make the tax schedule dependent on age. In the Mirrlees Review (Banks and Diamond (2010)) this is mentioned as one very promising way to improve the tax system. In contrast to other tagging schemes suggested in the literature, such as tagging by gender or individuals’ height, tagging by age is probably politically feasible.

As a matter of fact, there are already elements of age-dependence in the tax and transfer- systems of many countries. For example, in Sweden the payroll tax recently became differentiated by age. However, these age dependencies are based on ad hoc reasoning and are not the result of a thorough analysis. There is therefore a need for a systematic analysis and a characterization of an optimal age dependent income tax. Because there are important aspects of tagging by age that make it different from other forms of tagging, the analyses of tagging schemes in earlier articles are not directly applicable to the case of tagging by age. Earlier analyses of tagging schemes have considered atemporal models where each individual belongs to only one group. The study of an age-dependent tax requires instead an intertemporal model; moreover, in contrast to other tagging schemes, a given individual will over time belong to different (age) groups. Finally, an individual can via savings redistribute income between himself at different ages.

Age-dependent taxes have been studied in a few earlier papers. Kremer (2002) is an interesting early contribution. However, he did not consider intertemporal aspects, hence abstracted from savings decisions. Erosa and Gervais (2002) and Lozachmeur (2006) studied age-dependent taxes in dynamic models with a representative agent. Their papers did not therefore address the issue of income redistribution. Blomquist and Micheletto (2003, 2008) provided the first analysis of an age-dependent tax in an intertemporal model with heterogeneous individuals who face a stochastic wage process and can save. Using an overlapping generations (OLG) model they characterized the optimal marginal labor income taxes and the optimal tax on savings. They also showed that a strict Pareto improvement can be obtained by moving from an age-independent to an age-dependent labor income tax. However, the quantitative importance of the welfare gains was not assessed.

In this paper we quantify the potential welfare gains of making the income tax age-dependent, as well as characterize the shape of an optimal age-dependent income tax.[1] Our ambition is that the results should be policy relevant, and for this purpose we construct a model capturing important features of the real economy. We also make observational assumptions that only require information of the type used by current tax systems, plus information on the individuals' age. This means that we rule out the possibility of levying taxes that are dependent on the income history of individuals. Although analyses of history-dependent taxes are of interest, we believe that such tax systems are far off into the future.[2]

The potential gains of age-dependent nonlinear income taxes as compared to age-independent ones are threefold. First, the self-selection constraints are less severe under an age-dependent tax as mimicking across different age groups is ruled out. This is the benefit analyzed in Blomquist and Micheletto (2008) and Weinzierl (2008). In both these papers it is assumed that the interest rate and the wage rates are fixed. This means that the capital accumulation process and the role of the capital stock for the interest- and wage rates are neglected. In this article we take account of the capital accumulation. With unrestricted public debt it is possible to reach the optimal (golden rule) level of capital accumulation both under an age-dependent and an age-independent tax. However, in reality there might be restrictions on the debt policy. The tax instruments could then be used to control the capital stock. For the age-dependent tax it is the case that the capital stock can be perfectly controlled without interfering with the self-selection constraints or the redistributional goals. Thus, even if debt policy cannot be used to control the capital stock, under an age-dependent tax we are still able (and we still want) to achieve the golden rule. This is the second advantage of an age-

[1] Weinzierl (2008) provided a first contribution where a quantitative assessment of the welfare gains is attempted. However, as we will clarify below, there are many distinctive features between our model and the model by Weinzierl. For instance, we explicitly model the production side of the economy, and this allows us to consider the interaction between age-dependent taxes and the capital accumulation process. Moreover, we consider age-dependency both in the context of linear- as well as nonlinear taxation schemes, and we also deal with the interaction between age-dependency and optimal capital income taxation.

[2] See, e.g., Golosov et al. (2006) for a study where taxes can be conditioned on the earnings history of agents.

dependent tax. Contrarily, for the age-independent tax there is a conflict between controlling the capital stock while at the same time guaranteeing incentive-compatibility of the allocations intended for the various types of agents. This implies that, in the presence of restrictions on debt policy, an optimal nonlinear age-independent tax will in general not yield the golden rule. The third advantage of an age-dependent tax is computational. As will be illustrated below, it is computationally much more difficult to find an optimum for the nonlinear age-independent tax problem due to the larger set of incentive-compatibility constraints. When computing the optimal nonlinear income tax for actual economies, approximations of the wage processes must be calculated. For the age-independent tax the approximations must, for computational reasons, be much cruder than for the age-dependent tax. Although this advantage of the age-dependent tax is hard to quantify, in practice it might be quite important.

As we have mentioned above, one of the advantages of a nonlinear age-dependent tax is the reduction in the number of self-selection constraints, since under an age-dependent tax an individual cannot choose an income point intended for an agent belonging to a different age group. To model this feature in a proper way, we need a model where at each point in time the economy is populated by individuals of different ages. For this reason we use an OLG model as a workhorse for our analysis. Moreover, to endogenize the interest- and wage rates, we explicitly model the production side of the economy.

Our OLG model has individuals living for three periods, two where they work and one where they are retired. The last period is needed to generate a reasonable pattern of savings and capital accumulation. We can think of each period in our model as corresponding to something like 20 calendar years.

The shape of the optimal tax schedules depends on the first and second period wage distributions, and the Markov probabilities that relate wages in the two working periods. We have chosen to calibrate our model to wage data for both the United States and Swedish economy. These are two countries that sometimes are regarded as two extremes in terms of wage dispersion. Our calculations focus on how the age-dependent taxes should be designed, as well as the gains obtained by moving from an age-independent tax to an age-dependent tax.

Since we want to compare an optimal age-independent nonlinear income tax with an optimal age-dependent nonlinear income tax, computational feasibility requires us to keep the number of skill types in each period low, as for instance three types in the first period and four in the second. Even with this restriction the age-independent tax schedule in particular is still difficult to compute as in conjunction with our rich economic environment it admits complex mimicking behaviors, giving rise to a very large number of incentive constraints. The computational difficulties are also the reason why we will not analyze the transition from an age-independent to an age-dependent tax, but focus instead on steady-states.

The paper is organized as follows. In section 2 we present the basic structure of our model. In section 3, in order to convey in the simplest way the importance of the information on individuals' age, we consider the First Best solution to a simplified version of our model. We show that the social planner achieve the First Best only if it can use the information on both individuals' age and productivities. In section 4 we first set up the government's problems under an optimal age-dependent and age-independent *nonlinear* income tax, and discuss how the number of self-selection constraints depends on the number of skill types in the economy. We then move on to describe the government's problems under an optimal age-independent and age-dependent *linear* income tax. In section 5 we describe how we calibrate our model. In section 6 we present the results of our simulations. Among our results, we find that a move from a nonlinear age-independent tax to a nonlinear age-dependent tax yields a welfare gain of about 3% of GDP. We also find that the gains of using a nonlinear tax instead of a linear tax are substantial (about twice the size of the welfare gains from age-dependency), whereas setting the interest income tax at the optimal rate is of second order importance in most cases. Finally, section 8 offers concluding remarks.

2. The model

We consider a discrete time OLG economy with heterogeneous agents living for three periods, working during the first two periods and being retired in the last period. Agents are indexed by their productivity and age, and start out their lives as one of m^y possible productivity types. We denote a young agents' productivity by θ_i, with $i \in \{1,...,m^y\}$, and assume that the proportion of young agents of type i is p_i, with $\sum_i p_i = 1$. With probability p_{ij} an agent who is of productivity i in the first period of his worklife (when young) will have productivity θ_{ij}, with $j \in \{1,...,m^o\}$, in the second period (when middle-aged), where $p_{ij} \geq 0$, $\forall (i, j) \in \{1,...,m^y\} \times \{1,...,m^o\}$ and $\sum_j p_{ij} = 1$. Introducing the notation $I = \{1,...,m^y\}$ and $J = \{1,...,m^o\}$, we interpret θ_σ as referring to the productivity of a middle aged agent if σ is a tuple $(i, j) \in I \times J$, and as referring to the productivity of a young agent if σ is a single index $i \in I$.

Labor supply and consumption of a young agent of type i born in period t is denoted by $\ell_{i,t}$ and $c_{i,t}$ respectively, whereas $\ell_{ij,t}$ and $c_{ij,t}$ denote the labor supply and consumption of a j-type middle aged agent (therefore born in $t-1$) who was of type i when young. In the third period of their lifetime, agents are retired and we denote by $c_{ij,t}^{R}$ the consumption in period t of a retiree who was of type i when young and of type j when middle aged. Population grows at rate n starting from an initial size N_0. Thus, the size of the cohort born at time t is $N_t = N_0(1+n)^t$. Labor and consumption in efficiency units are defined as $\tilde{L}_t = \sum_i p_i \theta_i \ell_{i,t} + \frac{1}{1+n} \sum_{ij} p_i p_{ij} \theta_{ij} \ell_{ij,t}$ and $\tilde{C}_t = \sum_i p_i c_{i,t} + \frac{1}{1+n} \sum_{ij} p_i p_{ij} \left(c_{ij,t} + \frac{1}{1+n} c_{ij,t}^{R} \right)$, whereas total effective labor and consumption in the economy is given by $L_t = N_0 (1+n)^t \tilde{L}_t$ and $C_t = N_0 \left(1+n\right)^t \tilde{C}_t$ respectively.

Firms operate a CRS technology $F(K_t, L_t)$ admitting zero equilibrium profits. The capital depreciation rate, the interest rate and the wage rate per efficiency unit of labor are denoted respectively by δ, r_t and ω_t. With perfectly competitive markets, factors earn their marginal products and we have $r_t = f'(k_t) - \delta$ and $\omega_t = f(k_t) - k_t f'(k_t)$, where we have used the relationship $F\left(K_t, L_t\right) = L_t f\left(k_t\right)$, with $k_t \equiv K_t / L_t$. An agent's labor income is defined as the product of labor in efficiency units and the equilibrium wage rate: $y_{i,t} = \omega_t \theta_i \ell_{i,t} \equiv w_{i,t} \ell_{i,t}$ and $y_{ij,t} = \omega_t \theta_{ij} \ell_{ij,t} \equiv w_{ij,t} \ell_{ij,t}$.

We will consider labor income tax systems of the form $\{T_1(y_i), T_2(y_{ij})\}$, where $T_1(\bullet)$ is the tax function that applies to young agents and $T_2(\bullet)$ the tax function for middle aged agents.[3] This formulation admits taxes depending both on an agent's income and age, but does not admit taxes which depend on the income history of an agent.[4] For future use we define after-tax incomes as $b_i = y_i - T_1(y_i)$ and $b_{ij} = y_{ij} - T_2(y_{ij})$. A key assumption that we make is that savings can only be observed anonymously. That is, savings can be observed, but it cannot be observed who is the beneficiary of the savings. This implies that the returns on

[3] We will maintain this notation throughout the paper keeping in mind that under an age-dependent tax we will have $T_1(\bullet) \neq T_2\left(\bullet\right)$, whereas under an age-independent tax we will have $T_1(\bullet) = T_2\left(\bullet\right)$.

[4] A history-dependent tax function would have the more general non-separable form $T(y_i, y_{ij})$.

savings can only be taxed at a proportional rate τ_K.[5] In each period individual preferences are represented by a twice-differentiable, strictly quasi-concave utility function $u(c_t, \ell_t)$ where $u_c > 0$ and $u_\ell < 0$. Hence, the expected lifetime utility of an agent with productivity i when young is given by:

$$V_{i,t} = u(c_{i,t}, \ell_{i,t}) + \sum_j p_{ij} \left[\beta u(c_{ij,t+1}, \ell_{ij,t+1}) + \beta^2 u(c^R_{ij,t+2}) \right] = u(b_{i,t} - s_{i,t}, y_{i,t} / w_{i,t}) +$$

$$\sum_j p_{ij} \left[\beta u(b_{ij,t+1} + s_{i,t}\left(1 + r\left(1 - \tau_{K,t+1}\right)\right) - s_{ij,t+1}, y_{ij,t+1} / w_{ij,t+1}) + \beta^2 u\left(s_{ij,t+1}\left(1 + r\left(1 - \tau_{K,t+2}\right)\right)\right)\right].$$

For given values of $b_{i,t}, y_{i,t}, b_{ij,t+1}$ and $y_{ij,t+1}$, an individual chooses first- and second period savings $s_{i,t}$ and $s_{ij,t+1}$ maximizing the above expression. It should be noted that when a young individual plans his second period savings, he will plan to save different amounts depending on what productivity type he will turn out to be when middle aged.

We define aggregate savings in the following way. The aggregate savings of generation t is given by $N_t s_t = N_t \sum_i p_i s_{i,t}$ and the aggregate savings in period t of the generation born in period t-1 is given by $N_{t-1} s_{t-1} = N_{t-1} \sum_{ij} p_i p_{ij} s_{ij,t}$. Agents invest either in government bonds D_t or in physical capital I_t. Equilibrium in the capital market requires:

$$D_t + I_t = N_t s_t + N_{t-1} s_{t-1}. \tag{1}$$

The capital stock in period $t+1$, net of what is left after depreciation of the capital stock in period t, is equal to that part of investment that goes into physical capital in period t:

$$K_{t+1} - (1-\delta) K_t \equiv I_t = N_t s_t + N_{t-1} s_{t-1} - D_t. \tag{2}$$

Dividing (2) by $L_{t+1} = N_{t+1} \tilde{L}_{t+1}$ gives:

$$k_{t+1} - \left(1-\delta\right) \frac{k_t}{1+n} = \frac{s_t}{(1+n)\tilde{L}_{t+1}} + \frac{s_{t-1}}{(1+n)^2 \tilde{L}_{t+1}} - \frac{D_t}{N_{t+1}\tilde{L}_{t+1}},$$

and simplifying we get:

$$\tilde{L}_{t+1}(1+n) k_{t+1} - \left(1-\delta\right) k_t = s_t + \frac{s_{t-1}}{1+n} - d_{t+1}(1+n),$$

where we have defined:

[5] If one tried to tax savings through a nonlinear function, there would be large incentives for someone with a high marginal tax on savings to ask a friend, with a lower marginal tax, to save for him. Our assumption here parallels the one usually made about purchases of commodities: anonymous transactions can be observed and taxed by a proportional tax, but personal consumption levels are not publicly observable. See Hammond (1987) on the desirability of linear pricing when exchanges on side market are not observable by the government.

$$d_{t+1} \equiv D_t / N_{t+1}.$$

Defining debt in this way emphasizes that debt decided upon at t is at time $t+1$ held by the larger population N_{t+1}. As in Diamond (1965), we will restrict attention to debt which is constant per young worker. Thus, $d_t = d$, $\forall t$.

In period 0 the initial capital stock K_0 and the labor supplied by the young and middle aged agents are combined to produce output. The law of motion for capital is:

$$K_{t+1} = (1-\delta)K_t + F(K_t, L_t) - C_t. \tag{3}$$

Dividing (3) by L_t gives:

$$k_{t+1}\frac{N_{t+1}\tilde{L}_{t+1}}{N_t\tilde{L}_t} = (1-\delta)k_t + f\left(k_t\right) - \frac{C_t}{L_t}, \tag{4}$$

which can be simplified to:

$$(1+n)k_{t+1}\frac{\tilde{L}_{t+1}}{\tilde{L}_t} = (1-\delta)k_t + f\left(k_t\right) - \frac{C_t}{L_t}. \tag{5}$$

Combining the resource constraint (5), the private budget constraints relating before- to after-tax income, the definitions of C_t and L_t, and the capital market equilibrium condition, one can derive the following government's budget constraint:

$$\sum_i p_i T_{1,t}(w_{i,t}\ell_{i,t}) + \frac{1}{1+n}\left(\sum_{ij} p_i p_{ij} T_{2,t}(w_{ij,t}\ell_{ij,t})\right) + \frac{\tau_{K,t} r_t}{1+n}\left(\sum_i p_i s_{i,t-1} + \frac{\sum_{ij} p_i p_{ij} s_{ij,t-2}}{1+n}\right) + (1+n)d_{t+1} = (1+r_t)d_t.$$

We are now in a position to give the following definition:

Definition 1 *(Competitive Equilibrium)*

Given a tax policy $\{T_{1,t}(y_i), T_{2,t}(y_{ij}), \tau_{K,t}\}_{t=0}^{\infty}$, *a competitive equilibrium is a sequence of prices* $\{r_t, \omega_t\}_{t=0}^{\infty}$, *individual allocations* $\{\{c_{i,t}, c_{ij,t}, \ell_{i,t}, \ell_{ij,t}, s_{i,t}, s_{ij,t}\}_{t=0}^{\infty}\}_{(i,j)\in I\times J}$, *production plans* $\{K_t, L_t\}_{t=0}^{\infty}$ *and government debt* $\{D_t\}_{t=0}^{\infty}$ *such that individual allocations solve the individual maximization problems, factors are paid their marginal product, the production (feasibility) constraint is satisfied and factor markets clear.*

Definition 2 *(Stationary Equilibrium)*

A stationary equilibrium is such that, for all periods $t \geq t_{ss}$, $\{r_t, \omega_t\} = \{r, \omega\}$, $\{T_{1,t}(y_i), T_{2,t}(y_{ij}), \tau_{K,t}\} = \{T_1(y_i), T_2(y_{ij}), \tau_K\}$, $\{d_t\} = d$, $\{\tilde{C}_t, \tilde{L}_t\} = \{\tilde{C}, \tilde{L}\}$ *and* $\{k_t\} = \{k\}$.

Hence in a stationary equilibrium, prices are constant, tax and debt policy is constant, and all per capita quantities are constant. In this paper we will mainly focus on one particular steady state where the capital labor ratio k satisfies the golden rule $f'(k) = n + \delta$. This equilibrium arises when the government objective is the maximization of steady state welfare. We will assume such a stationary equilibrium exists. It will be achievable as long as the government can control the capital stock, either through sophisticated tax instruments or through (unrestricted) debt policy. Under this assumption, we can derive the following steady state analogues to the previously derived relationships:[6]

(i) $\tilde{L}(1+n)k - \left(1-\delta\right)k = s + \dfrac{s_{-1}}{1+n} - d(1+n);$

(ii) $(1+n)k = (1-\delta)k + f\left(k\right) - \dfrac{C}{L};$

(iii) $\displaystyle\sum_i p_i T_1(w_i \ell_i) + \frac{1}{1+n}\left(\sum_{ij} p_i p_{ij} T_2(w_{ij}\ell_{ij})\right) + \frac{\tau_K r}{1+n}\left(\sum_i p_i s_i + \frac{\sum_{ij} p_i p_{ij} s_{ij}}{1+n}\right) = 0.$

Before considering the optimal taxation problem solved by a government in a second-best setting where ability is a private information of agents, in the next section we characterize the first best solution to a simplified version of our OLG model. This will allow us to highlight that, in order to implement the first best, the information on individual's abilities is not enough. Only if the information on agents' age is used too, the first best can be achieved.

3. First Best

We consider a simplified economy where in the first period of their lives everyone has the same productivity θ_1. In the second period an agent remains at productivity $\theta_{11} = \theta_1$ with

[6] By the definition of the steady state, debt per worker is constant; hence debt grows at rate n but, since $r = n$, this exactly equals the interest amount which needs to be paid on debt from the previous period. Thus, debt does not appear in the government budget constraint in the golden rule steady state.

probability p_{11} and with probability p_{12} his productivity is $\theta_{12} > \theta_{11}$. We also let $\delta = 0$, abstract from the retirement period, and assume that the policy maker can control the capital K_t in period t. As our objective function we take the expected utility of a young agent:

$$U(c_1, \ell_1) + \beta\left[p_{11}U(c_{11}, \ell_{11}) + p_{12}U(c_{12}, \ell_{12})\right],$$

where β is the individuals' discount factor.

Note that $MRS_{c_1,c_{11}} = \left(\frac{dc_1}{dc_{11}}\right) = -\beta p_{11}\frac{U_{c_{11}}}{U_{c_1}}$ and $MRS_{c_{12},c_{11}} = \left(\frac{dc_{12}}{dc_{11}}\right) = -\frac{p_{11}}{p_{12}}\frac{U_{c_{11}}}{U_{c_{12}}}$. Also, $MRT_{c_1,c_{11}} = \left(\frac{dc_1}{dc_{11}}\right)_{prod} = -\frac{p_{11}}{1+n}$ and $MRT_{c_{12},c_{11}} = \left(\frac{dc_{12}}{dc_{11}}\right)_{prod} = -\frac{p_{11}}{p_{22}}$,

where for any variable x we have denote by U_x the derivative of U with respect to x.

Under the assumption that both productivity and age are observable by the planner, the consumption and labor supply can be freely set at different levels for young low skilled and old low skilled. Under this informational assumption, the Lagrangian for the First Best optimization problem can be written as:

$$\Lambda = U(c_1, \ell_1) + \frac{1}{R}\left[p_{11}U(c_{11}, \ell_{11}) + p_{12}U(c_{12}, \ell_{12})\right] + \mu\left[k + f\left(k\right) - (1+n)k - \tilde{C}/\tilde{L}\right]. \quad (6)$$

The first order condition with respect to k is:

$$\partial\Lambda / \partial k = \mu\left[1 + f\,'(k) - (1+n)\right] = 0, \quad (7)$$

which yields the familiar golden rule result $f\,'(k) = n$. Interpreting $f\,'(k)$ within a market framework we would have $r = n$.

For the purpose of illustrating the importance of the information on age, it is here sufficient to consider the first order conditions for c_1, c_{11} and c_{12}:

$$\frac{\partial\Lambda}{\partial c_1} = U_{c_1} - \frac{\mu}{\tilde{L}} = 0; \quad (8)$$

$$\frac{\partial\Lambda}{\partial c_{11}} = \beta p_{11}U_{c_{11}} - \mu\frac{p_{11}}{\tilde{L}(1+n)} = 0; \quad (9)$$

$$\frac{\partial\Lambda}{\partial c_{12}} = \beta p_{12}U_{c_{12}} - \mu\frac{p_{12}}{\tilde{L}(1+n)} = 0. \quad (10)$$

Combining (9) and (10) we obtain $U_{c_{11}} / U_{c_{12}} = 1$: the marginal utility of consumption should be equalized across all old agents. It also implies that $MRS_{c_{12},c_{11}} = MRT_{c_{12},c_{11}}$.

Equations. (8)-(10) also imply

$$\frac{U_{c_{12}}}{U_{c_1}} = \frac{U_{c_{11}}}{U_{c_1}} = 1/\beta(1+n), \tag{11}$$

and therefore $MRS_{c_1,c_{11}} = MRT_{c_1,c_{11}}$. Considering the first order conditions for hours of work we can also show that the hours of work are undistorted. This is done in Appendix A. However, if the policy maker cannot observe individuals' age, or is prevented from using this information, the consumption and labor supply bundle offered to young agents cannot be freely differentiated from the one offered to old low-skilled agents. In Appendix A we show that if $c_1 = c_{11}$ and $\ell_1 = \ell_{11}$ the no-distortion results cannot be obtained. Thus, to achieve the undistorted First Best information on age is necessary.

We are now ready to get back to our more general OLG model and consider age-dependency in various second-best settings.

4. Income taxation: theory

4.1 Age dependent optimal nonlinear income tax

In this section we assume that the government knows the age of individuals, the skill distribution at each age, and the Markov probabilities relating these distributions. We consider a discrete adaptation of the Mirrlees (1971) optimum income taxation model. The government can observe pre-tax income $y = w\theta\ell$ and w but neither θ nor ℓ. To pursue its redistributive goals the government has at its disposal a nonlinear labor income tax schedule at each age, $T_1(y_i)$ and $T_2(y_{ij})$, and a proportional tax on capital income, τ_K. Agents maximize expected utility based on the link between pre-tax earnings and post-tax earnings implied by the tax schedules. Using the notation introduced on page 5, the government's problem can equivalently be stated as choosing the allocations $\{b_i, y_i\}_{i\in I}$ and $\{b_{ij}, y_{ij}\}_{(i,j)\in I\times J}$ subject to a set of self-selection constraints and a public budget constraint. The self-selection constraints require that each agent (weakly) prefers the bundles intended for him to those that are intended for some other agent. This is a necessary condition on the allocation for it to be implementable by tax schedules that are common to all agents. Rather than choosing a single income point as in a static Mirrlees problem, in the dynamic problem we think of agents choosing a strategy. A strategy specifies which income point an agent chooses in each period of work and each state of the world (e.g. for each skill realization in the second period of life). Agents' strategies are independent of each other and there is no aggregate uncertainty. Each strategy also implies (unique) savings decisions consistent with the chosen income points and the agent's first order conditions for savings.

Formally, a strategy corresponding to agent i is a plan $\sigma^i = (\sigma_1^i, \sigma_2^i)$ where $\sigma_1^i \in I$ is the reported type when young, and σ_2^i is a *function* $\sigma_2^i : J \to I \times J$ specifying the income point chosen when middle aged for each second period skill realization $j \in J$. The set of all strategies available to agent i is denoted Γ_i. Truth-telling implies that a young agent of ability type i chooses the income point (b_i, y_i) in the first period and the income point (b_{ij}, y_{ij}) in the second period if j is his skill realization when middle aged. We denote the truthful strategy by $\tilde{\sigma}^i = (\tilde{\sigma}_1^i, \tilde{\sigma}_2^i)$ with $\tilde{\sigma}_1^i = i$ and $\tilde{\sigma}_2^i(j) = (i, j)$, $\forall j \in J$. An agent choosing any other strategy is called a *mimicker* or *deviating* agent. An allocation is said to be incentive-compatible or satisfying the self-selection constraints if the agent (weakly) prefers to adopt his truthful strategy rather than any deviating strategy. It is clear that there are many possible deviations from the truthful strategy as a truthful agent must not only report his true ability in period 1 but also report truthfully his skill realization in period 2. The possibility to save/borrow doesn't affect the number of incentive constraints but does enable agents to equalize, both for truthful- and mimicking strategies, the expected marginal utility of consumption when middle aged to the marginal utility of consumption when young. This makes deviating strategies more attractive as compared to a situation where agents cannot save, since in such a case the intertemporal allocation would be completely determined by the income points chosen when young and middle aged. Hence, with anonymous savings, income in any period is a weaker signal of ability and less redistribution can be achieved.

As we have assumed a flat tax τ_K on capital income, the intertemporal distortion will be the same both for truthful- and mimicking agents. A positive capital income tax might be desirable when mimickers tend to have a valuation of future consumption that is higher than that of non-deviating agents.

It is clear that the set of truthful strategies is just a small subset of the set of possible strategies $\Gamma = \bigcup_{i \in I} \Gamma_i$. Assume for instance that each agent has the same *number* of different possible skill realizations in the second period and denote this number by φ. With m^y possible ability types when young, the number of bundles offered by the government on the income tax schedule for the young is m^y. On the income tax schedule for the old, the efficient number of bundles offered by the government is $\Psi \equiv \sum_{i=1}^{m^y} \varphi = m^y \varphi$. Hence, the total number of strategies, including the strategies entailing truthful revelation, is given by:[7]

$$|\Gamma| = m^y \left[m^y \Psi^{\varphi} \right] = \left(m^y \right)^2 \Psi^{\varphi}.$$

To illustrate the formula above, suppose that $m^y = 2$ and that each type of young agents can be one of three possible types in the second period, meaning that $\varphi = 3$. This implies that the optimal number of bundles offered to the middle aged agents is $\Psi = 6$. In this case $|\Gamma| = 864$ and only two of these strategies entail truthful revelation (one for each type in the first period). This calculation forms the basis for setting up the set of self-selection constraints in the age-dependent (hereafter, AD) nonlinear income taxation problem.[8] Truthful revelation necessitates that choosing any of the 862 “mimicking” strategies should yield lower utility for the agent than choosing the truthful strategy associated with his type.

We are now in a position to formally state the problem of the government. As an objective function we choose the maximin social welfare function, given that it will be the one that we will adopt later on in our simulations. The government’s problem is therefore:

---

[7] Proof: We have defined $\sigma^i = (\sigma_1^i, \sigma_2^i)$ with $\sigma_1^i \in I$ and $\sigma_2^i : J \to I \times J$. Note that the number of functions from J to $I \times J$ is $|I \times J|^{|J|} = \Psi^{\varphi}$ by the rule of cardinal exponentiation. Hence the number of strategies for each agent i is given by $|\{\sigma^i\}| = |\{\sigma_1^i\} \times \{\sigma_2^i\}| = |\{\sigma_1^i\}||\{\sigma_2^i\}| = m^y |\{\sigma_2^i\}| = m^y \Psi^{\varphi}$ and summing over all types when young (which are m^y in number) the formula is obtained.

[8] For the general case when each ability type when young has a specific number of possible skill realizations when middle aged, the formula for calculating the total number of strategies is the following. Suppose that there is a function $\varphi(i)$ that for every ability type $i \in I$ when young gives the corresponding number of possible ability types when middle aged. For instance, $\varphi(i) = j$ would mean that, if an agent is of ability type i when young, his ability type when middle aged can take j different realizations. With m^y possible ability types when young, the number of bundles offered by the government on the income tax schedule for the young is m^y. On the income tax schedule for the middle aged, the efficient number of bundles offered by the government is $\Psi \equiv \sum_{i=1}^{m^y} \varphi(i)$. Hence, the total number of strategies, including the strategies entailing truthful revelation, is given by $|\Gamma| = m^y \left[\sum_{i=1}^{m^y} \Psi^{\varphi(i)} \right]$.

$$\max_{d,\{b_i,y_i\}_{i\in I},\{b_{ij},y_{ij}\}_{i\in I,j\in J},\tau_K}\left(\min_i V_i(\tilde{\sigma}^i)\right)$$

subject to

$$\forall\ i\in I\ :\ V_i(\tilde{\sigma}^i)\geq V_i(\sigma^i),\forall\sigma^i\in\Gamma_i \qquad \text{(IC1)}$$

$$\sum_i p_i(y_i-b_i)+\frac{1}{1+n}\sum_{ij}p_i p_{ij}(y_{ij}-b_{ij})+\frac{\tau_K r}{1+n}\left(\sum_i p_i s_i+\frac{\sum_{ij}p_i p_{ij}s_{ij}}{1+n}\right)=0 \quad \text{(Public Budget)}$$

$$\tilde{L}=\left(\sum_i p_i(y_i/\omega)+\frac{1}{1+n}\sum_{ij}p_i p_{ij}(y_{ij}/\omega)\right) \qquad \text{(Labor Market)}$$

$$\tilde{L}(1+n)k-(1-\delta)k=s+\frac{s_{-1}}{1+n}-d(1+n)\,, \qquad \text{(Capital Market)}$$

where:

$$V_i(\sigma^i)=u\left(b_{\sigma_1^i}-s_{\sigma^i},y_{\sigma_1^i}/w_i\right)+\beta\sum_j p_{ij}\begin{Bmatrix}u\left(b_{\sigma_2^i(j)}+(1+r(1-\tau_K))s_{\sigma^i}-s_{\sigma^i}(j),y_{\sigma_2^i(j)}/w_{ij}\right)+\\ \beta^2u((1+r(1-\tau_K))s_{\sigma^i}(j))\end{Bmatrix}$$

and

$$V_i(\tilde{\sigma}^i)=u\left(b_i-s_i,y_i/w_i\right)+\beta\sum_j p_{ij}\begin{Bmatrix}u\left(b_{ij}+(1+r(1-\tau_K))s_i-s_i(j),y_{ij}/w_{ij}\right)+\\ \beta^2u((1+r(1-\tau_K))s_i(j))\end{Bmatrix},$$

with s_{σ^i} denoting the savings chosen in the first period by a type i agent who adopts the strategy σ^i and $s_{\sigma^i}(j)$ denoting the savings prescribed in the second period by this strategy if he turns out being of skill j when middle aged (with $s_i(j)$ denoting the savings undertaken in the second period under a truthful reporting strategy). The (IC1) set of constraints ensures that every agent $i\in I$ prefers the truthful strategy $\tilde{\sigma}^i$ over any other available strategy.

It is of interest to note that the optimal nonlinear AD tax achieves the golden rule capital-labor ratio even if debt policy is unavailable. We show this formally in Appendix B. This means that there is an indeterminacy in the above optimization problem in the sense that there is an infinite number of combinations of taxes and debt that all yield the same global optimum. The intuition for this result is that under an AD nonlinear tax the savings of the young does not matter for self-selection purposes. More precisely, under an AD nonlinear tax one can marginally change the after-tax labor incomes of the young, and at the same time adjust the after-tax labor incomes of the middle aged, in such a way that the lifetime tax payment of all agents is left unaffected, all young agents change their savings by the same amount, the public budget is still balanced and all the self-selection constraints continue to be satisfied if they were satisfied before the implementation of the reform. In this sense one can claim that the absolute level of private savings does not matter for self-selection purposes; therefore, savings can be controlled (by a proper choice of the labor income taxes when young and when middle aged) with the sole purpose of achieving the golden rule capital stock.

4.2 Age independent optimal nonlinear income tax

Under an age-independent (hereafter, AID) nonlinear income tax the government's problem changes only because the set of incentive constraints is larger. It is still the case that the number of income points on the schedule for the young is m^y and the one on the schedule for the middle aged is Ψ. However, under a nonlinear AID tax it is possible for an agent to choose an income point intended for an agent of a different age. We write $\sigma^i = (\sigma_1^i, \sigma_2^i)$ where $\sigma_1^i \in (I \times J \cup I)$ is the reported type when young, and σ_2^i is a *function* $\sigma_2^i : J \to (I \times J \cup I)$ specifying the income point chosen when middle aged for each $j \in J$. The set of all strategies available to a young agent of type i is denoted by $\hat{\Gamma}_i$ and the set of incentive constraints can be written as:

$$\forall \ i \in I \ : V_i(\tilde{\sigma}^i) \geq V_i(\sigma^i), \forall \sigma^i \in \hat{\Gamma}_i.$$

The total number of possible strategies in the AID system is:[9]

$$\left|\hat{\Gamma}\right| = (m^y + \Psi)\left[\sum_{i=1}^{m^y} (m^y + \Psi)^{\varphi}\right].$$

This rule can be derived as follows. Consider an agent with a specific skill type when young. In the first period he can choose among $(m^y + \Psi)$ different income points, and this explains the first factor in the expression above. For any given choice in the first period, a strategy must specify a point to be chosen *for each possible skill realization in the second period*; with φ possible skill realizations there are $(\Psi + m^y)^{\varphi}$ ways to choose an income point in the second period. Taking into account that this reasoning applies to each of the m^y possible types of young agents, the formula above is obtained.

[9] The formal proof is analogous to the proof given in footnote 7 for the AD case, taking into account that the number of points to choose from when young and middle aged is now $\Psi + m^y$. For the general case when each ability type when young has a specific number of possible skill realizations when middle aged (meaning that φ is type-specific), we would get $\left|\hat{\Gamma}\right| = (m^y + \Psi)\left[\sum_{i=1}^{m^y} (m^y + \Psi)^{\varphi(i)}\right]$.

In contrast to the nonlinear AD tax, the nonlinear AID tax does in general not achieve the golden rule if debt policy is not available. This is shown in Appendix C. In the absence of debt policy, or when there are restrictions on debt policy, a conflict arises between controlling the capital stock and guaranteeing incentive-compatibility of the allocations intended for the various types of agents. The reason is that, under a nonlinear AID tax, it is not possible to change the after-tax labor incomes of the young, and at the same time adjust the after-tax labor incomes of the middle aged, in such a way that the lifetime tax payment of all agents is left unaffected, all young agents change their savings by the same amount, the public budget is balanced and all the self-selection constraints continue to be satisfied. In this sense we can claim that, under a nonlinear AID tax, the absolute level of savings matters for self-selection purposes (or, put differently, the level of individual savings becomes relevant in order to deter mimicking behaviors). Therefore, we can no longer select the after-tax labor incomes of the young and the middle aged with the sole purpose of attaining the golden rule. Hence, for the nonlinear AID tax, only if unrestricted debt policy can be used can the capital stock be controlled independently of the self-selection constraints.

4.3 Linear income tax: age dependent and age independent

An interesting question is how the gains from age dependency depend on the social planner's ability to fully optimize a nonlinear tax schedule; moreover, it is also of interest to compare the welfare gains descending from age-dependency with the welfare gains that can be obtained by moving from a linear labor income tax to a nonlinear tax. For this purposes, in this section we restrict $T_1(\bullet)$ and $T_2(\bullet)$ to be affine functions, so that labor income is taxed at the flat rates τ_1 and τ_2 in the first- and second-period respectively. In addition, there is a demogrant G_1 which is paid to all young workers and a demogrant G_2 which is paid to all middle aged workers. In the AID scenario we impose $\tau_1 = \tau_2$ and $G_1 = G_2$, whereas no such restriction is made in the AD setting. Both in the AID and the AD linear tax systems we restrict the demogrants to be positive ($G_i \geq 0$ i=*1,2*). The reason for imposing such a restriction is a concern for realism: a uniform lump-sum tax is very unlikely to be politically implementable.[10] For later purposes it is worth noticing that age-differentiated demogrants are redundant when the government has access to an unrestricted debt policy. In that case, the restriction that $G_1 = G_2$ is harmless from the social welfare perspective.

[10] The attempt made by Margaret Thatcher to introduce a poll tax in UK can be regarded as an example.

The tax rate on capital income is always denoted by τ_K. Combining the resource constraint and the private budget constraints, together with the definitions of C_t and L_t, and the capital market equilibrium condition, one can derive the following government's budget constraint:

$$\sum_i p_i(\tau_{1,t} w_{i,t}\ell_{i,t} + \frac{\tau_{K,t} r_t s_{i,t-1}}{1+n}) + \frac{1}{1+n}\left(\sum_{ij} p_i p_{ij}\left(\tau_{2,t} w_{ij,t}\ell_{ij,t} + \frac{\tau_{K,t} r_t s_{ij,t-2}}{1+n}\right)\right) + (1+n)d_{t+1} = (1+r_t)d_t +$$
$$+G_{1,t} + \frac{1}{1+n}G_{2,t}.$$

The steady state private budget constraints are in this case:

$$c_i \equiv w_i \ell_i (1-\tau_1) - s_i + G_1;$$
$$c_{ij} \equiv w_{ij}\ell_{ij}(1-\tau_2) + s_i(1+r(1-\tau_K)) + G_2 - s_{ij};$$
$$c_{ij}^R \equiv s_{ij}(1+r(1-\tau_K)).$$

5. Calibration and computational approach

5.1 Parameterization

In the benchmark parameterization each period corresponds to 20 years. We use a parameterization similar to the one employed in Conesa et al. (2009). Annual depreciation is set to 8%, population growth to 1.1%. We then calculate the 20 year analogues of these numbers which yields $n = (1.011)^{20} - 1$ and $\delta = 1 - 0.92^{20}$. We also assume $\beta = 0.988^{20}$. Production is Cobb-Douglas and the share of capital in production is $\alpha = 1/3$. The production scale factor A is chosen so that the equilibrium rental price for one efficiency unit of labor is equal to one. Agents maximize their expected lifetime utility given an instantaneous utility function defined as:

$$u(c,\ell) = \frac{c^{1-\gamma}}{1-\gamma} - \frac{\ell^{\kappa}}{\kappa}.$$

We choose $\kappa = 2.25$ (implying a Frisch elasticity of 0.8) and γ=0.9 as benchmark values. For sensitivity analysis we also consider $\gamma = 0.7$ and $\gamma = 1.3$ as well as $\kappa = 4$ (implying a Frisch elasticity of 0.33) and $\kappa = 2$ (implying a Frisch elasticity of 1).[11] The government's exogenous revenue requirement is set to zero.

[11] There is substantial empirical uncertainty regarding the Frisch elasticity. Microeconometric evidence suggests a low value of around 0.1. However, other estimates (see for instance Imai and Keane (2004) and Keane (2009)) are as high as 4.

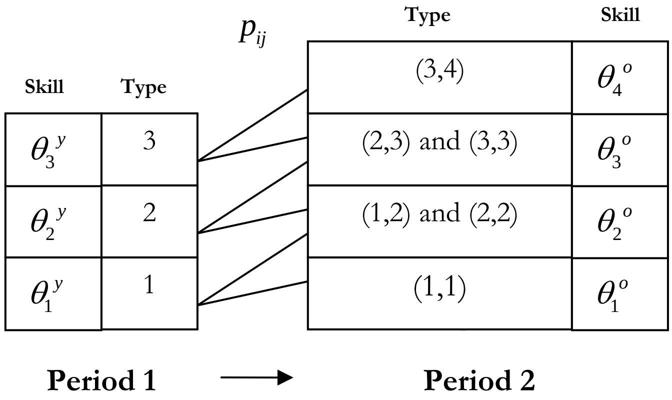
5.2 Wage Process

In static optimal nonlinear tax analysis it is well known that the optimal tax schedule depends on the distribution of skills in the economy. In a dynamic setting the equivalent of the skill distribution is the skill process. The skill process is a more complex object which consists of three parts, the skill distributions in the first- and second periods, and the transition probabilities linking these distributions. It is worth noting that there are various combinations of transition probabilities and wage levels which result in the same expected lifetime income paths for all agents, but these different skill processes might yield different welfare gains when studying policy reforms. The welfare gains and the shape of an optimal AD tax will crucially depend on the skill process. There are primarily two features of the wage process which are important. The first is the degree of overlap of the wage distributions for the young and for the middle aged. The second is the persistence of the wage process, which captures how likely it is for an agent classified as low skilled in the first period to remain low skilled also in the second period. Both these aspects of the wage process are likely to differ between economies and therefore the welfare gains of AD taxes will be different for different countries. In this paper we calibrate our model economy to the skill processes of two actual economies, the US and Sweden. These economies are often regarded as two extremes in terms of wage dispersion and wage mobility. In addition to being informative regarding the welfare gains achievable in real economies, the two calibrations tell us something about the sensitivity of our results to the choice of the skill process.

Since the number of self-selections constraints increases quickly in the number of types, especially for the nonlinear AID tax, for computational reasons we can only have a small number of types. We have chosen to allow for three skills in the first period and four in the second. To keep the model manageable we also restrict some of the transition probabilities by assuming $p_{ij} = 0$ for $|i - j| > 1$. The dynamics of the skill process is illustrated in Figure 1 below.

Besides assessing the welfare gain of age-dependency, our simulations will also shed light on the welfare gains of using an optimal nonlinear tax instead of a linear one. Given the many suggestions for a “flat tax” this is also of large interest. We have performed experiments in a simple static model and found that the welfare gain of a nonlinear tax upon a linear tax is, up to a certain point, monotonically increasing in the number of types used to represent the skill distribution. This is likely to be due to the fact that the quality of a linear approximation of a nonlinear function diminishes when the number of data points increase (e.g. a line perfectly approximates two points but not three). Given that computational considerations restrict us to consider few types, our results should therefore be interpreted as a lower bound for the power of the nonlinear tax schedules.[12]

Figure 1 (Skill Process)



5.2.1. The United States

In order to estimate the transition probabilities we have used the National Longitudinal Study of Youth (NLSY79) published by the US Bureau of Labor Statistics. Agents who were 25 years old during the period 1982-1988 were sorted into different wage bins and this was done also for these agents 20 years later in order to construct wage bins for the young and old.[13] By tracing the movements among the different wage bins, the transition probabilities are estimated; the Markov probabilities are given as the relative frequencies of each of the possible wage paths represented by movements in and out of the wage bins. By construction, the (ex-ante) probability of being any one of the three possible types when young is 1/3. [14]

The wage bins for young and middle aged are constructed as follows. Young agents are divided into three groups separated by the 33rd and 66th percentile; middle aged agents are divided into four groups separated by the 25th, 50th and 75th percentile. The wage of each group is represented by the mean wage in that group. This enables us to classify individuals as having one out of six possible wage paths.

[12] We conjecture that the same is true for the welfare gain of age dependence which is a *comparison* between two nonlinear tax schedules. The reason is that our welfare gain of age dependence is defined as the difference between the welfare gain of a nonlinear AD tax in comparison with a linear tax benchmark and the welfare gain of a nonlinear AID tax in comparison with the same linear benchmark. If both these terms are magnified by the same factor in a less restrictive model, the welfare gains of age dependence in our model with a limited number of types should be regarded as a lower bound of the welfare gains achievable in a richer environment.

[13] The reason we have not focused on a single year is to smooth out potential year and cohort-specific shocks to individual wages.

[14] The final sample size is 6981. Individuals with an hourly wage rate of less than SEK 10 or greater than SEK 300 (in 1976 prices) were dropped from the sample.

Although wage data is available in the NLSY79, to get a more accurate picture of the actual wage distribution in the US and for the purpose of obtaining the wage rates only, we have chosen to use extracts from the more representative Current Population Survey (CPS). We have constructed hourly wage rates for individuals who were 25 and 45 year old during the years 2004-2008. Focusing on males who were not self-employed and who earned more than an approximate federal minimum wage during the period, a measure of the wage rate was obtained by dividing earnings by hours worked. The wages in the CPS are top-coded (wages are truncated so that all earners above a threshold are assigned this threshold as their wage rate). We have dealt with this by assigning all agents subject to top-coding to a wage rate equal to the mean of the observations above the top-code. The mean is not known and has to be estimated. We have first estimated a Pareto distribution using the original data and then calculated the mean of this distribution above the top code.[15] The wage rates and the associated transition matrix for US are given in Table 1 below.

Table 1 Hourly wages and transition matrix for the United States Economy 1984-1988 and 2004-2008

	Middle aged		1	2	3	4
Young	Wages		10.319	16.079	22.624	44.813
		Probabilities				
1	8.723		0.413	0.587	0	0
2	13.02		0	0.531	0.469	0
3	23.82		0	0	0.547	0.453

(Wages in 2003 US Dollars)

5.2.2. Sweden

The wage process for Sweden is derived from a representative panel of the Swedish population (LINDA) which covers around 3 percent of the population each year. It is a combination of income tax registers, population censuses and other sources. We obtain the hourly wage by dividing the yearly income by an estimate of the total number of working hours for a full-time employee (currently 1880). To estimate the transition probabilities we follow the same general procedure as was used for the US data. Thus, the wage distribution in the early period is divided into three wage bins and all men who were 25 years old during the period 1976-1980 are sorted into these bins. The same thing is done for the wage distribution of these same individuals 25 years later (when they are 50). At the later period the

[15] We have adopted the procedure outlined in Schmitt (2003) which is applied to the CEPR extracts. The procedure of applying a Pareto distribution to top incomes in the context of top-coded earnings data has also been used by Saez and Veall (2003).

wage distribution is divided into four wage bins. The Markov probabilities are then given as the relative frequencies of each of the possible wage paths represented by movements in and out of the wage bins.[16] The wage rates and the associated transition matrix for Sweden are given in Table 2 below.

Table 2 (Wages and transition matrix for the Swedish Economy 1976-1980 and 2001-2005)

Young \ Middle aged			1	2	3	4
Young	Wages		8.792	14.930	18.536	25.922
		Probabilities				
1	4.530		0.342	0.658	0	0
2	12.195		0	0.514	0.486	0
3	17.985		0	0	0.684	0.316

(Wages expressed in US Dollars by taking wages expressed in 2003 Swedish Kronor and then converting them into US Dollars based on the average exchange rate over the period 2001-2005)

There are a few points which we would like to emphasize. For the purpose of obtaining the transition probabilities we use data from two points in time. However, for both countries, the wages reported in table 1 and table 2 are taken from the most recent of the two sample periods. Given our assumptions, this is correct for two reasons. First, if we were to use the earlier period for the wages of the young, and the more recent sample period for the middle aged, the wages for the latter would include wage changes (for instance productivity growth) occurring in the data, but not part of the model and which do not represent the evolution of skills that our skill process is meant to represent.[17] The second reason is that in the AID tax optimum the self-selection constraints are related to a tax schedule common to both young and old agents, and agents should properly be modeled as having the possibility only to mimic other agents alive in the *same period*.

Comparing the wage processes in Table 1 and 2 we find that the wage distribution is more equal in the Swedish economy.[18] With respect to upward wage mobility we find that it is larger in Sweden in the lower part of the wage distribution, but that upward mobility for

[16] The final sample size is 6981. Individuals with an hourly wage rate of less than SEK 10 or greater than SEK 300 (in 1976 prices) were dropped from the sample.

[17] Recall that our model economy is assumed to be in a steady state which means that the wage distributions for young and old are assumed to be identical in the early and the late period.

[18] The estimated Gini coefficient for the US data is 0.267 for young agents, 0.329 for old agents and 0.327 for the combined population of young and old agents. For the Swedish data it is 0.196 for young agents, 0.268 for old agents and 0.262 for the combined population of 25 and 45 year old agents.

those with high wages as young is larger in the US. Thus, both the wage distributions and transition probabilities differ between the two economies.

5.3 Computational approach

The optimal tax problems in this paper are constrained nonlinear optimization problems and have been solved using sophisticated interior-point/barrier methods.[19] In cases where the problems are not entirely concave, the algorithms have been combined with global optimization heuristics initiating the solver from many different starting points with the ambition of finding all local optima, and selecting the one which results in the highest objective function value (all variables subject to optimization were properly normalized to ensure the efficiency of this procedure). The main computational difficulty lies in the structure of (and the number of) self-selection constraints. In some cases these difficulties were resolved by solving a sub-problem with a reduced number of constraints, and checking whether the optimal solution to the sub-problem is feasible with the full number of constraints. Care has also been taken to check the Lagrange multipliers for the possible failure of constraint qualifications (see Judd and Su (2006)).

6. Results

6.1 Social welfare function and welfare gains measure

In most numerical optimal tax studies in the Mirrlees tradition, the government maximizes a concave transformation $W(U)$ of the vector of individual utilities U. However, it is the combined curvature of W and U which determines the preference for redistribution, hence changing preference parameters (through its effect on the concavity of U) does not only have the intended effect on behavior but at the same time changes the redistributive taste of the government. To avoid confounding the curvature of W and U, and in order to obtain the cleanest measure of the welfare gain of policy reform, we adopt the maximin social welfare function, widely applied in the theoretical optimal taxation literature. Thus, $W(U) = U_{\min}$, where $U_{\min}$ denotes the expected lifetime utility of the least well off agent in the economy.

To calculate a consumption-based measure of the welfare gains attainable by more sophisticated tax schemes, we consider an equivalent-variation-type of welfare gain measure,

[19] The problem was set up in the modeling language AMPL and then solved using the nonlinear optimization solver KNITRO by Ziena Optimization Inc.

taking as a benchmark the solution to the government's problem under an AID *linear* income tax with unrestricted debt policy. To obtain the welfare gain measure, we proceed as follows. We look for the minimum amount of extra revenue that should be injected, in the optimal AID linear taxation problem (with unrestricted debt policy), into the government's budget in order to achieve the same social welfare level as under a more sophisticated tax system. Once we have found this minimum amount of extra revenue, we divide it by the aggregate GDP at the AID linear pure tax optimum (with unrestricted public debt policy) in order to get a revenue-based measure of the welfare gains.

6.2. Welfare gains: an overview

In this sub-section we provide an overview of the welfare gains obtained by an AD tax. We do this for the benchmark parametrization of the utility function. We present results both when debt policy is unrestricted and when public debt is restricted to be non-negative (or, equivalently, when per capita debt is restricted to be zero, given that the optimal unrestricted per capita debt would be negative). The results are summarized in tables 3 and 4 below.

Table 3 Welfare gains (as percentage of GDP) for US

Tax system	Welfare gain
Linear AID Benchmark	-
Linear AID, Zero Debt	-2.54%
Linear AD, Zero Debt	$\approx$-0.01%
Linear AD	$\approx$0.01%
Nonlinear AID, Zero Debt	6.34%
Nonlinear AID	6.91%
Nonlinear AD	8.88%

AD improves by **2.54%** due to slackening of IC + effect on capital accumulation

AD improves by **1.97%** due to slackening of IC

We start by commenting on the results for the US economy. The perhaps most striking result in table 3 is the large welfare gain obtained when moving from a linear- to a nonlinear income tax. Comparing the linear AID with the nonlinear AID, both without restrictions on debt policy, the welfare gain amounts to about 6.9% of GDP. Comparing the linear AID with the nonlinear AID, both considered under the zero debt restriction, the welfare gain is almost 9%, and the same welfare gain is obtained comparing the linear AD with the nonlinear AD.

Let's now focus on the welfare gains from age-dependency. For the linear tax the gains are virtually zero if the policy maker can freely use debt policy, so that the golden rule can be achieved. If the policy maker is restricted in the use of public debt, the welfare gain of moving from the AID to the AD is about 2.5% of GDP. The reason is that the AD tax almost achieves the Golden Rule whereas the AID does not. The reason why the AD does not fully achieve the golden rule is that we have imposed a restriction that the demogrants must be nonnegative.

The AD nonlinear tax achieves the golden rule even if public debt cannot be used. This means that the gains of moving from a nonlinear AID tax to the nonlinear AD are of two kinds. One is due to the fact that the golden rule can always be achieved under the AD; another is that the AD tax mitigates the self-selection constraints. The total gain in moving from the nonlinear AID with a zero debt restriction to the nonlinear AD amounts to 2.54% of GDP (8.88%-6.34%). This total gain can be decomposed into one part due to the power of the AD to achieve the golden rule even without public debt; this part amounts to 0.57% of GDP (6.91%-6.34%). The other part is due to the fact that the AD mitigates some of the self-selection constraints; this part amounts to 1.97% of GDP (8.88%-6.91%).

In table 4 we show the corresponding welfare gains obtained for the Swedish economy. It should be noted that the only difference between our representations of the US and Swedish economies refers to the wage processes. All other aspects of the economies are the same. Hence, the difference in results between table 3 and table 4 depends entirely on the difference in the wage processes. The gains of moving from a linear- to a nonlinear tax are of the same magnitude as for the US economy, i.e. around 7-9% of GDP. The gains of moving from a nonlinear AID with a zero debt restriction to a nonlinear AD is around 3.34%, i.e. about 0.8 percentage points higher than for US. A gain of about 1.78% of GDP is due to the power of the AD to achieve the Golden Rule and a gain of about 1.56% is due to the power to mitigate self-selection constraints.

Table 4 Welfare gains (as percentage of GDP) for Sweden

Tax system	Welfare gain
Linear AID Benchmark	-
Linear AID, Zero Debt	-3.30%
Linear AD, Zero Debt	0.19%
Linear AD	0.57%
Nonlinear AID, Zero Debt	5.98%
Nonlinear AID	7.76%
Nonlinear AD	9.32%

AD improves by **3.34%** due to slackening of IC + effect on capital accumulation

AD improves by **1.56%** due to slackening of IC

After having given an overview of the welfare gains we move on to describe the characteristics of the optimal tax systems.

6.3 US: Nonlinear taxation with unrestricted debt policy

Table 5 below presents the most relevant characteristics of the optimal nonlinear AD and the optimal nonlinear AID labor income tax. The values for the AID scenario are calculated under the assumption that the government is unrestricted in its debt policy. Since the AD tax achieves the golden rule even when public debt is zero, the AD tax is computed under the assumption that public debt is zero. We can see that, as compared with the results for the nonlinear AID setting, the AD labor income tax entails a shift of the tax burden on labor income from the young- to the middle-aged-workers: all the average labor income tax rates on young workers are lowered whereas all the average labor income tax rates on middle-aged-workers are increased.

With respect to the effects on labor supply, we find that all young workers, with the exception of the least skilled agents, increase their labor supply when moving from the AID to the nonlinear AD tax. For the middle-aged-workers, we find that all agents, with the exception of the group with the second highest wages, increase their labor supply when moving from the AID to the nonlinear AD tax. The labor supply pattern is also reflected in how the marginal tax rates on the labor income vary when moving to an AD tax.[20] Among the young, the

[20] As is common practice in the optimal taxation literature, the marginal income tax rates are calculated using the first order conditions characterizing the agents' behavior. More precisely, since an optimizing agent equalizes the marginal rate of substitution (MRS) between pre-tax labor income and consumption to one minus the marginal tax rate on labor income, it is possible to express the marginal labor income tax rate implicitly faced by an agent at the equilibrium allocation as 1-MRS. It should be noticed that the implicit marginal tax rate on labor

marginal tax rate on labor income is raised for the least skilled workers, whereas it declines for the other types of young workers. Among the middle aged, the marginal tax rate on labor income increases for the group with the second highest wages, and declines for all the remaining types of middle-aged-workers.

The shift from a nonlinear AID to a nonlinear AD tax entails for US a more generalized reduction in the optimal marginal tax rates than what we find for Sweden (see table 10 below). This appears consistent with the welfare gain results displayed in tables 3 and 4. There, we have seen that the welfare gain of the shift is larger for Sweden when we consider the “full” welfare gain, that takes into account both the gains descending from incentive-compatibility effects and those descending from capital accumulation effects. However, we have also seen that in the case of US the component related to incentive-compatibility effects explains more than 3/4 of the “full” welfare gain, whereas for Sweden the corresponding fraction is about 1/2. Because of that, the magnitude of the welfare gain is larger for US when we restrict attention to the role of an AD tax as an instrument to slacken binding self-selection constraints (1.97% for US as compared to 1.56% for Sweden). Since high marginal income tax rates signal the need to distort agents’ behavior in order to prevent mimicking (or, equivalently, in order to attain incentive-compatibility), slackening the self-selection constraints allows the government to lower the distortions required to ensure incentive-compatibility. Taking into account that the stronger the effect on the self-selection constraints and the larger the reduction in the marginal income tax rates that the government can afford, we can rationalize why the adoption of a nonlinear AD tax brings about in US a more generalized reduction in the marginal tax rates on labor income.

income depends not only on the (y,b)-bundle under consideration but also on the savings behavior of agents. This is because the marginal utility of consumption depends both on the after-tax labor income and on the level of savings.

Table 5 US Nonlinear Taxation, Benchmark case

Age Dependent Tax		Welfare Gain = 8.88%			Capital Tax = 49.19%		
		Young			Middle aged		
Type	$E(U)$	y_i	$T(y_i)$	$T'(y_i)$	y_{ij}	$T(y_{ij})$	$T'(y_{ij})$
(1,1)	30.175	4.278	-18.134	55.79%	6.301	0.621	59.78%
(1,2)					15.061	6.626	49.64%
(2,2)	30.355	8.815	-15.120	50.58%	15.061	6.626	49.74%
(2,3)					22.447	11.214	56.81%
(3,3)	30.792	30.004	-3.274	17.38%	22.447	11.214	51.05%
(3,4)					65.302	40.606	41.93%
Age Independent Tax		Welfare Gain = 6.91%			Capital Tax = 4.63%, Debt = -6.038		
		Young			Middle aged		
Type	$E(U)$	y_i	$T(y_i)$	$T'(y_i)$	y_{ij}	$T(y_{ij})$	$T'(y_{ij})$
(1,1)	30.079	5.239	-8.363	49.36%	5.239	-8.362	68.19%
(1,2)					11.625	-3.910	65.05%
(2,2)	30.280	8.454	-6.173	57.92%	11.625	-3.910	65.99%
(2,3)					26.206	5.045	44.74%
(3,3)	30.748	26.206	5.045	38.55%	26.206	5.045	41.51%
(3,4)					63.281	31.000	47.77%

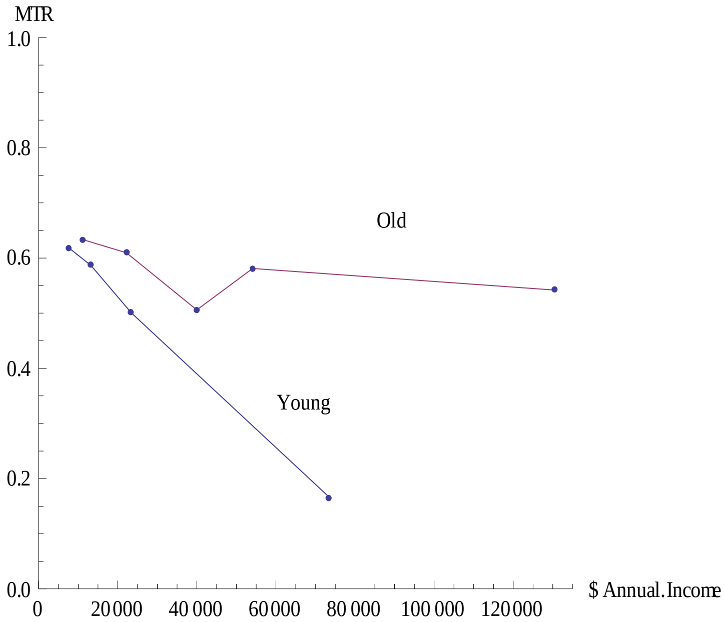
A somewhat unexpected feature of the results displayed in table 5 is the partial pooling that occurs among the middle-aged-workers. When calculating the solution to the government's problems, we did not impose any restriction requiring that the same (y,b)-bundle should be offered to all middle-aged-workers of a given ability type, irrespective of their ability type when young.[21] This restriction would be natural in a model that doesn't allow people to save, because then the slope of the indifference curves in the (y,b)-space of a middle-aged-worker would only depend on his current wage rate and not on his wage rate (ability type) when young. However, when people are allowed to freely save/borrow, the slope of the indifference curves in the (y,b)-space of a middle-aged-worker depend in general both on his current wage rate and on his past wage rate. This means that it is in principle possible for the government to separate the middle-aged-workers of a given ability type depending on their type when young. This possibility notwithstanding, pooling is obtained at an optimum both under a nonlinear AID and under a nonlinear AD tax.

[21] Formally, we didn't impose the restriction $\left(y_{ij}, b_{ij}\right) = \left(y_{kj}, b_{kj}\right)$.

The column headed $E(U)$ in table 5 shows the expected utility of the various agents. We see that, even though our welfare criterion maximizes the expected utility for the young type with lowest expected utility, it is still the case that for all types the nonlinear AD tax gives a higher expected utility than the nonlinear AID tax.

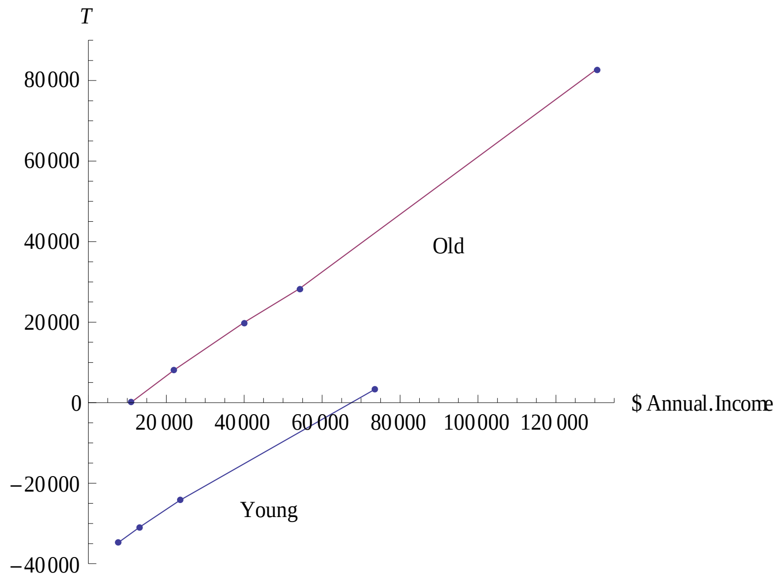
For computational reasons we had to restrict ourselves to three types in the first period and four types in the second period. However, exploiting the fact that there are fewer self-selection constraints for the AD tax, we can make calculations for the AD tax also for a four by five case. Although for this more general case we cannot calculate the welfare gains of age-dependency because we do not calculate the AID optimum, the four by five case allows us to provide a more detailed characterization of the AD tax. As for the three by four case we have maintained the restriction on transition probabilities. We graphically illustrate the tax regime below. Figure 2 shows how the marginal tax rate varies with income for the young and the middle aged. The young consistently face a lower marginal tax than the middle aged. For static optimal tax models there are two typical profiles for the marginal tax rates. One is a U-shaped profile, which we obtain for the middle-aged-workers. The other is a continuously declining profile, which is the one that we obtain for the young workers.

Figure 2. Marginal tax rates for young and middle aged, extended model



In figure 3 we show how the tax varies with income, with the middle-aged-workers paying much higher taxes than the young. This is of course to drive savings to a level compatible with the golden rule. (For an AD tax under a regime with unrestricted public debt the vertical location of the tax schedules in figure 3 would be indeterminate.)

Figure 3. **Tax functions for young and middle aged, extended model**



6.4 US: Nonlinear taxation with restrictions on debt policy

In the simulations for the AID tax under no restriction on debt policy, we find that public debt should be negative. That is, the government should be a net lender to individuals. This is not a phenomenon we observe in actual economies. Rather, in most economies the public sector borrows from the private sector and the stock of public debt is sometimes quite high. We therefore believe it is of interest to study the case where we impose the restriction that public debt must be non-negative. In table 6 we show simulation results obtained for the nonlinear AID tax under this restriction. The nonlinear AID tax cannot achieve the golden rule in this case; what is obtained is a steady state where the capital/labor ratio is lower than the one prescribed by the golden rule.

Table 6 US Nonlinear Age Independent Tax, Benchmark case with zero debt restriction

Age Independent Tax (d=0)		Welfare Gain = 6.34%			Capital Tax = 65.4%		
		Young			Middle aged		
Type	$E(U)$	y_i	$T(y_i)$	$T'(y_i)$	y_{ij}	$T(y_{ij})$	$T'(y_{ij})$
(1,1)	30.050	4.401	-9.052	40.72%	4.401	-9.052	62.73%
(1,2)					9.643	-5.709	59.72%
(2,2)	30.250	7.299	-7.261	48.90%	9.643	-5.709	60.66%
(2,3)					21.958	1.090	35.43%
(3,3)	30.715	21.958	1.090	28.43%	21.958	1.090	31.90%
(3,4)					53.401	21.467	38.51%

Comparing these results with those obtained for the AID tax with unrestricted debt we see a dramatic increase in the tax on capital income, which increases from 4.6% to 65.4%. We provide comments on this in section 6.8 below.

6.5 US: Sensitivity analysis

In this sub-section we perform some sensitivity analysis. We begin by considering alternative values for the coefficient of relative risk aversion and the Frisch elasticity. More precisely, we consider how the welfare gains change when $\gamma = 0.7$ or $\gamma = 1.3$, and when $\kappa = 2$ or $\kappa = 4$. The results are presented in table 7 where we compare the nonlinear AID tax with unrestricted debt policy to the nonlinear AD tax (thus, we are only capturing the welfare gains due to alleviation of incentive constraints). From this table we see that the welfare gains are increasing in γ and decreasing in κ; equivalently, the welfare gains are increasing in the value of the coefficient of relative risk aversion and of the Frisch elasticity.[22]

[22] Given that these welfare gain comparisons have been based on the assumption that the government faces no restriction on debt policy, they only reflect the welfare gain component ascribable to the beneficial effect on the incentive-compatibility constraints thwarting the government's redistributive policy.

Table 7 Welfare Gains Nonlinear Taxation – Sensitivity Analysis

Variation in γ	$\gamma = 0.7; \kappa = 2.25$ AD = 8.72% AID = 6.81% Diff = 1.91%	$\gamma = 0.9; \kappa = 2.25$ AD = 8.88% AID = 6.91% Diff = 1.97%	$\gamma = 1.3; \kappa = 2.25$ AD = 9.02% AID= 6.75% Diff = 2.27%
Variation in κ	$\gamma = 0.9; \kappa = 2.00$ AD = 8.30% AID = 6.06% Diff = 2.24%	$\gamma = 0.9; \kappa = 2.25$ AD = 8.88% AID = 6.91% Diff = 1.97%	$\gamma = 0.9; \kappa = 4.00$ AD = 9.21% AID = 8.04% Diff = 1.17%

To assess the sensitivity of results with respect to the wage process, we have also performed some experiments with changes in the transition probabilities. In particular, we have assumed a “high persistency” (or “low uncertainty”) scenario where for all young agents the probability of maintaining the same skill rank when middle aged is equal to 0.98, whereas the probability to move upwards is equal to 0.02. To save space we summarize here the main results without providing tables.

For the US benchmark parameterization case, the welfare gain of a nonlinear AD tax is equal to 9.43% (as compared to the benchmark linear AID tax with unrestricted debt policy), whereas for the nonlinear AID it is equal to 7.43%. Thus, both welfare gains becomes larger than under the “correct” transition probabilities, but the welfare gain of moving from a nonlinear AID to an AD tax is virtually unaffected (since it is 2% and it was 1.97%).

Interestingly, we also find that type (1,2) and type (2,2) are no longer pooled together in the nonlinear AID case. Doing the same experiment but with the parameterization γ=1.3 and κ=2.25 we were able to eliminate all pooling under the nonlinear AD tax, whereas for the nonlinear AID tax pooling was eliminated only for types (1,2) and (2,2). The welfare gain of a nonlinear AD tax becomes in this case 9.3% and 7.26% is the welfare gain of a nonlinear AID tax. Thus, the welfare gains become larger (and the difference is more substantial than in the benchmark parametrization scenario) also in this case. However, the welfare differential between the nonlinear AD tax and the nonlinear AID tax becomes 2.04% as compared to 2.27% that is obtained with the “correct” transition probabilities. Thus, the welfare gain of switching from a nonlinear AID to an AD tax appears to become smaller with "high persistency" only when the coefficient of relative risk aversion is fairly large.

Finally, when the government is empowered with nonlinear income taxes, the capital tax rate remains substantially different from zero even under "high persistency" in the transition probabilities.

6.6 US Linear taxation

In this sub-section we present the simulation results for the case when age-dependency is nested upon a linear taxation system. In table 8 we present the results for the different linear tax optima. In the presence of unrestricted debt policy linear AD taxes offer only a modest welfare gain as compared to the benchmark economy. However, when there are restrictions on debt policy, the possibility of having AD demogrants offers a significant advantage as it allows the linear AD tax to partially replicate the intergenerational transfer implicit in the optimum with the more powerful unrestricted debt instrument. This shows that the advantage of AD taxes is not limited to situations where sophisticated nonlinear tax instruments are available. A compelling argument for AD taxes exists also in linear taxation framework.

If, for any given assumption about the availability of debt policy, we compare the values for the AD scenario with those for the AID scenario, we can see that age-dependency entails an increase in the degree of progressivity of the labor income tax schedule faced by young workers and a reduction in the degree of progressivity of the labor income tax schedule faced by middle-aged-workers. Under the non-negativity of debt restriction, we can even see that the optimal AD labor income tax on middle-aged-workers is proportional, rather than linear. Here it is our restriction that demogrants cannot be negative that is binding. Removing this restriction would generate a regressive labor income tax schedule on middle-aged-workers.

In table 8 we also present results for alternative values of the utility function parameters. We can see that the optimal capital tax rate is increasing in the value of the coefficient of relative risk aversion and of the Frisch elasticity.

Table 8 US Linear Tax System

Optimum	Parameterization	τ_K	τ_1	τ_2	G_1	G_2	d	R_{WF}
AID	$\gamma = 0.7$, $\kappa = 4$	4.43%	64.77%	64.77%	12.232	12.232	-6.555	-
	Benchmark	**10.84%**	**54.13%**	**54.13%**	**9.269**	**9.269**	**-6.443**	**-**
	with d=0	*66.04%*	*47.58%*	*47.58%*	*8.726*	*8.726*	-	*-2.55%*
	$\gamma = 1.3$, $\kappa = 2$	17.84%	54.56%	54.56%	5.908	5.908	-3.977	-
AD	$\gamma = 0.7$, $\kappa = 4$	2.16%	65.94%	63.90%	12.204	12.204	-6.667	0.007%
	Benchmark	**7.25%**	**55.20%**	**53.54%**	**9.238**	**9.238**	**-6.540**	**0.006%**
	with d=0	*15.58%*	*54.68%*	*53.21%*	*16.787*	*0*	-	*-0.009%*
	$\gamma = 1.3$, $\kappa = 2$	13.17%	55.51%	54.11%	5.883	5.883	-4.047	0.004%

6.7 Sweden: Nonlinear taxation with and without restrictions on debt policy

Table 9 below presents the most relevant characteristics of the optimal nonlinear AD and the optimal nonlinear AID labor income tax for the Swedish economy. Although the results are not the same as for US, by and large the results are qualitatively similar.

Table 9 Sweden Nonlinear Taxation, Benchmark case

Age Dependent Tax		Welfare Gain = 9.32%			Capital Tax = 49.5%		
		Young			Middle aged		
Type	$E(U)$	y_i	$T(y_i)$	$T'(y_i)$	y_{ij}	$T(y_{ij})$	$T'(y_{ij})$
(1,1)	29.429	1.181	-16.667	68.24%	4.816	1.574	69.52%
(1,2)					18.370	10.443	34.02%
(2,2)	29.613	9.587	-10.537	44.08%	18.370	10.443	33.38%
(2,3)					22.573	12.149	40.15%
(3,3)	30.017	23.096	-4.467	7.46%	22.573	12.149	30.63%
(3,4)					35.393	18.271	28.11%
Age Independent Tax		Welfare Gain = 7.76%			Capital Tax = 61.4%, Debt = -6.631		
		Young			Middle aged		
Type	$E(U)$	y_i	$T(y_i)$	$T'(y_i)$	y_{ij}	$T(y_{ij})$	$T'(y_{ij})$
(1,1)	29.356	1.232	-9.134	65.72%	5.113	-5.908	69.35%
(1,2)					20.304	2.799	21.70%
(2,2)	29.607	7.766	-4.166	56.09%	20.304	2.799	24.19%
(2,3)					23.697	4.084	37.31%
(3,3)	29.968	20.304	2.799	24.84%	23.697	4.084	29.78%
(3,4)					31.939	8.634	44.86%

In table 10 we present the results obtained for the Swedish wage process when public debt is restricted to be non-negative. Once again, the results are qualitatively similar to those obtained for the US economy.

Table 10 Sweden Nonlinear Age Independent Tax, Benchmark case with zero debt restriction

Age Independent Tax (d=0)		Welfare Gain = 5.98%			Capital Tax = 88.9%		
		Young			Middle aged		
Type	$E(U)$	y_i	$T(y_i)$	$T'(y_i)$	y_{ij}	$T(y_{ij})$	$T'(y_{ij})$
(1,1)	29.269	0.955	-9.052	57.44%	3.712	-6.853	64.34%
(1,2)					15.318	-1.341	5.47%
(2,2)	29.517	6.229	-7.261	42.50%	15.318	-1.341	8.48%
(2,3)					18.803	-0.578	16.67%
(3,3)	29.887	15.318	-1.090	9.09%	18.803	-0.578	8.20%
(3,4)					24.418	1.836	32.20%

6.8 Sweden: Linear taxation

In table 11 we present results for the Swedish wage process under the assumption that a linear tax system is adopted. Here also the results are similar to those obtained for the US economy. However, one difference concerns the size of the optimal capital income tax rate which is considerably higher for Sweden than for the US when an AID tax with unrestricted debt policy is considered.

Table 11 Sweden Linear Tax System, Benchmark case

Optimum	Debt	τ_K	τ_1	τ_2	G_1	G_2	d	R_{WF}
AID	Yes	42.53%	50.70%	50.70%	6.998	6.998	-5.696	-
	No	77.46%	43.34%	43.34%	6.520	6.520	-	-3.30%
AD	Yes	3.57%	60.80%	43.24%	6.563	6.563	-6.552	0.57%
	No	41.55%	58.14%	41.31%	11.978	0	-	0.19%

Having completed the presentation of our numerical results, in the next sub-section we provide some further comments on the results that we have obtained with respect to the optimal level of the capital income tax rate.

6.9 Capital income taxation

There is a long standing interest in the question of whether there should be a positive tax on capital income. Our simulations shed light on this issue. We consider several different specifications of our model and in all specifications we obtain a positive tax on interest income.[23] However, the numerical magnitude varies quite a lot: from 4.6% for the US nonlinear AID tax under unrestricted debt policy up to 88.9% for the Swedish nonlinear AID tax with public debt restricted to be non-negative. There are basically two different mechanisms that in our model generate the result that there should be a positive interest income tax. One is that such a tax helps to mitigate self-selection constraints, another is that it can help capital accumulation and move the economy closer to the golden rule.

The first mechanism is best illustrated by the tax systems (economies) where the set of policy instrument is so rich that the golden rule capital/labor ratio is obtained without any help from interest income taxation. These are the nonlinear AID tax under unrestricted public debt and the nonlinear AD tax systems. For these tax systems a non-zero tax on interest income occurs only if it mitigates self-selection constraints. The Atkinson-Stiglitz result of a zero commodity tax really requires both that labor is weakly separable from commodities in the utility function *and* that labor is weakly separable from commodities in the individuals' budget constraints (this latter requirement has received too little emphasis in the literature). In our model labor is weakly separable from commodities in the utility function. However, since the wage process is such that the second period wage rate is uncertain for the individual, the budget constraint becomes non-separable. The standard result from the optimal commodity tax literature becomes applicable. One should tax more heavily the commodities for which the mimicker's consumption is higher than the consumption of the mimicked's. Hence, a positive capital income tax rate becomes desirable if, at the binding self-selection constraints to the government's problem, the savings of the mimickers tend to exceed the savings of the agents being mimicked. The exact formula for the optimal interest tax rate in our model is given in Appendix D. From the simulations we see that the optimal value of the capital income tax rate jumps around quite a lot. However, this is hardly surprising given that τ_K depends solely on the difference between the amount of savings of a mimicker, and that of the agent being mimicked, at the various self-selection constraints that are binding at the solution to the

[23] This is in line with many recent studies like, for example, Conesa et al. (2009) who find that the tax on capital income should be around 30%. Early studies, like Judd (1985) and Chamley (1986), using a model with a representative consumer with infinite life, obtained the result that the tax on capital income should be zero.

government's problem. Since a change in the wage process or a switch from a nonlinear AID labor income tax to a nonlinear AD labor income tax is likely to bring about significant changes in the set of self-selection constraints that are binding at an optimum, one should not expect to find any clear pattern in how the optimal value for τ_K is affected.[24]

Let's now consider the second mechanism that can explain the desirability of distorting the agents' intertemporal consumption choices. That a tax/subsidy on interest income can help capital accumulation is perhaps best illustrated in the linear tax model with zero debt. For an AD tax with no restrictions on the demogrants, savings, and hence the capital stock, can be perfectly controlled by setting the demogrants in the proper way. For the linear AID tax the demogrant must be the same in both periods. However, a tax on interest income can serve as an imperfect substitute for AD demogrants. As our model is set up, individuals save when young and pay tax on interest income in subsequent periods. Increasing the tax on interest income and using the proceeds to increase the demogrant implies a redistribution of resources to the young, hence it will induce increased savings. Of course, an increase in the tax on interest income will distort the relative price between consumption in different periods, which limits the extent to which one would like to use this mechanism to increase the steady state capital stock. The large increase in the interest tax rate from 10.8% to 66% that obtains for the US linear AID tax as we impose a zero debt restriction can best be understood as a way to increase individual savings and hence the capital stock. For the Swedish economy we get an increase from 42.5% to 77.5% as the zero debt restriction is imposed for the linear AID tax.

For the nonlinear AID tax under a zero debt restriction there can be a non-zero tax on interest income both because such a tax can mitigate self-selection constraints, and because it can help capital accumulation. For the US, we see that when the zero debt restriction is imposed, the interest tax rate increases from 4,6% to 65.4% under an optimal nonlinear AID labor income tax. However, it is not clear that one should attribute all of this increase to the capital accumulation mechanism as the set of self-selection constraints also will change as we introduce the zero debt condition.

We have seen that the size of the optimal tax on interest income varies to a very large extent as the specification of the model varies. An implication of this is that we should not expect to be able to set the tax on interest income at the correct (optimal) level in real

[24] This point is further investigated in Appendix D where we provide an analytical expression for the condition characterizing the optimal level of the interest income tax rate.

economies. However, this does not necessarily imply large welfare losses. For the US economy we have done simulations looking at how the welfare gains (with respect to the benchmark linear AID tax scenario) are reduced when we deprive the government of the capital tax instrument. Compared to the welfare gains of having a nonlinear instead of linear income tax or an AID instead of an AD tax, the gains of setting the tax on interest income at the optimal level are in most cases of second order importance; setting $\tau_K = 0$ we find that the welfare gain of a nonlinear AD tax is lowered from 8.88% to 8.60% and the welfare gain of a nonlinear AID tax is lowered from 6.91% to 6.89%.[25] Of a similar order of magnitude are the reductions in the welfare gain values when linear income taxation schemes are considered. However, there is one exception to this pattern of results. For the US nonlinear AID tax under a zero debt restriction the welfare loss of restricting the capital tax rate to zero amounts to about 3%.

8. Concluding remarks

In this paper we have quantitatively assessed the welfare gains of an age-dependent tax. Our vehicle of analysis has been an overlapping generations model with heterogeneous agents, facing uncertainty regarding their future earnings capacities, and choosing labor supply and consumption optimally over their lifetime. For computational reasons we have not considered transitional dynamics but have focused on steady states. We have calibrated our model and estimated transitional wage paths using detailed wage data for both US and Sweden.

Our calculations show that the welfare gains from age-dependent taxes are substantial, especially when there are restrictions on the use of debt policy so that the tax instruments play a role in achieving the optimal level of capital accumulation. For US, under our benchmark specification of the utility function, the welfare gain of switching from an optimal nonlinear age-independent income tax to an optimal nonlinear age-dependent income tax is about 2.5% of total output. This gain can be decomposed into incentive effects arising from alleviating self-selection constraints (which account for 2%) and capital accumulation effects (which account for 0.5%). The results for the Swedish wage process are similar. Here the total welfare gain is around 3.5%. Slightly more than half of this stems from a slackening of self-selection constraints and slightly less than half from capital accumulation effects. Sensitivity analysis shows that the welfare gains of switching to a nonlinear age-dependent income tax are increasing in the coefficient of relative risk-aversion and in the value of the Frisch

[25] The results are similar in magnitude for Sweden.

elasticity. The optimal age-dependent tax is characterized by a shift of tax burden from young to middle-aged-workers, and the optimum features a declining profile of marginal tax rates for young workers and a U-shaped profile for middle-aged workers. Age-dependency allows in particular the government to lower the marginal tax rate for young high income earners.

It can be of interest to put the welfare gains obtained from moving to an age-dependent tax into perspective. Although sizable, the most powerful reform is the switch from linear to nonlinear taxation. In light of recent flat tax proposals, our model suggests that the implied welfare loss of replacing an optimal nonlinear income with a linear tax can be as large as 7-9% of total output. This result holds for several alternative specifications of the model and for both the US and Swedish wage processes. Another issue which has attracted considerable interest is the taxation of interest income, which has typically been the focus of intertemporal models of optimal taxes. In our simulations the optimal tax rate on interest income exhibits substantial variability across different specifications, ranging from a minimum of 4.6% to a maximum of 88.9%. An implication is that we should not expect to be able to set the tax on interest income at the correct (optimal) level in real economies. However, this does not necessarily imply large welfare losses. For the US economy our simulations show, at least when the government is unrestricted in its debt policy, that the welfare contribution of choosing optimally the level of the capital tax rate is of second order importance, ranging from 0 to 0.3% of total output.

REFERENCES

Akerlof, G. A. (1978), "The Economics of Tagging", *American Economic Review* 68, 8-19.

Alesina A, Ichino A. and Karabarbounis, L. (2010), "Gender-Based Taxation and the Division of Family Chores" Forthcoming in *American Economic Journal: Economic Policy*.

Banks, J. and Diamond P. (2010), "The Base for Direct Taxation", in Tax By Design, The Mirrlees Review, Oxford University Press.

Blomquist S. and Micheletto L. (2003), "Age Related Optimal Income Taxation", Working Paper 2003:7, Department of Economics, Uppsala University

Blomquist S. and Micheletto L. (2008), "Age Related Optimal Income Taxation", *Scandinavian Journal of Economics* 110, 45-71.

Boadway, R. and Pestieau, P. (2006), "Tagging and redistributive taxation", *Annales d'Economie et de Statistique* 83/84, 123-150.

Chamley, C. (1986), “Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives”, *Econometrica* 54, 607-622.

Cremer, H., Gahvari, F. and Lozachmeur, J.M., (2010), “Tagging and Income Taxation: Theory and an Application”, *American Economic Journal: Economic Policy* 2 (1), 31-50.

Conesa, J. C., Kitao S. and Krueger D. (2009), “Taxing Capital? Not a Bad Idea after All”, *American Economic Review* 99, 25-48.

Diamond, P. (1965), “National debt in a neoclassical growth model”, *American Economic Review* 41, 1126-1150.

Erosa, A. and Gervais, M. (2002), “Optimal Taxation in Life-Cycle Economies”, *Journal of Economic Theory* 105, 338-369.

Golosov, M., Tsyvinski, A. and Werning, I. (2006), “New Dynamic Public Finance: A User’s Guide”, in *NBER Macroeconomic Annual 2006,* MIT press.

Hammond, P.J. (1987), “Markets as constraints: multilateral incentive compatibility in continuum economies”, *Review of Economic Studies* 54, 399-412.

Imai, S. and Keane, M. P. (2004), “Intertemporal labor supply and human capital accumulation”, *International Economic Review* 45, 601-641.

Immonen, R., Kanbur, R., Keen, M. and Tuomala, M. (1998), “Tagging and taxing: the optimal use of categorical and income information in designing tax/transfer schemes”, *Economica* 65, 179-192.

Judd, K. (1985), “Redistributive taxation in a simple perfect foresight model”, *Journal of Public Economics* 28, 59-83.

Judd, K. and Su, C. L. (2006), “Optimal income taxation with multidimensional agents’ types”, Computing in Economics and Finance 2006 471, Society for Computational Economics.

Keane, M. P. (2009), “Income taxation in a life cycle model with human capital”. Unpublished paper.

Kremer, M. (2002), “Should taxes be independent of age?” Unpublished paper Harvard University.

Lozachmeur, J. M. (2006), "Optimal Age Specific Income Taxation", *Journal of Public Economic Theory* 8, 697-711.

Mankiw, N. G. and Weinzierl, M. (2010), "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution", *American Economic Journal: Economic Policy* 2 (1), 155-176.

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

Saez, E. and Veall, M. R. (2003), "The evolution of high incomes in Canada 1920-2000", NBER working paper no. 9607.

Schmitt, J. (2003), "Creating a consistent hourly wage series from the Current Population Surveyís Outgoing Rotation Group, 1979-2002", mimeo CEPR.

Weinzierl, M. (2008), "The Surprising Power of Age-Dependent Taxes". Mimeo, Harvard University.

Appendix A. Information on age is needed to achieve First Best

Full information First Best.

If there is information on each individual's productivity and age, the policy maker can assign values for $c_1, \ell_1, c_{11}, \ell_{11}, c_{12}, \ell_{12}$. Rewrite the Lagrangian in an equivalent way as:

$$\Lambda = U(c_1, \ell_1) + \beta\left[p_{11}U(c_{11}, \ell_{11}) + p_{12}U(c_{12}, \ell_{12})\right] + \lambda\left[\left(k + f\left(k\right) - (1+n)k\right)\tilde{L} - \tilde{C}\right].$$

The first order conditions with respect to ℓ_1, ℓ_{11} and ℓ_{12} are respectively:

$$U_{\ell_1} = -\lambda\left[\frac{\partial\tilde{L}}{\partial\ell_1}\left(k + f\left(k\right) - \left(1+n\right)k\right) + \tilde{L}\left(1 + f' - 1 - n\right)\frac{\partial k}{\partial\tilde{L}}\frac{\partial\tilde{L}}{\partial\ell_1}\right];$$

$$\beta p_{11}U_{L_{11}} = -\lambda\left[\frac{\partial\tilde{L}}{\partial\ell_{11}}\left(k + f\left(k\right) - \left(1+n\right)k\right) + \tilde{L}\left(1 + f' - 1 - n\right)\frac{\partial k}{\partial\tilde{L}}\frac{\partial\tilde{L}}{\partial\ell_{11}}\right];$$

$$\beta p_{12}U_{L_{12}} = -\lambda\left[\frac{\partial\tilde{L}}{\partial\ell_{12}}\left(k + f\left(k\right) - \left(1+n\right)k\right) + \tilde{L}\left(1 + f' - 1 - n\right)\frac{\partial k}{\partial\tilde{L}}\frac{\partial\tilde{L}}{\partial\ell_{12}}\right].$$

The golden rule condition $r = n$ allows simplifying the previous conditions as follows:

$$U_{\ell_1} = -\lambda\left[\frac{\partial\tilde{L}}{\partial\ell_1}\left(f\left(k\right) - nk\right)\right];$$

$$\beta p_{11}U_{\ell_{11}} = -\lambda\left[\frac{\partial\tilde{L}}{\partial\ell_{11}}\left(f\left(k\right) - nk\right)\right];$$

$$\beta p_{12}U_{\ell_{12}} = -\lambda\left[\frac{\partial\tilde{L}}{\partial\ell_{12}}\left(f\left(k\right) - nk\right)\right].$$

Noticing that $f\left(k\right) - nk = w$ is (under the golden rule condition $r = n$) the marginal product in efficiency unit of labor which, in a perfectly competitive factor market, would be equal to the wage rate per efficiency unit of labor, we can write:

$$U_{\ell_1} = -\lambda\frac{\partial\tilde{L}}{\partial\ell_1}w = -\lambda w\theta^1;$$

$$\beta p_{11}U_{\ell_{11}} = -\lambda\frac{\partial\tilde{L}}{\partial\ell_{11}}w = -\lambda w\frac{p_{11}\theta^1}{1+n};$$

$$\beta p_{12}U_{\ell_{12}} = -\lambda \frac{\partial \tilde{L}}{\partial \ell_{12}} w = -\lambda w \frac{p_{12}\theta^2}{1+n}.$$

Combining each of the first order conditions above with the corresponding f.o.c. for consumption gives:

$$MRS_{C_1,\ell_1} = \frac{U_{\ell_1}}{U_{C_1}} = -w\theta^1 = MRT_{C_1,\ell_1};$$

$$MRS_{C_{11},\ell_{11}} = \frac{U_{\ell_{11}}}{U_{C_{11}}} = -w\theta^1 = MRT_{C_{11},\ell_{11}};$$

$$MRS_{C_{12},\ell_{12}} = \frac{U_{\ell_{12}}}{U_{C_{12}}} = -w\theta^2 = MRT_{C_{12},\ell_{12}}.$$

That is, hours of work are undistorted.

Restricted First Best: No information on age.

We now assume that the policy maker has information on individuals' productivities but not on their age. We keep the assumption that the capital stock can be controlled. However, since age is not observed, consumption and labor for young and old low-skilled must be the same, i.e. $c_1 = c_{11}$ and $\ell_1 = \ell_{11}$. The Lagrangian now becomes:

$$\Lambda = U(c_1,\ell_1) + \frac{1}{R}\left[p_{11}U(c_1,\ell_1) + \pi^h U(c_{12},\ell_{12})\right] + \mu\left[k + f\left(k\right) - (1+n)k - \tilde{C}/\tilde{L}\right].$$

We still have the golden rule result:

$$\partial\Lambda / \partial k = \mu\left[1 + f'(k) - (1+n)\right] = 0.$$

The F.O.C. for c_1 and c_{12} are:

$$\frac{\partial\Lambda}{\partial c_1} = \left(1+\beta p_{11}\right)U_{c_1} - \frac{\mu}{\tilde{L}}\left(1 + \frac{p_{11}}{1+n}\right) = 0;$$

$$\frac{\partial\Lambda}{\partial c_{12}} = \beta p_{12}U_{c_{12}} - \frac{\mu}{\tilde{L}}\frac{p_{12}}{(1+n)} = 0.$$

The two equations above imply:

$$\frac{U_{c_{11}}}{U_{c_{12}}}=\frac{U_{c_1}}{U_{c_{12}}}=\frac{1+n+p_{11}}{p_{11}+1/\beta} \text{ and } MRS_{c_{12},c_{11}}=-\frac{p_{11}}{p_{12}}\frac{1+n+p_{11}}{p_{11}+1/\beta}\neq-\frac{p_{11}}{p_{12}}=MRT_{c_{12},c_{11}}$$

Since $c_1=c_{11}$ and $\ell_1=\ell_{11}$ it follows that $U_{c_{11}}=U_{c_1}$ and $MRS_{c_1,c_{11}}=-\beta p_{11}\neq-\frac{p_{11}}{1+n}=MRT_{c_1,c_{11}}$

Had we studied the F.O.C. for hours of work we would have found that hours of work also are distorted.

Appendix B. The nonlinear AD tax always achieves the golden rule.

In the absence of debt policy, the aggregate production constraint can be written as:

$$Y_t=N_t\tilde{L}_t f\left(k_t\right)=N_t\left[\sum_i p_i c_{i,t}+\frac{1}{1+n}\sum_{ij}p_i p_{ij}\left(c_{ij,t}+\frac{1}{1+n}c^R_{ij,t}\right)\right]+K_{t+1}-\left(1-\delta\right)K_t$$

and the capital market equilibrium condition is:

$$k_{t+1}-\left(1-\delta\right)\frac{k_t}{1+n}=\frac{s_t}{(1+n)\tilde{L}_{t+1}}+\frac{s_{t-1}}{(1+n)^2\tilde{L}_{t+1}}$$

Taking the steady-state version of the capital market equilibrium condition we get:

$$k=\frac{1}{n+\delta}\left(\frac{s}{\tilde{L}}+\frac{s_{-1}}{(1+n)\tilde{L}}\right)$$

Substituting the equation above into the steady-state version of the aggregate production constraint, and dividing by N_t , one obtains:

$$\tilde{L}f\left(\frac{1}{n+\delta}\left(\frac{s}{\tilde{L}}+\frac{s_{-1}}{(1+n)\tilde{L}}\right)\right)=\left[\sum_i p_i c_i+\frac{1}{1+n}\sum_{ij}p_i p_{ij}\left(c_{ij}+\frac{1}{1+n}c^R_{ij}\right)\right]+\frac{K_{t+1}-\left(1-\delta\right)K_t}{N_t}$$

Taking into account that $\frac{K_{t+1}-\left(1-\delta\right)K_t}{N_t}=\frac{k_{t+1}L_{t+1}-\left(1-\delta\right)k_tL_t}{N_t}=\left(1+n\right)k\tilde{L}-\left(1-\delta\right)k\tilde{L}$, and also that $\left(n+\delta\right)\tilde{L}k=s+\frac{s_{-1}}{1+n}$, we finally obtain the following steady-state version of the aggregate production constraint:

$$\tilde{L}f\left(\frac{1}{n+\delta}\left(\frac{s}{\tilde{L}}+\frac{s_{-1}}{(1+n)\tilde{L}}\right)\right)-\left[\sum_i p_i c_i+\frac{1}{1+n}\sum_{ij}p_i p_{ij}\left(c_{ij}+\frac{1}{1+n}c^R_{ij}\right)\right]-s-\frac{s_{-1}}{1+n}\geq 0$$

In the absence of debt policy we can therefore state the government's problem under a nonlinear age-dependent tax as follows:

$$\max_{\{b_i,y_i\}_{i\in I},\{b_{ij},y_{ij}\}_{i\in I,j\in J},\tau_K}\left(\min_i V_i(\tilde{\sigma}^i)\right)$$

subject to:

$$\forall\ i \in I\ : V_i(\tilde{\sigma}^i) \geq V_i(\sigma^i), \forall \sigma^i \in \Gamma_i$$

$$\tilde{L}f\left(\frac{1}{n+\delta}\left(\frac{s}{\tilde{L}}+\frac{s_{-1}}{(1+n)\tilde{L}}\right)\right)-\left[\sum_i p_i c_i+\frac{1}{1+n}\sum_{ij} p_i p_{ij}\left(c_{ij}+\frac{1}{1+n}c^R_{ij}\right)\right]-s-\frac{s_{-1}}{1+n}\geq 0\,.$$

Starting from an optimum where all the first order conditions with respect to the various government's instruments ($\{b_i, y_i\}_{i\in I}, \{b_{ij}, y_{ij}\}_{i\in I, j\in J}, \tau_K$) are satisfied, consider the following experiment: increase marginally all the after-tax labor incomes assigned to the various young workers, b_i, and decrease by $1+r(1-\tau_K)$ all the after-tax labor incomes assigned to the various middle-aged workers, b_{ij}. To assess the effect of this policy experiment, what is required is to multiply by $-\left[1+r(1-\tau_K)\right]$ the set of the first order conditions with respect to the various b_{ij}, and then to sum them up with the set of first order conditions with respect to the various b_i. Taking into account that, under a nonlinear AD tax, young workers are prevented from choosing a bundle on the income tax that applies to the middle-aged-workers, and vice versa, one can easily recognize that the effect of the policy experiment that we are considering is to leave unaffected the expected lifetime utilities of all agents. The reason is that for all agents, including those who were planning to behave as mimickers, the present value of lifetime disposable income is not affected. The only behavioral effect is given by a change in the savings behavior of young workers. Specifically, young workers marginally increase their savings in the first period in order to keep consuming the same amount of material goods and leisure in all periods and possible states of the world. Thus, the policy experiment that we are considering is going to have no effect both on the government's objective function (and this happens irrespective on whether the social welfare function is of the maximin type, as in our example, or of any other type) and on the set of self-selection constraints faced by the government. The only effect that we are left to consider is the effect on the aggregate production constraint. Rewriting the aggregate production constraint as

$$\tilde{L}f\left(\frac{1}{n+\delta}\left(\frac{s}{\tilde{L}}+\frac{s_{-1}}{(1+n)\tilde{L}}\right)\right)-\left\{\sum_i p_i\left(b_i-s_i\right)+\frac{1}{1+n}\sum_{ij} p_i p_{ij}\left[b_{ij}+s_i\left(1+r\left(1-\tau_K\right)\right)-s_i\left(j\right)+\frac{1}{1+n}c^R_{ij}\right]\right\}$$
$$-s-\frac{s_{-1}}{1+n}\geq 0,$$

and denoting by μ the Lagrange multiplier associated with the production constraint, the effect of the reform is given by:

$$-\mu\left[\sum_i p_i - \frac{1+r\left(1-\tau_K\right)}{1+n}\sum_{ij} p_i p_{ij}\right]$$

$$-\mu\left[-\frac{f'}{n+\delta}-\sum_i p_i + \frac{1+r\left(1-\tau_K\right)}{1+n}\sum_{ij} p_i p_{ij}+1\right]\left\{\sum_i \frac{\partial s}{\partial b_i}-\left[1+r\left(1-\tau_K\right)\right]\sum_{ij}\frac{\partial s}{\partial b_{ij}}\right\}=0,$$

where the fact that the expression above is equal to zero descends from the assumption that we started from an optimum where all the first order conditions to the government's problem were satisfied.

Since from our previous discussion we have concluded that $\sum_i \frac{\partial s}{\partial b_i}-\left[1+r\left(1-\tau_K\right)\right]\sum_{ij}\frac{\partial s}{\partial b_{ij}}=0$, the equation above can be simplified to:

$$-\mu\left[\sum_i p_i - \frac{1+r\left(1-\tau_K\right)}{1+n}\sum_{ij} p_i p_{ij}\right]-\mu\left[-\frac{f'}{n+\delta}-\sum_i p_i + \frac{1+r\left(1-\tau_K\right)}{1+n}\sum_{ij} p_i p_{ij}+1\right]=0,$$

or, equivalently, upon further simplifications, to:

$$\frac{f'}{n+\delta}=1 \Rightarrow f'=n+\delta \Rightarrow r=n,$$

given that, with perfectly competitive markets, factors earn their marginal products implying that $r=f'-\delta$.

We can therefore conclude that, by combining the various first order conditions to the government's problem under a nonlinear AD tax, one can derive the golden rule condition even in the absence of debt policy.

Appendix C. The nonlinear AID tax and the conditions for golden rule.

In Appendix B we showed that debt policy is redundant to achieve the golden rule condition when the government is empowered with a nonlinear AD tax. The proof crucially relied on the fact that an AD tax makes impossible for workers to implement mimicking strategies where they choose when young an income bundle intended for middle aged, or vice versa. Given that this condition is no longer satisfied under an AID tax, it should not be surprising that the first order conditions of the government's problem do not allow to recover the golden rule condition in the absence of debt policy. To realize that this is the case one can follow a procedure similar to that employed in Appendix C. Starting from an optimum where all the first order conditions with respect to the various government's instruments $(\{b_i, y_i\}_{i\in I}, \{b_{ij}, y_{ij}\}_{i\in I, j\in J}, \tau_K)$ are satisfied, consider the following experiment: increase marginally all the after-tax labor incomes assigned to the various young workers, b_i, and decrease by $1+r(1-\tau_K)$ all the after-tax labor incomes assigned to the various middle-aged-workers, b_{ij}. Under an AID tax nothing guarantees that for all agents, including those who were planning to behave as mimickers, the present value of lifetime disposable income is left unaffected by the reform. In particular, the reform would change the present value of lifetime disposable income for those agents planning to mimic by choosing a deviating strategy which entails picking at some age an income bundle intended for agents of different age. For mimickers planning to implement such deviating strategies, the reform would not be welfare-neutral, but lead to either an increase or a reduction in expected lifetime utilities.[26] This implies that we are prevented from combining the first order conditions with respect to the various b_i and b_{ij} to attain the golden rule condition $r=n$. Formally, the effect of the reform would take the following form:

$$-\mu\left[\sum_i p_i - \frac{1+r(1-\tau_K)}{1+n}\sum_{ij} p_i p_{ij}\right] - \mu\left[-\frac{f'}{n+\delta} - \sum_i p_i + \frac{1+r(1-\tau_K)}{1+n}\sum_{ij} p_i p_{ij} + 1\right]$$
$$+self-selection\ term = 0,$$

where the self-selection term captures how the self-selection constraints faced by the government are affected by the circumstance that the reform changes the expected utility associated with some deviating strategies.[27] Due to the presence of a non-zero self-selection term in the above equation, the golden rule condition $r=n$ (which would make the first line of the equation equal to zero, as we have seen in Appendix B) cannot be a solution to the equation.[28]

Appendix D. The optimal interest tax rate under a nonlinear income tax.

To obtain an analytical expression for the optimal interest income tax rate one can adapt the procedure followed by Blomquist and Micheletto (2008). For this purpose, we should multiply by $r\left[s_i + \frac{s_i(j)}{1+r(1-\tau_K)}\right]$ the various first order conditions with respect to b_{ij} and then sum up the resulting set of equations with the first order condition with respect to τ_K. Notice

[26] An increase (resp.: reduction) would occur when the deviating strategy entails picking when middle-aged (resp.: young) a bundle intended for a young (resp.: middle-aged) worker.
[27] The fact that the expression is equal to zero descends also in this case from the assumption that we started from an optimum where all the first order conditions to the government's problem were satisfied.
[28] The self-selection term would only vanish if, at the solution to the government's problem under a nonlinear AID tax, there were no binding self-selection constraints where a worker of a given age is tempted to pick a bundle intended for a worker of a different age. Notice that in this special case the welfare gain of switching from a nonlinear AID to a nonlinear AD tax would be nil.

that the policy experiment of marginally increasing τ_K, while at the same time raising all the various b_{ij} by $r\left[s_i + \frac{s_i(j)}{1+r(1-\tau_K)}\right]$, leaves unaffected the expected utility of all non-deviating agents. This is however not the case for mimickers. The reason is that the change in the various b_{ij} is based on the savings behavior of non-deviating agents, since it is tailored to keep unchanged their expected utility. However, deviating agents will in general save to a different extent than non-deviating agents, and therefore the adjustment in the various b_{ij} will in general change their expected utility. Thus, the reform affects both the expected utility of mimickers and the resource constraint faced by the government. Moreover, starting from an optimum where all the first order conditions to the government's problem are satisfied, the sum of the effects of the reform on the set of self-selection constraints and on the resource constraint has to be equal to zero. The resulting equation can be used to provide an implicit characterization for the optimal level of the interest income tax rate. After some manipulations, and denoting by $\lambda_{i(\sigma^i)}$ the Lagrange multiplier associated to the self-selection constraint requiring a young type i worker not to engage in the deviating strategy σ^i, one obtains the following formula, where a tilde has been used to denote Hicksian demands:[29]

$$\frac{\mu}{r}\left[\frac{f'}{n+\delta}-\frac{1+r(1-\tau_K)}{1+n}\right]\left(\frac{\partial\tilde{s}}{\partial\tau_K}+\frac{1}{1+n}\frac{\partial\tilde{s}_{-1}}{\partial\tau_K}\right)=\sum_i\sum_{\sigma^i}\lambda_{i(\sigma^i)}\sum_{ij}\frac{\partial V_i(\sigma^i)}{\partial b_{ij}}\left[s_i-s_{\sigma^i}+\frac{s_i(j)-s_{\sigma^i}(j)}{1+r(1-\tau_K)}\right].$$

Given that under a nonlinear AD tax we have $r=n \Rightarrow f'=n+\delta$ and that the same condition holds under a nonlinear AID tax when debt policy is available, in these cases the condition implicitly characterizing the optimal interest income tax rate can be simplified to:

$$\mu\frac{\tau_K}{1+n}\left(\frac{\partial\tilde{s}}{\partial\tau_K}+\frac{1}{1+n}\frac{\partial\tilde{s}_{-1}}{\partial\tau_K}\right)=\sum_i\sum_{\sigma^i}\lambda_{i(\sigma^i)}\sum_{ij}\frac{\partial V_i(\sigma^i)}{\partial b_{ij}}\left[s_i-s_{\sigma^i}+\frac{s_i(j)-s_{\sigma^i}(j)}{1+r(1-\tau_K)}\right].$$

From the condition above it is apparent that, in the presence of a nonlinear tax on labor income, the value of the optimal interest income tax rate crucially hinges on the difference between the savings behavior of a deviating- and a non-deviating agent, for all the self-selection constraints that are binding at a solution to the government's problem. Given that a switch from a nonlinear AID to a nonlinear AD tax is likely to produce significant changes in

[29] The formula can be viewed as a generalization of eq. (9) in Blomquist and Micheletto (2008).

WORKING PAPERS
Uppsala Center for Fiscal Studies
Editor: Håkan Selin

2009:1	Sören Blomquist and Håkan Selin, Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates. 31 pp.
2009:2	Luca Micheletto, Optimal nonlinear redistributive taxation and public good provision in an economy with Veblen effects. 26 pp.
2009:3	Håkan Selin, The Rise in Female Employment and the Role of Tax Incentives. An Empirical Analysis of the Swedish Individual Tax Reform of 1971. 38 pp.
2009:4	Håkan Selin, Marginal tax rates and tax-favoured pension savings of the self-employed Evidence from Sweden. 32 pp.
2009:5	Tobias Lindhe and Jan Södersten, Dividend taxation, share repurchases and the equity trap. 27 pp.
2009:6	Che-Yan Liang, Nonparametric Structural Estimation of Labor Supply in the Presence of Censoring. 48 pp.
2009:7	Sören Blomquist, Vidar Christiansen and Luca Micheletto, Public Provision of Private Goods and Nondistortionary Marginal Tax Rates: Some further Results. 42 pp.
2009:8	Laurent Simula and Alain Trannoy, Optimal Income Tax under the Threat of Migration by Top-Income Earners. 26 pp.
2009:9	Laurent Simula and Alain Trannoy, Shall We Keep Highly Skilled at Home? The Optimal Income Tax Perspective. 26 pp.
2009:10	Michael Neugart and Henry Ohlsson, Economic incentives and the timing of births: Evidence from the German parental benefit reform 2007, 21 pp.
2009:11	Laurent Simula, Optimal Nonlinear Income Tax and Nonlinear Pricing: Optimality Conditions and Comparative Static Properties, 25 pp.
2009:12	Ali Sina Onder and Herwig Schlunk, Elderly Migration, State Taxes, and What They Reveal, 26 pp.
2009:13	Ohlsson, Henry, The legacy of the Swedish gift and inheritance tax, 1884-2004, 26 pp.
2009:14	Onder, Ali Sina, Capital Tax Competition When Monetary Competition is Present, 29 pp.

2010:1 Sören Blomquist and Laurent Simula, Marginal Deadweight Loss when the Income Tax is Nonlinear. 21 pp.

2010:2 Marcus Eliason and Henry Ohlsson, Timing of death and the repeal of the Swedish inheritance tax. 29 pp.

2010:3 Mikael Elinder, Oscar Erixson and Henry Ohlsson, The Effect of Inheritance Receipt on Labor and Capital Income: Evidence from Swedish Panel Data. 28 pp.

2010:4 Jan Södersten and Tobias Lindhe, The Norwegian Shareholder Tax Reconsidered. 21 pp.

2010:5 Anna Persson and Ulrika Vikman, Dynamic effects of mandatory activation of welfare participants. 37 pp.

2010:6 Ulrika Vikman, Does Providing Childcare to Unemployed Affect Unemployment Duration? 43 pp.

2010:7 Per Engström, Bling Bling Taxation and the Fiscal Virtues of Hip Hop. 12 pp.

2010:8 Niclas Berggren and Mikael Elinder, Is tolerance good or bad for growth? 34 pp.

2010:9 Heléne Lundqvist, Matz Dahlberg and Eva Mörk, Stimulating Local Public Employment: Do General Grants Work? 35pp.

2010:10 Marcus Jacob and Martin Jacob, Taxation, Dividends, and Share Repurchases: Taking Evidence Global. 42 pp.

2010:11 Thomas Aronsson and Sören Blomquist, The Standard Deviation of Life-Length, Retirement Incentives, and Optimal Pension Design. 24 pp.

2010:12 Spencer Bastani, Sören Blomquist and Luca Micheletto, The Welfare Gains of Age Related Optimal Income Taxation. 48 pp.