Abstract

Using a calibrated overlapping-generations model we quantify the welfare gains of an age-dependent labor income tax. Agents face uncertainty regarding future abilities and can transfer consumption across periods through savings. The welfare gain of switching from an age-independent to an age-dependent nonlinear tax varies between 2.4% and 4% of GDP. Part of the welfare gain is due to capital accumulation effects and part descends from relaxing incentive-compatibility constraints. The welfare gain is of about the same magnitude as the welfare gain that can be achieved by moving from a linear- to a nonlinear labor income tax. Finally, the welfare loss from tax-exempting interest income is negligible under an optimal age-dependent labor income tax.

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

JEL Classification: H21; H23; H24.
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) 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, age is a non-manipulable individual characteristic and the distribution of wage rates differs by age, the average wage being higher for older cohorts and the dispersion of wages wider. Hence, age represents a potentially relevant candidate to be used for tagging purposes.\(^1\) Moreover, as compared to other tagging schemes which have been considered in the literature, such as tagging by gender or individuals’ height,\(^2\) tagging by age is likely to be politically less controversial.\(^3\) The reason is that it seems more coherent with the horizontal equity principle which, albeit not easy to formulate in an unambiguous way, appears as an important constraint when reasoning about the practical implementability of a tax reform proposal (see Mankiw and Weinzierl (2010)).\(^4\)

Age-dependent income taxation has been analyzed in a few earlier papers.\(^5\) Kremer (2002) is an interesting early contribution conjecturing that marginal tax rates should be lower for young workers. However, he adopts a static model and abstracts from the individuals’ savings decisions. Dynamic models of age-dependent taxes are considered by Erosa and Gervais...

---

\(^1\)Indeed, in one of the chapters of the recently published Mirrlees Review, Banks and Diamond (2010) argue that indexing the income tax schedule on the age of the taxpayer represents a very promising avenue to improve the tax system.

\(^2\)Gender-based tax schemes have been studied by many authors, starting with the seminal contribution by Boskin and Sheshinski (1983). A recent example is Alesina et al. (2011). Tagging by height has been analyzed in a provocative article by Mankiw and Weinzierl (2010).

\(^3\)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.

\(^4\)Under an age-dependent income tax all agents face during their lifetime the different income tax schedules offered to the different age groups. With a gender-based income tax or a height-based income tax, instead, agents are assigned once forever to one of the different income tax schedules.

\(^5\)Even though it was not explicitly about age-dependent income taxation, Diamond and Mirrlees (1978) can be regarded as the earliest contribution related to this topic. Their paper studied optimal social insurance policies in a dynamic setting with elastic labor supply along the extensive margin.
(2002) and Lozachmeur (2006). However, working within a representative agent framework, these papers cannot not address income redistribution issues. The first paper to consider age-dependent nonlinear taxation in a dynamic Mirrleesian setting with heterogeneous agents and private savings is the Blomquist and Micheletto paper (2003, 2008). Using an overlapping generations (OLG) model where agents face a stochastic wage process, they characterize the optimal marginal labor income tax rates and the optimal proportional tax on the return on savings. They also show that a strict Pareto improvement can be obtained by moving from an age-independent nonlinear labor income tax to an age-dependent labor income tax. However, the offered results are only of qualitative nature.

Developing the analysis contained in Blomquist and Micheletto (2008), the main purpose of this paper is to explore the quantitative policy prescriptions descending from the implementation of an optimal age-dependent nonlinear labor income tax and to quantify its potential welfare gains.

The most related contribution is a paper recently published by Weinzierl (2011), who provides the first quantitative assessment of the welfare gains from age-dependent nonlinear income taxes. However, as we will clarify below, there are at least three distinctive features between our model and the one by Weinzierl. A first difference is that we explicitly model the production side of the economy, and this allows us to consider the interaction between taxes and the capital-accumulation process. Second, whereas Weinzierl focuses exclusively on nonlinear tax schemes, we consider the effects of tagging by age in the context of both linear- and nonlinear taxation schemes. Finally, we also allow the government to optimally choose a linear tax on the return on savings. This enables us to investigate how its optimal value changes under different assumptions about the structure of the labor income tax schedule and how large are the losses associated with tax-exempting interest income.

Two other related contributions are those by Conesa et al. (2009) and by Fukushima (2010). Conesa et al. (2009) quantitatively characterize the optimal proportional capital and nonlinear labor income tax in an OLG model with idiosyncratic, uninsurable income shocks and permanent productivity differences of households. Restricting their attention to three parameters income tax functions à la Gouveia and Strauss (1994), they show that the optimal tax system features a substantially higher welfare gain as compared to the current US tax system. Even though they do not consider age-dependent taxes, they emphasize the role played by the capital income tax as an imperfect substitute for age-dependent labor income taxes and find substantial welfare losses associated with tax-exempting interest income. Fukushima (2010) extends the Conesa et al. (2009) analysis by allowing the government to levy nonlinear income taxes that are arbitrarily age- and history-dependent. He finds that the restriction imposed by Conesa et al. (2009) on the shape of the admissible tax functions entails a substantial welfare loss as
compared to the optimal age- and history-dependent tax schedule.

In terms of degree of sophistication, the armory of tax instruments that we consider in this paper are intermediate between those used by Conesa et al. (2009) and those used by Fukushima (2010). In particular, and contrary to Fukushima (2010), we rule out the possibility of levying fully history-dependent income taxes. These would require that, at any given age, the tax liability that applies on a given level of earned income depends on the entire taxpayer’s past income history. This is a much stronger informational requirement than that implicit in an age-dependent income tax, where the tax liability that applies on a given level of income is a function of only two arguments, i.e. the level of income plus the taxpayer’s age. Albeit the so called new dynamic public finance literature initiated by Golosov et al. (2003) has shown that, in dynamic economies where agents’ skills may change over time, optimal taxation requires in general history-dependent taxes, we see two reasons to narrow the attention to pure age-dependent income taxes. The first is that, as suggested by Diamond and Saez (2011), the structure of an optimal fully history-dependent tax arrangement is likely to be too complex to be relevant for actual public policies. Thus, although analysis of history-dependent taxes are of interest, we believe that such tax systems are far off into the future. The second is that recent contributions by Fahri and Werning (2011), Golosov et al. (2011) and Weinzierl (2011) seems to suggest that age-dependent taxes can capture a substantial share of the total welfare gains achievable through complex history-dependent tax instruments.

The potential gains of an age-dependent nonlinear labor income tax as compared to an age-independent one are twofold. First, by shutting down some mimicking strategies, it reduces the number of incentive-compatibility constraints faced by the government. This is the standard benefit of tagging analyzed in the context of age by Blomquist and Micheletto (2008) and Weinzierl (2011). Second, in a general equilibrium environment with endogenous capital accumulation and restrictions on the management of public debt policy, only an age-dependent tax can always achieve the golden-rule level of capital without interfering with the redistributive goals of the government.

---

6The new dynamic public finance literature has been rapidly growing in the past few years. It is here impossible to offer a comprehensive summary of the various contributions belonging to this literature. A useful survey article is provided by Golosov et al. (2006) and a more exhaustive account by Kocherlakota (2010).

7Fahri and Werning (2011) consider a dynamic Mirrleesian environment in a life-cycle context with idiosyncratic shocks and study the optimal insurance arrangement. Assuming a utilitarian social planner, they show that age-dependent linear income taxes entail an almost negligible welfare loss (around 0.15% of lifetime consumption) as compared to the fully optimal history-dependent tax system. Using a different stochastic process for skills and assuming different individuals’ preferences, Golosov et al. (2011) reach similar conclusions but point out that the welfare loss from age-dependent linear income taxes is increasing in the degree of social aversion to inequality. Weinzierl (2011), also using a utilitarian planner, shows that age-dependent nonlinear labor income taxes capture between two-thirds and 89% of the gains from the full history-dependent optimum.
Since neither Blomquist and Micheletto (2008) nor Weinzierl (2011) model the production side of the economy, they cannot capture this additional benefit.\footnote{This remark also applies to the papers by Fahri and Werning (2011) and Golosov et al. (2011).}

In this paper, instead, the effects of taxes on the capital-accumulation process are explicitly taken into account and we can then decompose the overall welfare gains from an age-dependent nonlinear income tax into incentive-compatibility-effects and capital-accumulation-effects. In a realistically calibrated model based on US data, we show that in general an optimal age-independent tax does not lead the economy to the golden-rule, unless the government is allowed to run a sizable negative debt, implying that it (indirectly) owns a large fraction of the total capital stock. As we will explain later in more details, the reason is that relying on private savings to achieve the golden-rule requires a shift in the tax burden from the young to the middle-aged that is going to exacerbate some of the binding incentive-compatibility constraints faced by the government in the design of a nonlinear income tax which redistributes from those agents who are relatively well-off on a lifetime basis to those who are less well-off. On the other hand, an optimal nonlinear age-dependent labor income tax always leads the economy to the golden rule, irrespective of the redistributive objectives pursued by the government and independently on the assumptions on the availability of public debt policy. This represents another mechanism by which an age-dependent nonlinear income tax relaxes the efficiency-equity trade-off as compared to an age-independent income tax. Since negative public debt is not something that we observe in real-world economies, where governments are usually net borrowers towards the private sector and the stock of public debt is sometimes quite high, to quantitatively capture the importance of this second gain of age-dependent taxes we consider age-independent income taxes under two different scenarios. One in which public debt policy is unconstrained so that the optimal age-independent tax achieves the golden-rule condition, and one in which public debt is constrained to be non-negative, implying that the economy does not reach the golden-rule level of capital under the optimal age-independent tax.

As mentioned above, one of the advantages of a nonlinear age-dependent tax is the reduction in the number of incentive-compatibility 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. In particular, we consider a three-period OLG economy with heterogeneous workers in which individuals work for two periods and then retire. 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. We let agents’ skills evolve over time and, in our baseline scenario, we assume that agents face uncertainty about their future market ability. Several forces determine the shape of the optimal tax schedules. Of these, a crucial role is played by the per-period wage distributions as well as by the probabilities that relate wages in the two working periods. As a benchmark for our numerical simulations, we calibrate our model to US wage data. However, in order to highlight how results are affected by the structure of the wage process, we also perform sensitivity analysis based on Swedish wage data. Other sensitivity experiments include checking for the impact of different degree of uncertainty, of different assumptions about the degree of social aversion to inequality, and of different values for the coefficient of relative risk aversion and of the Frisch elasticity.

Dealing with nonlinear income taxes in a multi-period setting, 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 nonlinear age-independent tax schedule is still difficult to compute as it admits complex mimicking behaviors, giving rise to a very large number of incentive constraints. The computational difficulties are also the reason why we do not analyze the transition from an age-independent to an age-dependent tax, but focus instead on steady-states.

Our main results can be summarized as follows. For a max-min social welfare function, a move from a nonlinear age-independent tax to a nonlinear age-dependent tax yields an overall welfare gain of about 4% of GDP. Half of the gain comes from relaxing the incentive constraints faced by the government.

Taking a less extreme social welfare function, as for instance a utilitarian, lowers the overall welfare gain to about 3.3% of GDP and slightly increases the relative importance of capital-accumulation effects as compared to incentive-effects.

Replacing US wage data with Swedish data lowers the overall welfare gain under a max-min social welfare function from 4% to 3% of GDP. The reduction in the magnitude of the welfare gain is entirely attributable to a diminished gain from slackening incentive constraints. This is coherent with the idea that a higher wage inequality tightens the incentive-compatibility constraints faced by the government in the design of a nonlinear income tax. As wage inequality is lower in Sweden as compared to US, the gain from age-dependency due to relaxing incentive constraints becomes smaller.

Retaining the benchmark per-period wage distributions but assuming that agents face no uncertainty about their future wage rate does not affect

---

9We choose Sweden for comparative purposes since US and Sweden are two countries that sometimes are regarded as two extremes in terms of wage dispersion.
the magnitude of the overall welfare gain obtained by making the nonlinear income tax age-dependent. However, the share of the gain due to capital-accumulation effects becomes more important, accounting for almost three-fourths of the overall welfare gain. The intuition for the result is twofold. On one hand, in the absence of uncertainty agents no longer save for precautionary motives, and this makes it harder for the economy to reach the golden-rule without the government issuing negative debt. On the other hand, removing uncertainty reduces the number of incentive-compatibility constraints faced by the government in the design of the income tax, both under an age-dependent and under an age-independent tax. This tends to mechanically weaken the gains of age-dependency as a mimicking-deterring device.

We also find that the gains of using a nonlinear labor income tax instead of a linear tax are roughly of the same order of magnitude as those due to tagging by age (at least in the baseline scenario where agents face uncertainty).

Finally, setting the proportional interest income tax at the optimal rate is of second order importance under an optimal age-dependent labor income tax. Under an age-independent labor income tax, on the other hand, the welfare loss due to tax-exempting interest income can be sizeable, especially when the labor income tax is linear and public debt is constrained to be non-negative. For our benchmark specification, we find a welfare loss of about 2% of GDP under a nonlinear age-independent labor income tax, and of 3% under a linear age-independent tax, in both cases assuming that public debt is restricted to be non-negative.

The paper is organized as follows. In section 2 we present the basic structure of our model. In section 3 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 4 we describe how we calibrate our model. In section 5 we present the results of our simulations. Finally, section 6 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 $\theta_i$, with $i \in \{1, \ldots, m^y\}$, and the proportion of young agents of type $i$ by $p_i$, with $\sum_i p_i = 1$. With probability $p_{ij}$ an agent who is
of productivity $i$ in the first period of his work-life (when young) has productivity $\theta_{ij}$, 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 $\theta_\sigma$ as referring to the productivity of a middle-aged agent if $\sigma$ is a tuple $(i, j) \in I \times J$, and as referring to the productivity of a young agent if $\sigma$ 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 at time $t$ 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 in efficiency units and consumption are defined as

$$L_t = \sum_i p_i \theta_i \ell_{i,t} + \frac{1}{1 + n} \sum_{ij} p_i p_{ij} \theta_{ij} \ell_{ij,t}$$

and

$$C_t = \sum_i p_i c_{i,t} + \frac{1}{1 + n} \sum_{ij} p_i p_{ij} \left( c_{ij,t} + \frac{1}{1 + n} c_{ij,t}^R \right),$$

whereas total effective labor and consumption in the economy is given by $L_t = N_0(1 + n)^t \cdot \bar{L}_t$ and $C_t = N_0(1 + n)^t \cdot \bar{C}_t$ 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 $\delta$, $r_t$ and $\omega_t$. 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 $F(K_t, L_t) = L_t f(k_t)$, 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: $y_{it} = \omega_t \theta_i \ell_{i,t}$ and $y_{ij,t} = \omega_t \theta_{ij} \ell_{ij,t}$. We consider labor income tax systems of the form $\{T_1(y_i), T_2(y_{ij})\}$, where $T_1(\cdot)$ is the tax function that applies to young agents and $T_2(\cdot)$ the tax function for middle-aged agents. 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 (these would have the more general non-separable form $T(y_i, y_{ij})$). We maintain this notation throughout the paper keeping in mind that under an age-dependent tax we have $T_1(\cdot) \neq T_2(\cdot)$, whereas under an age-independent tax we have $T_1(\cdot) = T_2(\cdot)$. 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 the true beneficiary of the savings is. This implies that the returns on savings can only be
taxed at a proportional rate $\tau_K$. 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:

$$V_{i,t} = u(c_{i,t}, \ell_{i,t}) + \sum_j p_{ij} \left[ \beta u(c_{ij,t+1}, \ell_{ij,t+1}) + \beta^2 u(c_{ij,t+2}^R) \right],$$

which can be written

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

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 makes plans for second period savings, these plans are contingent on his productivity type 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:

$$D_t + I_t = N_t s_t + N_{t-1} s_{t-1}.$$  \hspace{1cm} (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 $t$:

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

Our assumptions on tax instruments provide for a version of the so-called dual income tax system introduced in the Nordic countries (Finland, Sweden, Norway and Iceland, as well as Denmark where a diluted version of the system was applied) in the 1990s. One argument for adopting a dual income tax system was to prevent arbitrage opportunities: 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. Notice also that our assumption on the observability of savings 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.
Dividing (2) by $L_{t+1} = N_{t+1}\tilde{L}_{t+1}$ gives:

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

and simplifying:

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

where we have defined:

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

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:

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

Dividing (3) by $L_t$ gives:

$$k_{t+1} \frac{N_{t+1}\tilde{L}_{t+1}}{N_tL_t} = (1 - \delta)k_t + f(k_t) - \frac{C_t}{L_t}, \quad (4)$$

which can be simplified to obtain:

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

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 government’s budget constraint:

$$\sum_i p_i T_{1,t}(w_{i,t},\ell_{i,t}) + \frac{1}{1 + n} \left( \sum_{ij} p_i p_{ij} T_{2,t}(w_{ij,t},\ell_{ij,t}) \right) + \frac{\tau K,\rho_t}{1 + n} \left( \sum_i p_i s_{i,t-1} + \sum_{ij} p_{ij}s_{ij,t-1} \right) + (1 + n)d_{t+1} = (1 + r_t)d_t. \quad (6)$$

We can now provide the following definitions:
Definition 1 (Competitive Equilibrium) Given a tax policy 
\[ T_{1,t}(y_i), T_{2,t}(y_{ij}), \tau_{K,t} \] \( t = 0 \), a competitive equilibrium is a sequence of prices \( \{r_t, \omega_t\} \), individual allocations \( \{c_{i,t}, c_{ij,t}, \ell_{i,t}, \ell_{ij,t}, s_{i,t}, s_{ij,t}\} \), production plans \( \{K_t, L_t\} \), and government debt \( \{D_t\} \) 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 a competitive equilibrium with the property that, for all periods \( t \geq t_{ss} \),
\[ \{T_{1,t}(y_i), T_{2,t}(y_{ij}), \tau_{K,t}\} = \{T_{1}(y_i), T_{2}(y_{ij}), \tau_{K}\}, \]
\[ \{r_t, \omega_t\} = \{r, \omega\}, d_t = d, \{\tilde{C}_t, \tilde{L}_t\} = \{\tilde{C}, \tilde{L}\} \quad \text{and} \quad k_t = k. \]

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 focus on steady states under alternative assumptions about the properties of the labor income tax schedule and about the availability of debt policy. Apart from the assumptions on the armory of instruments the policy-maker has access to, the features of the specific steady state reached by the economy depend also, in general, on the objective function maximized by the government. Our assumption will be that it maximizes the steady state expected lifetime utility of a typical cohort, evaluated according to an inequality-averse social welfare function. In the numerical simulations we mainly focus on the case where the social welfare function is of the max-min type but we also consider, for sensitivity analysis purposes, the case of a utilitarian social welfare function. Irrespective of the degree of social aversion to inequality, if the government has access to unrestricted public debt policy, it will achieve a steady-state where the golden-rule condition for capital accumulation is satisfied: \( f'(k) = n + \delta \). However, if the government faces restrictions in the use of debt policy, the steady-state equilibrium will not necessarily satisfy the golden-rule condition. The reason is that, if public debt cannot be freely adjusted to meet the golden-rule condition, there might be a tension in the design of the optimal income tax between the pursuit of the equity objectives embedded in the social welfare function, on one hand, and the pursuit of the golden-rule condition, which purely reflects efficiency considerations, on the other hand. In general, this tension manifests itself when the income tax is age-independent. Given the structure of the economy and the evolution of skills over the individuals’ life-cycle that we consider in our numerical simulations, and given the redistributive goals pursued by the government, the steady state of the economy does not reach the golden-rule under an age-independent tax unless the government runs a sizable negative debt, implying
that it (indirectly) owns a large fraction of the total capital stock. For this reason, to capture the magnitude of the welfare gains of an age-dependent tax due to capital accumulation effects, we consider two solutions to the government’s problem under an age-independent tax: one in which public debt policy is fully unrestricted and one in which public debt is restricted to be non-negative. This seems to us a reasonable approach given that in most real-world economies the public sector is a net borrower towards the private sector and the stock of public debt is sometimes quite high.

We are now ready to set up the government’s problem for the various tax systems that will be considered in the numerical simulations.

3 Tax systems and government’s problems

3.1 Age-dependent nonlinear income tax

We consider a discrete adaptation of the Mirrlees (1971) optimum income taxation model. The government knows the skill distribution at each age and the Markov probabilities relating these distributions. Moreover, it can observe pre-tax incomes \( y = wθℓ \), whereas neither individual skills \( θ \) nor labor supplies \( ℓ \) are publicly observable. This prevents the government from levying personalized lump-sum taxes/transfers. Instead, to pursue its redistributive goals the government has at its disposal a nonlinear labor income tax schedule which can be conditioned on the age of the taxpayer (assumed to be observable). Thus, \( T_1(y_i) \) is the nonlinear income tax applying to young agents and \( T_2(y_{ij}) \) is the one applying to middle-aged agents. Another instrument at disposal of the government is a proportional tax on the return on savings, \( τ_K \). 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 8, 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 agents choose a strategy. A strategy specifies which income point is chosen by an agent in each period of work and for each state of the world (in our case, 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 an agent \( i \) is a plan \( σ^i = (σ^i_1, σ^i_2) \) where \( σ^i_1 \in I \) is the reported type when young and \( σ^i_2 \) is a functional \( σ^i_2 :
determining the income point chosen when middle-aged as a function of the second period skill realization \( j \in J \). The set of all strategies available to agent \( i \) is denoted \( \Gamma_i \). 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 available deviating strategy. The fact that agents are allowed to save/borrow doesn’t affect the total number of incentive constraints but enables 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/borrow. In fact, if agents were unable to save/borrow, the intertemporal allocation would be completely determined by the income points chosen when young and middle-aged. Hence, with anonymous savings, labor income is in any period a weaker signal of ability and less redistribution can be achieved.

Given our assumption that the return on savings can only be taxed at a proportional rate \( \tau_K \), the intertemporal distortion will be the same both for truthful- and mimicking agents. A positive tax on the return on savings becomes desirable when mimickers tend to value future consumption more than truth-telling agents.

Even under an age-dependent (hereafter, AD) nonlinear income tax, the set of incentive constraints faced by the government is quite large, especially under the assumption that agents face uncertainty with respect to their second-period skill. In fact, in the second period each worker can choose among all income points on the tax schedule for middle-aged agents, irrespective of the income point which was chosen in the first period on the tax schedule for young agents.\(^{11}\) Moreover, uncertainty about second-period skill increases substantially the number of strategies which can be adopted by an agent. In the absence of uncertainty, and for any given income point chosen in the first period, each agent has available as many strategies as the number of income points offered by the government on the income tax schedule for the middle-aged. Under uncertainty about second-period skill, on the other hand, and for any given income point chosen in the first period, each agent has available many more strategies than the number of income points offered by the government on the tax schedule for the middle-aged: this is because

\(^{11}\) This stands in contrast with the case where the government uses history-dependent taxes. In that case the set of income points available for an agent in the second period depends on the income point chosen in the first period.
a strategy must specify which income point is chosen in the second period for each possible skill realization. Assume for instance that each agent has the same number of different possible skill realizations in the second period and denote it by $\varphi$. With $m^y$ possible ability types when young, we show in Appendix A.1 that the total number of strategies $\Gamma = \bigcup_{i \in I} \Gamma_i$, including the strategies entailing truthful revelation, is given by:

$$|\Gamma| = m^y [m^y \Psi \varphi] = (m^y)^2 \Psi \varphi,$$

where $\Psi \equiv \sum_{i=1}^{m^y} \varphi = m^y \varphi$.

To illustrate the formula above, suppose that $m^y = 3$ and that each type of young agents can be one of two possible types in the second period, meaning that $\varphi = 2$. In this case $|\Gamma| = 324$ and only three of these strategies entail truthful revelation (one for each skill type in the first period).

We are now in a position to formally state the government’s problem. As an objective function we choose a concave social welfare aggregator $W(\cdot)$ whose arguments are the steady state expected lifetime utilities of the $m^y$ types of agents forming a representative cohort. The government’s problem is therefore:

$$\max_{d, \{b_i, y_i\}_{i \in I}, \{b_{ij}, y_{ij}\}_{i \in I, j \in J, \tau} \in \mathcal{K}} W \left( V_1(\tilde{\sigma}^1), \ldots, V_{m^y}(\tilde{\sigma}^{m^y}) \right)$$

subject to:

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

$$\sum_i p_i (y_i - b_i) + \frac{1}{1+n} \sum_{ij} p_i p_{ij} (y_{ij} - b_{ij}) +$$

$$\frac{\tau \kappa r}{1+n} \left( \sum_i p_i s_i + \frac{\sum_{ij} p_i p_{ij} s_{ij}}{1+n} \right) + (1+n) d = (1 + r) d$$

and

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

where we have defined:

$$\tilde{L} = \left( \sum_i p_i (y_i/\omega) + \frac{1}{1+n} \sum_{ij} p_i p_{ij} (y_{ij}/\omega) \right),$$
\[ V_i(\sigma^i) \equiv u\left(b_{\sigma^i_i} - s_{\sigma^i_i}, y_{\sigma^i_i}/w_i\right) + \beta \sum_j p_{ij} \left\{ u\left(b_{\sigma^i(j)} + (1 + r(1 - \tau K))s_{\sigma^i_i} - s_{\sigma^i_i(j)}, y_{\sigma^i(j)}/w_{ij}\right) + \beta^2 u((1 + r(1 - \tau K))s_{\sigma^i(j)}) \right\} \]

and

\[ V_i(\tilde{\sigma}^i) \equiv u(b_i - s_i, y_i/w_i) + \beta \sum_j p_{ij} \left\{ u\left(b_{ij} + (1 + r(1 - \tau K))s_i - s_i(j), y_{ij}/w_{ij}\right) + \beta^2 u((1 + r(1 - \tau K))s_i(j)) \right\}. \]

In the last two expressions above \( s_{\sigma^i_i} \) denotes the savings chosen in the first period by a type \( i \) agent who adopts the strategy \( \sigma^i \), whereas \( s_{\sigma^i_i(j)} \) denotes the savings prescribed in the second period by this strategy if the agent’s skill realization when middle-aged is \( j \) (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 important to notice that under an optimal AD nonlinear tax the economy reaches a steady state where the golden-rule capital-labor ratio is achieved even if debt policy is unavailable. We show this formally in Appendix A.2. As a consequence, the above optimization problem admits a continuum of equivalent solutions since 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 level of savings of the young does not matter for self-selection purposes. More precisely, under an AD nonlinear tax one can always 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 present value lifetime tax payment of all agents is left unaffected, all young agents change their savings by the same amount, the public budget is kept 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 and without interfering with the redistributive goals of the government.

Before moving to the case of an age-independent nonlinear tax, a final remark is warranted about the way we have written the public budget constraint in the government’s problem above. In principle, there are two different ways to formulate this constraint, both of which are reasonable. One
possibility is to write the government’s budget constraint as a per-period constraint. This has been our approach as well as the approach taken in the classic contributions by Samuelson (1958) and Diamond (1965). The reason why we find this approach appealing is that we believe that it closely reflects the constraint perceived by real-world governments in their decision-making process. As an alternative, one could have formulated the government’s budget constraint as a generational budget constraint. This is the approach which prevails in the new dynamic public finance literature and is usually justified on the grounds that it limits the extent to which inter- versus intra-cohort redistributive motives are mixed. Nonetheless, it is our view that policy-relevance considerations make the per-period approach a case worth focusing on.

3.2 Age-independent nonlinear income tax

Under an age-independent (hereafter, AI) nonlinear income tax, the government faces a larger set of incentive constraints than under an AD nonlinear tax. The reason is that an AI tax leaves open the possibility for a worker to choose also among income points intended for people of different age-groups. A strategy corresponding to an agent $i$ is now a plan $\sigma^i = (\sigma^i_1, \sigma^i_2)$ where $\sigma^i_1 \in (I \times (J \cup I))$ is the reported type when young and $\sigma^i_2$ is a functional $\sigma^i_2 : J \rightarrow (I \times (J \cup I))$ determining the income point chosen when middle-aged as a function of the second period skill realization $j \in J$. Denoting by $\hat{\Gamma}_i$ the set of all strategies available to a young agent of type $i$, the set of incentive constraints faced by the government in the design of the income tax can be written as:

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

As shown in Appendix A.1, under the assumption that each agent has the same number, denoted by $\varphi$, of different possible skill realizations in the second period, the total number of possible strategies under an AI nonlinear tax is:

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

where $\Psi$ is still defined as $\Psi \equiv \sum_{i=1}^{m^y} \varphi = m^y \varphi$. To illustrate the difference with the case of an AD tax, suppose again that $m^y = 3$ and that each type

---

12It is also true that the model usually considered in the new dynamic public finance literature is a model with a single cohort living for $T$ periods. In such a setting, the generational budget constraint approach appears the only available option to write the government’s budget constraint.

13Notice also that the two approaches are equivalent when the golden-rule condition is satisfied.
of young agents can be one of two possible types in the second period, so that \( \varphi = 2 \). In this case \( |\hat{\Gamma}| = 2187 \), entailing a huge increase in the number of strategies as compared to the AD case.

In contrast to the case of the AD nonlinear tax, under an optimal AI tax the economy does not necessarily reach a steady state where the golden-rule is satisfied. Whether it does or not depends crucially on the assumptions about the availability of public debt policy. If debt policy is fully unrestricted, it can be used by the government to achieve the golden-rule while the tax policy can be chosen with the sole purpose of fulfilling the intra-generational redistributional objectives. However, if there are restrictions on the use of debt policy, the tax instruments are burdened with the pursuit of two objectives, achieving the golden-rule and redistributing among agents of a given cohort, which are not necessarily coherent. This is formally shown in Appendix A.3. As our simulations in section 5 show, public debt must be large and negative for the golden-rule to be satisfied at the solution to the government’s problem under an AI tax. If the government is prevented from running a sizable negative debt (owning a large fraction of the total capital stock), 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 an AI 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 present value lifetime tax payment of all agents is left unaffected, all young agents change their savings by the same amount, the public budget is kept balanced, and all the self-selection constraints continue to be satisfied. In this sense one can claim that, under an AI nonlinear tax, the absolute level of private savings matters for self-selection purposes (or, put differently, the level of individual savings becomes relevant in order to deter mimicking behaviors). Therefore, unless debt policy is fully unrestricted, the government cannot control the capital stock without affecting the self-selection constraints. This difference between an AD tax and an AI tax represents a second source of welfare gains from tagging by age.

### 3.3 Linear income taxes: age dependent and age independent

An interesting question is how the gains from tagging by age depend on the government’s ability to use general (nonlinear) tax schedules; moreover, it is also of interest to compare the welfare gains descending from tagging by age with those which can be obtained by moving from a linear labor income tax to a nonlinear tax. For these purposes, in our numerical simulations we also consider the case where the labor income tax is a linear affine function. Under an AD linear tax, labor income is taxed at the flat rates \( \tau_1 \) and \( \tau_2 \) in the first- and second-period respectively. In addition, there is a demigrant
which is paid to all young workers and a demogrant \( G_2 \) which is paid to all middle-aged workers. In the AI scenario we impose \( \tau_1 = \tau_2 \) and \( G_1 = G_2 \). Both in the AD and the AI linear tax systems we restrict the demogrants to be positive \( (G_i \geq 0 \; i=1,2) \). The reason for such a constraint is a concern for realism: a uniform lump-sum tax is very unlikely to be politically implementable.\(^{14}\) For later purposes it is also worth noticing that age-differentiated demogrants are redundant when the government has access to unrestricted public debt policy. In that case, the restriction that \( G_1 = G_2 \) is harmless from the social welfare perspective. Finally, we always assume that the return on savings can be taxed at a proportional rate \( \tau_K \).

Under a linear tax system the steady state private budget constraints are:

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

and the government’s problem can be formally stated as:

\[
\max_{d, \tau_1, \tau_2, G_1, G_2, \tau_K} W \left( V_1 (\tilde{\sigma}^1), \ldots, V_m (\tilde{\sigma}^m) \right)
\]
subject to the budget constraint:\(^{15}\)

\[
\sum_i p_i (\tau_{1,t} w_i, t \ell_i, t + \tau_{K,t} r_t s_i, t - 1) + \frac{1}{1 + n} \left( \sum_{ij} p_{ij} \left( \tau_{2,t} w_{ij}, t \ell_{ij}, t + \frac{\tau_{K,t} r_t s_{ij, t - 1}}{1 + n} \right) \right) + (1 + n) d_{t+1} = (1 + \tau_t) d_t + G_{1,t} + \frac{1}{1 + n} G_{2,t}
\]

and the capital market equilibrium condition:

\[
\tilde{L} (1 + n) k - (1 - \delta) k \tilde{L} = s + \frac{s-1}{1 + n} - d (1 + n).
\]

Having described the various tax systems that we consider in our numerical simulations, we are now ready to describe our calibration and computational approach.

---

\(^{14}\)The attempt made by Margaret Thatcher to introduce a poll tax in the UK can be regarded as an example.

\(^{15}\)The budget constraint for the linear tax case can be obtained by combining the resource constraint, the private budget constraints, the definitions of \( C_t \) and \( L_t \), and the capital market equilibrium condition.
4 Calibration and computational approach

4.1 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 $\kappa = 3$ (implying a Frisch elasticity of 0.5) and $\gamma=2$ as baseline values. For sensitivity analysis we also consider $\gamma = 0.9$ and $\gamma = 2.5$ as well as $\kappa = 4$ (implying a Frisch elasticity of 0.33) and $\kappa = 1.5$ (implying a Frisch elasticity of 2).\textsuperscript{16} The government’s exogenous revenue requirement is set to zero.

4.2 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 consists of three parts, the skill distributions in the first- and second periods, and the transition probabilities linking these distributions. For our baseline numerical simulations we calibrate the skill process on US data. To capture how agents’ skills evolve over the life-cycle, we look at how, within a given cohort, individuals’ wages change over time. However, since our model abstracts from productivity growth, we look at how agents change over time their relative position in their cohort-specific wage distribution. That is, in both the first- and second period each individual’s percentile position in the wage distribution is calculated. We then see how individuals move between percentiles in the wage distribution over time. In our model the length of a period is 20 years. Hence, we need a panel data set following individuals over a long time. For this purpose we use the National Longitudinal Study of Youth (NLSY79) published by the US Bureau of Labor.

\textsuperscript{16}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. See Keane (2011) for a recent survey.
Statistics. We focus on agents who were 25 years old during the period 1982-1988 and determine how they move in the wage distribution by calculating their wages 20 years later during the period 2002-2008 (when they are 45).\textsuperscript{17} Since the number of self-selection constraints severely limits the number of types we can handle in our numerical simulations, we divide the wage distribution for the young into three groups and the wage distribution for the middle-aged into four groups. Showing the proportions that move from one group to another is the most general description of the underlying process. Any other description imposes functional form assumptions. To keep the model manageable we also restrict some of the transition probabilities by assuming $p_{ij} > 0$ for $0 \leq j - i \leq 1$.

The procedure outlined above, based on wage distributions taken from two separate periods of time, serves us to obtain the transition probabilities that we use in the simulations. However, the wage rates that we use in the simulations to describe the skill distribution for young and middle-aged agents are obtained following a different approach. In particular, we use for this purpose the wage distributions which \emph{in a given year} characterize young and middle-aged workers and calibrate these wage distributions using the 2003 wave of the Current Population Survey (CPS).\textsuperscript{18} The reason for doing so is the following. In a real-world economy the only possibility for mimicking is by choosing on the tax schedule an income point intended for someone else who is alive \emph{in the same period}. Because of that, if one wants to provide a realistic quantitative assessment of the power of age-dependency as a mimicking-deterring device, one has to derive the optimal tax function(s) based on the wage distributions for young and middle-aged agents who are active on the labor market at the same time. Although in principle we could have used the NLSY to calculate these wage distributions, we use the CPS to obtain a more accurate estimate and benefit from a larger sample size.

The wage rates and the associated transition matrix that we use in our benchmark simulations are given in Table 1.\textsuperscript{19}

\textsuperscript{17}The reason we have not focused on a single year is to smooth out potential year and cohort-specific shocks to an individual’s positions in the wage distribution.
\textsuperscript{18}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 estimating a Pareto distribution on the original data and then using the Pareto distribution to describe the wage distribution above the top code. The procedure is similar to the procedure outlined in Schmitt (2003) which is applied to the CEPR CPS extracts. The principle 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).
\textsuperscript{19}More specifically, the relative position of young agents is separated by the 33\textsuperscript{rd} and 66\textsuperscript{th} percentile and the relative position of middle-aged agents is determined by the 25\textsuperscript{th}, 50\textsuperscript{th} and 75\textsuperscript{th} percentile. With the restriction on the transition probabilities that we impose, this enables us to classify individuals as having one out of six possible wage paths.
Table 1: Transition matrix and hourly wages for the United States Economy.¹

<table>
<thead>
<tr>
<th></th>
<th>Middle-aged</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Wages</td>
</tr>
<tr>
<td>Young</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.41</td>
<td>0.59</td>
<td>0</td>
<td>0</td>
<td>8.19</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.43</td>
<td>0.57</td>
<td>0</td>
<td>12.38</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.59</td>
<td>0.41</td>
<td>19.58</td>
</tr>
<tr>
<td>Wages</td>
<td>9.67</td>
<td>15.42</td>
<td>21.40</td>
<td>34.75</td>
<td></td>
</tr>
</tbody>
</table>

¹ Wages expressed in 2003 US Dollars.

The numbers in the matrix can be given different interpretations. One interpretation, that is always correct, is that they show the proportion of agents belonging to a given group in the first period that move to a particular group in the second period. However, this interpretation is consistent with different assumptions about how individuals perceive their wage process. One possibility is that individuals perceive these proportions as probabilities. Thus, if a proportion $p_{ij}$ moves from group $i$ in first period to group $j$ in second period, the assumption would be that individuals who belong to group $i$ in first period regard $p_{ij}$ as the probability that they will end up in group $j$ in the second period. However, other assumptions can be made. For instance, one can assume that agents face no uncertainty so that a proportion $p_{ij}$ of those who belong to group $i$ in the first period know that they will be of type $j$ in the second period. Since the agents’ behavioral response to taxes varies depending on how they perceive their wage process, and this is likely to have an impact on the welfare gains descending from different tax schemes, we retain as a benchmark for our simulations the assumption that agents face uncertainty but, for sensitivity purposes, we also consider the case where young agents know their skill type in the second period.

To assess the sensitivity of the results to the structure of the underlying wage process, we compare the outcomes from our benchmark scenario with those generated under a different wage process. For this purpose, and following a procedure similar to the one outlined above for the US case, we construct a wage process which is based on Swedish data.²⁰

Finally, to have a perspective on the importance of the welfare gains from age-dependency, we compare them with those achievable by moving from a linear to a nonlinear labor income tax. Given the many suggestions for a “flat tax” this is in itself an interesting and policy relevant issue.

²⁰The details of the procedure, as well as the table summarizing the wage process for Sweden, are provided in Appendix B.
4.3 Computational approach

The optimal tax problems that we consider are constrained nonlinear optimization problems and have been solved using sophisticated interior-point/barrier methods. 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)).

5 Results

5.1 Social welfare function and welfare gains measure

As a benchmark for our simulations we use a social welfare aggregator of the max-min type and assume that the government aims at maximizing the minimum expected lifetime utility of a young worker. Thus, we take an ex-ante version of the max-min criterion. As an alternative, one could have taken an ex-post version of the criterion and, accordingly, let the government maximize the minimum realized lifetime utility of a worker at the end of life. Even though our focus is on the ex-ante version of the max-min criterion, we perform some sensitivity analysis to check how results are affected if one adopts the ex-post view. To assess the importance of the degree of social aversion to inequality, we also run the baseline version of our model for a purely utilitarian social welfare function.

To obtain a revenue-based measure of the welfare gains attainable by more sophisticated tax schemes, we consider an equivalent-variation-type of welfare gain measure, taking as a benchmark the solution to the government’s problem under an AI linear income tax with unrestricted public debt policy. More precisely, we proceed as follows. We first calculate the minimum amount of extra revenue that should be injected into the government’s budget, in the optimal AI linear tax problem with unrestricted debt policy, 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 AI linear tax optimum (with unre-

---

21The problem was set up in the modeling language AMPL and then solved using the nonlinear optimization solver KNITRO by Ziena Optimization Inc.
stricted public debt policy) to get a revenue-based measure of the welfare gains.

5.2 Welfare gains: an overview

In this section we provide an overview of the welfare gains obtained by an AD tax. We do this for the baseline parameterization of the utility function ($\gamma = 2$ and $\kappa = 3$). We present results both when debt policy is unrestricted and when public debt is restricted to be non-negative. The results are summarized in table 2.

Table 2: Overview of welfare gains.\(^1\)

<table>
<thead>
<tr>
<th>Instruments</th>
<th>max-min SWF(^2)</th>
<th>utilitarian SWF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear AI Benchmark</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Linear AI Restr. Debt</td>
<td>-2.07%</td>
<td>-2.99%</td>
</tr>
<tr>
<td>Linear AD</td>
<td>(\approx 0.01)%</td>
<td>(\approx 0.01)%</td>
</tr>
<tr>
<td>Nonlinear AI Restr. Debt</td>
<td>2.60%</td>
<td>-0.95%</td>
</tr>
<tr>
<td>Nonlinear AI</td>
<td>4.54%</td>
<td>1.15%</td>
</tr>
<tr>
<td>Nonlinear AD</td>
<td>6.65%</td>
<td>2.38%</td>
</tr>
<tr>
<td>Total gain of nonlinear age dependent taxes</td>
<td>4.05%</td>
<td>3.33%</td>
</tr>
<tr>
<td>Gain due to slackening IC constraints</td>
<td>2.11%</td>
<td>1.23%</td>
</tr>
<tr>
<td>Gain due to capital accumulation effects</td>
<td>1.94%</td>
<td>2.10%</td>
</tr>
</tbody>
</table>

\(^1\) expressed as percentage of GDP.
\(^2\) (ex-ante).

Looking at the welfare gains from age-dependency we can see that, for the linear tax case, they are virtually zero if the policy maker can freely use debt policy. On the other hand, under the constraint that public debt must be non-negative, the welfare gains of moving from the AI linear to the AD linear tax amount to about 2% of GDP for the max-min SWF and to about 3% for the utilitarian SWF. The reason is that the optimal linear AD tax allows the government to almost achieve the golden-rule even when public debt is constrained to be non-negative, whereas this is not the case for the linear AI tax. The reason why the AD tax might not fully achieve the golden-rule is that we have imposed the restriction that the demogrants must be nonnegative.

Irrespective on the constraints on the use of debt policy, the golden-rule condition for capital accumulation is always satisfied under an optimal AD nonlinear tax. This means that the gains of moving from a nonlinear AI tax to a nonlinear AD tax can be decomposed into capital-accumulation effects and mimicking-deterring effects, the latter being related to the fact that the AD tax mitigates the self-selection constraints faced by the government in the tax design problem. Under the restriction that public debt must be
non-negative, the total gain of moving from the optimal AI nonlinear tax to the optimal AD tax amounts to 4.05% of GDP (6.65%-2.60%) under a max-min SWF, and to 3.33% of GDP (2.38%-(-0.95%)) under a utilitarian SWF. Roughly half of the total gain comes from mimicking-deterring effects under a max-min SWF, whereas under a utilitarian SWF they account for a slightly lower share (37%). Thus, lowering the degree of social aversion to inequality appears to yield slightly lower welfare gains; however, the magnitude of this effect is dampened once the endogeneity of the capital-labor ratio is accounted for. In fact, neglecting the capital-accumulation effects would imply that welfare gain of tagging by age is reduced by 42% when using a utilitarian SWF instead of a max-min SWF; considering both the capital-accumulation effects and the mimicking-deterring effects implies that the welfare gain is reduced by only 18%.

As we can see from table 2, for the max-min SWF, the welfare gain of tagging by age under a nonlinear income tax is a bit lower than that which can be obtained by moving from a linear AI to a nonlinear AI tax. Under a utilitarian SWF, on the other hand, the welfare gain of making the nonlinear tax age-dependent is larger than that obtained by moving from a linear AI to a nonlinear AI tax. This seems to suggest that reducing the degree of social aversion to inequality lowers the magnitude of the welfare gains that can be achieved by using a nonlinear tax more than it lowers the welfare gains that can be reaped by making the nonlinear tax age-dependent.

We now move on to describe the characteristics of the optimal tax systems.

### 5.3 Nonlinear taxation

Tables 3 and 4 present the most relevant features of the optimal nonlinear AD and the optimal nonlinear AI labor income tax for, respectively, the case of an ex-ante max-min SWF and a utilitarian SWF. The values for the AI scenario are in the middle panel of tables 3 and 4 calculated under the assumption that the government is unrestricted in its debt policy, whereas in the bottom panel of these tables a non-negativity constraint on public debt has been imposed. Since debt policy is a redundant instrument under an optimal AD tax, the values for the AD tax are computed under the assumption that public debt is zero.

We can see that, as compared with the results for the nonlinear AI setting, the AD 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.

Under a max-min SWF all young workers, with the exception of the least skilled agents, increase their labor supply when moving from the AI to the nonlinear AD tax. Middle-aged workers, on the other hand, tend to reduce
Table 3: Results for the ex-ante max-min SWF

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.516</td>
<td>0.31</td>
<td>2.57</td>
<td>-6.99</td>
<td>0.70</td>
<td>0.34</td>
<td>3.26</td>
<td>0.31</td>
<td>0.74</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.492</td>
<td>0.41</td>
<td>5.05</td>
<td>-5.11</td>
<td>0.57</td>
<td>0.48</td>
<td>7.36</td>
<td>3.61</td>
<td>0.62</td>
</tr>
<tr>
<td>(2, 3)</td>
<td>-0.492</td>
<td>0.41</td>
<td>5.05</td>
<td>-5.11</td>
<td>0.57</td>
<td>0.48</td>
<td>7.36</td>
<td>3.61</td>
<td>0.62</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.444</td>
<td>0.50</td>
<td>9.75</td>
<td>-1.97</td>
<td>0.38</td>
<td>0.56</td>
<td>11.99</td>
<td>6.72</td>
<td>0.46</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.444</td>
<td>0.50</td>
<td>9.75</td>
<td>-1.97</td>
<td>0.38</td>
<td>0.56</td>
<td>11.99</td>
<td>6.72</td>
<td>0.46</td>
</tr>
</tbody>
</table>

$\omega = 1.00$  
$r = 0.24$  
$K = 8.38$  
$L = 14.33$

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.531</td>
<td>0.36</td>
<td>2.98</td>
<td>-3.56</td>
<td>0.58</td>
<td>0.32</td>
<td>3.12</td>
<td>-3.46</td>
<td>0.79</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.498</td>
<td>0.34</td>
<td>4.27</td>
<td>-2.57</td>
<td>0.69</td>
<td>0.52</td>
<td>8.09</td>
<td>0.24</td>
<td>0.56</td>
</tr>
<tr>
<td>(2, 3)</td>
<td>-0.498</td>
<td>0.34</td>
<td>4.27</td>
<td>-2.57</td>
<td>0.69</td>
<td>0.52</td>
<td>8.09</td>
<td>0.24</td>
<td>0.56</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.452</td>
<td>0.41</td>
<td>8.09</td>
<td>0.24</td>
<td>0.62</td>
<td>0.55</td>
<td>12.54</td>
<td>3.17</td>
<td>0.46</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.452</td>
<td>0.41</td>
<td>8.09</td>
<td>0.24</td>
<td>0.62</td>
<td>0.55</td>
<td>12.54</td>
<td>3.17</td>
<td>0.46</td>
</tr>
</tbody>
</table>

$\omega = 1.00$  
$r = 0.24$  
$K = 7.99$  
$L = 13.68$

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.545</td>
<td>0.36</td>
<td>2.53</td>
<td>-3.93</td>
<td>0.51</td>
<td>0.34</td>
<td>2.81</td>
<td>-3.72</td>
<td>0.73</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.511</td>
<td>0.34</td>
<td>3.62</td>
<td>-3.13</td>
<td>0.64</td>
<td>0.53</td>
<td>6.98</td>
<td>-0.79</td>
<td>0.45</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>-0.511</td>
<td>0.34</td>
<td>3.62</td>
<td>-3.13</td>
<td>0.64</td>
<td>0.53</td>
<td>6.98</td>
<td>-0.79</td>
<td>0.45</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.465</td>
<td>0.42</td>
<td>6.99</td>
<td>-0.78</td>
<td>0.55</td>
<td>0.58</td>
<td>10.63</td>
<td>1.44</td>
<td>0.38</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.465</td>
<td>0.42</td>
<td>6.99</td>
<td>-0.78</td>
<td>0.55</td>
<td>0.58</td>
<td>10.63</td>
<td>1.44</td>
<td>0.38</td>
</tr>
</tbody>
</table>

$\omega = 0.85$  
$r = 0.66$  
$K = 4.90$  
$L = 13.81$

AD  Welfare Gain = 6.65%  
$\tau_K = 0.11$

AI  Welfare Gain = 4.54%  
$\tau_K = 0.40, d = -2.43$

AI, Restr. Debt  Welfare Gain = 2.60%  
$\tau_K = 0.84, d = 0.00$
Table 4: Results for the utilitarian SWF

| Type | E(U)  | \( \ell_i \) | \( y_i \) | \( T(y_i) \) | \( T'(y_i) \) | \( \ell_{ij} \) | \( y_{ij} \) | \( T(y_{ij}) \) | \( T'(y_{ij}) \) |
|------|-------|--------------|---------|---------|---------|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| (1, 1) | 0.51 | 4.91 | 2.32 | 0.41 |
| (1, 2) | 0.60 | 9.24 | 4.96 | 0.26 |
| (2, 2) | 0.60 | 9.24 | 4.96 | 0.26 |
| (2, 3) | 0.58 | 12.52 | 6.34 | 0.29 |
| (3, 3) | 0.58 | 12.52 | 6.34 | 0.22 |
| (3, 4) | 0.61 | 21.36 | 11.01 | 0.06 |

\( \omega = 1.00 \) \( r = 0.24 \) \( K = 9.26 \) \( L = 15.82 \)

- AD Welfare Gain = 2.38% \( \tau_K = 0.13 \)

| Type | E(U)  | \( \ell_i \) | \( y_i \) | \( T(y_i) \) | \( T'(y_i) \) | \( \ell_{ij} \) | \( y_{ij} \) | \( T(y_{ij}) \) | \( T'(y_{ij}) \) |
|------|-------|--------------|---------|---------|---------|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| (1, 1) | 0.49 | 4.76 | -2.11 | 0.45 |
| (1, 2) | 0.59 | 9.19 | 0.45 | 0.25 |
| (2, 2) | 0.59 | 9.19 | 0.45 | 0.30 |
| (2, 3) | 0.56 | 11.95 | 1.70 | 0.42 |
| (3, 3) | 0.56 | 11.95 | 1.70 | 0.32 |
| (3, 4) | 0.50 | 17.41 | 5.66 | 0.58 |

\( \omega = 1.00 \) \( r = 0.24 \) \( K = 8.80 \) \( L = 15.05 \)

- AI Welfare Gain = 1.15% \( \tau_K = 0.13, d = -2.79 \)

| Type | E(U)  | \( \ell_i \) | \( y_i \) | \( T(y_i) \) | \( T'(y_i) \) | \( \ell_{ij} \) | \( y_{ij} \) | \( T(y_{ij}) \) | \( T'(y_{ij}) \) |
|------|-------|--------------|---------|---------|---------|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| (1, 1) | 0.50 | 4.04 | -2.70 | 0.37 |
| (1, 2) | 0.60 | 7.75 | -0.78 | 0.15 |
| (2, 2) | 0.60 | 7.75 | -0.78 | 0.21 |
| (2, 3) | 0.56 | 10.01 | 0.05 | 0.35 |
| (3, 3) | 0.56 | 10.01 | 0.05 | 0.25 |
| (3, 4) | 0.51 | 14.78 | 3.33 | 0.52 |

\( \omega = 0.84 \) \( r = 0.69 \) \( K = 5.24 \) \( L = 15.19 \)

- AI, Restr. Debt Welfare Gain = -0.95% \( \tau_K = 0.73, d = 0.00 \)
their labor supply under a max-min SWF. The only exception is represented by the top-skilled middle-aged workers, who increase substantially their labor supply. These labor supply patterns are also reflected in how the marginal labor income tax rates vary when moving to an AD tax. Among the young, the marginal tax rate 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 drops substantially for the top-skilled workers. Under a utilitarian SWF, instead, both young workers (with, again, the exception of the least-skilled agents) and middle-aged workers increase their labor supply under an AD tax. Once again, the pattern of labor supply changes is mirrored in a coherent pattern of marginal tax rates, which decline for all agents except for the least-skilled among the young workers. Age-dependency increases aggregate labor supply by approximately 4% under a max-min SWF and by approximately 5% under a utilitarian SWF.

Looking at the results for the AI tax when debt policy is unrestricted, we can see that, both in the max-min and the utilitarian case, the golden-rule level of capital is achieved thanks to a large and negative value for the public debt. Constraining debt to be non-negative implies that the steady state capital-labor ratio is lowered by approximately 40% in both the max-min case and the utilitarian case, with a 15% drop in the equilibrium wage rate and a huge increase (179%) in the rental price for capital.

Another feature of the results which is worth emphasizing refers to the value of the optimal tax rate on the return on savings. Tables 3 and 4 show that the optimal value for $\tau_K$ is always positive and, moreover, that it is decreasing in the degree of sophistication of the other instruments at disposal of the government. It varies from about 10% under the optimal nonlinear AD tax to about 80% under the optimal nonlinear AI tax when the non-negativity constraint on debt is imposed. In section 5.5 we provide a more thorough discussion of the role played by this instrument in our model, highlighting that it can both help moving the economy closer to the golden-rule and relaxing the incentive constraints faced by the government. Even though the pursuit of each of these two goals might require varying $\tau_K$ in opposite directions, this instrument can serve the role of an imperfect substitute for the impossibility to levy age-dependent taxes or to run a negative public debt. Under this respect it seems reasonable that its importance increases when the sophistication of the other disposable instruments diminishes. Given

\footnote{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 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.}
that the optimal value for $\tau_K$ is very high under a nonlinear AI tax with constrained debt policy, we have also solved the model imposing the additional restriction that $\tau_K$ is equal to zero. This should tell us whether the high tax rates on the return on savings which appear in the bottom panels of tables 3 and 4 are primarily due to mimicking-deterring effects or capital-accumulation effects. We have found that, when the additional constraint $\tau_K=0$ is imposed, the steady state capital-labor ratio drops further (by approximately 15%). This suggests that, in the AI case with non-negative debt, the high tax rate on the return on savings is indeed used, at least partly, as an instrument to move the economy closer to the golden-rule.

Finally, the column headed $E(U)$ in tables 3 and 4 shows the expected utility of the various agents. As we can see, both under an ex-ante max-min and a utilitarian SWF, the nonlinear AD tax raises the expected lifetime utility for all types of agents. This Pareto-improving feature of age-dependency is likely to further strengthen its attractiveness as a policy proposal.

As we have discussed in section 4, computational reasons, especially relevant when it comes to calculate the optimal nonlinear AI solution, prevent us from using a model with more than three types of agents in the first period and four in the second. However, exploiting the fact that there are fewer self-selection constraints for the AD tax, we are able to solve the model for an optimal AD tax also for a six by seven case. Although for this more general case we cannot calculate the welfare gains of age-dependency, since we do not calculate the AI optimum, the six by seven case allows us to provide a more detailed characterization of the structure of the AD tax. We graphically illustrate the tax regime below. Panel a) of figure 1 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 is the one that 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.

In panel b) of figure 1 we show how labor income taxes vary with income, with the middle-aged workers paying much higher taxes than the young. This is a way for the government to drive private savings to a level compatible with the golden-rule condition.

\footnotesize

\begin{itemize}
\item\footnote{These results are available upon request.}
\item\footnote{This potential role for a tax on the return on savings was already emphasized, albeit in a setting without intragenerational heterogeneity, by Atkinson and Sandmo (1980) and Kotlikoff and Summers (1987).}
\item\footnote{The same restriction on transition probabilities as for the three by four case has been maintained.}
\item\footnote{Since public debt is a redundant instrument under the nonlinear AD tax, we have drawn the figure normalizing public debt at zero. In general, for an AD tax under a regime with unrestricted public debt the vertical location of the tax schedules in figure 1, panel b), would be indeterminate.}
\end{itemize}
Figure 1: Extended age dependent nonlinear income tax model ($Y =$ annual labor income expressed in thousands of US Dollars, $T =$ income tax paid, $MTR =$ marginal tax rate).

(a) Marginal tax rates for young (lower graph) and middle aged (upper graph).

(b) Tax functions for young (lower schedule) and middle aged (upper schedule).

The remainder of this section will be devoted to perform some sensitivity analysis. In particular, we consider four variations to our baseline scenario. First, we analyze what difference it makes if the government maximizes an ex-post max-min SWF rather than an ex-ante SWF as we have assumed so far. Second, we investigate how results are affected by the assumption that agents face uncertainty with respect to their skill in the second period. Third, we consider how results are affected by different assumptions about the wage process. Finally, we try different values for the coefficient of relative risk aversion and the Frisch elasticity.

**Ex-ante versus ex-post max-min social welfare function**

Table 5 provides the counterpart of table 3 for the case where the government maximizes an ex-post max-min SWF. This means that the government maximizes the minimum realized lifetime utility of a worker at the end of life (instead of the minimum expected lifetime utility as in the ex-ante max-min SWF). As one can see, the only noteworthy difference is that the overall welfare gain from age-dependency is further increased: moving from a nonlinear AI tax to a nonlinear AD tax delivers a welfare gain of 4.68% of GDP, whereas the corresponding figure for the ex-ante max-min case was 4.05%. The higher value for the welfare gain comes almost entirely from mimicking-deterring effects.
Table 5: Ex Post max-min

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.519</td>
<td>0.32</td>
<td>2.59</td>
<td>-6.90</td>
<td>0.70</td>
<td>0.22</td>
<td>2.13</td>
<td>-0.79</td>
<td>0.89</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.516</td>
<td>0.41</td>
<td>5.08</td>
<td>-5.01</td>
<td>0.57</td>
<td>0.47</td>
<td>7.26</td>
<td>3.53</td>
<td>0.63</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>-0.507</td>
<td>0.41</td>
<td>5.08</td>
<td>-5.01</td>
<td>0.57</td>
<td>0.47</td>
<td>7.26</td>
<td>3.53</td>
<td>0.63</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.465</td>
<td>0.50</td>
<td>9.76</td>
<td>-1.88</td>
<td>0.39</td>
<td>0.56</td>
<td>12.06</td>
<td>6.79</td>
<td>0.46</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.420</td>
<td>0.50</td>
<td>9.76</td>
<td>-1.88</td>
<td>0.39</td>
<td>0.56</td>
<td>12.06</td>
<td>6.79</td>
<td>0.46</td>
</tr>
</tbody>
</table>

$\omega = 1.00 \quad r = 0.24 \quad K = 8.32 \quad L = 14.23$

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.537</td>
<td>0.35</td>
<td>2.91</td>
<td>-3.65</td>
<td>0.60</td>
<td>0.30</td>
<td>2.91</td>
<td>-3.65</td>
<td>0.81</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.528</td>
<td>0.29</td>
<td>3.60</td>
<td>-3.08</td>
<td>0.79</td>
<td>0.52</td>
<td>8.06</td>
<td>0.35</td>
<td>0.54</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>-0.519</td>
<td>0.29</td>
<td>3.60</td>
<td>-3.08</td>
<td>0.79</td>
<td>0.52</td>
<td>8.06</td>
<td>0.35</td>
<td>0.58</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.492</td>
<td>0.41</td>
<td>8.06</td>
<td>0.35</td>
<td>0.63</td>
<td>0.60</td>
<td>12.83</td>
<td>3.50</td>
<td>0.45</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.432</td>
<td>0.41</td>
<td>8.06</td>
<td>0.35</td>
<td>0.63</td>
<td>0.60</td>
<td>12.83</td>
<td>3.50</td>
<td>0.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.552</td>
<td>0.36</td>
<td>2.52</td>
<td>-3.93</td>
<td>0.52</td>
<td>0.31</td>
<td>2.53</td>
<td>-3.91</td>
<td>0.78</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.543</td>
<td>0.29</td>
<td>3.06</td>
<td>-3.50</td>
<td>0.75</td>
<td>0.53</td>
<td>6.90</td>
<td>-0.68</td>
<td>0.47</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>-0.531</td>
<td>0.29</td>
<td>3.06</td>
<td>-3.50</td>
<td>0.75</td>
<td>0.53</td>
<td>6.95</td>
<td>-0.65</td>
<td>0.51</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.489</td>
<td>0.42</td>
<td>7.01</td>
<td>-0.62</td>
<td>0.55</td>
<td>0.60</td>
<td>10.88</td>
<td>1.71</td>
<td>0.37</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.444</td>
<td>0.42</td>
<td>7.01</td>
<td>-0.62</td>
<td>0.55</td>
<td>0.60</td>
<td>10.88</td>
<td>1.71</td>
<td>0.35</td>
</tr>
</tbody>
</table>

$\omega = 0.85 \quad r = 0.67 \quad K = 4.81 \quad L = 13.70$
Uncertainty

In order to assess the effect of uncertainty on our results, we run our model assuming that each agent correctly anticipates when young his skill type when middle-aged. For this purpose, we keep the wage process illustrated in table 1 but give a different interpretation to the values appearing within the matrix. More precisely, we assume that these values only represent the proportion of agents moving from a skill position to another and that these proportions are not perceived by agents as probabilities. As a social welfare function, we assume that the government maximizes a max-min objective function. The results of this exercise are summarized in table 6.

Looking at the welfare gains, we can see that the assumption that agents face no uncertainty about their future skill type does not change the magnitude of the welfare gains achievable by making the nonlinear tax age-dependent. We still obtain a welfare gain of about 4% of GDP as in our benchmark simulations. There are however two main things which differ. The first is the relative importance of capital-accumulation versus mimicking-deterring effects in the decomposition of the overall welfare gain. In the no uncertainty scenario we find that capital-accumulation effects account for two-thirds of the total welfare gain, whereas in the baseline case with uncertainty they accounted for approximately one half of the total welfare gain. Thus, it appears that eliminating uncertainty reduces the power of age-dependency as a mimicking-deterring device but at the same time strengthens its importance as an instrument to achieve the golden-rule. This can be explained as follows. In the absence of uncertainty agents no longer save for precautionary motives; this makes it harder for the economy to reach the golden-rule relying solely on private savings and therefore it makes more precious the role of an age-dependent tax as a device to boost savings and attain the golden-rule. On the other hand, as discussed in section 3, removing uncertainty about second period skill significantly reduces the number of strategies which can be adopted by an agent. This implies that the total number of incentive-compatibility constraints faced by the government, both under an AD and under an AI tax, is substantially reduced. It also implies that the importance of age-dependency as a mimicking-deterring device is likely to be weakened.\footnote{For instance, the total number of incentive-compatibility constraints in our three-by-four model with uncertainty is 2184 for the nonlinear AI case and 321 for the nonlinear AD case. In this case tagging by age eliminates 1863 IC-constraints (corresponding to about 85\% of the total IC-constraints under the AI tax). In the absence of uncertainty, these are substantially reduced.}

\footnote{Given the absence of uncertainty, the ex-ante and ex-post max-min criteria coincide.}
\footnote{As one can see from table 6, when public debt is constrained to be non-negative, the steady state capital-labor ratio under the optimal nonlinear AI tax drops by approximately 45\% as compared with the golden-rule level achieved under the AD tax. This lowers by 17\% the equilibrium wage rate (per efficiency unit of labor) and generates a huge increase (208\%) in the rental price for capital.}
Table 6: Deterministic Model

<table>
<thead>
<tr>
<th>Type</th>
<th>$E(U)$</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>0.526</td>
<td>33</td>
<td>2.73</td>
<td>-6.75</td>
<td>0.70</td>
<td>0.22</td>
<td>2.17</td>
<td>-0.38</td>
<td>0.89</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>0.521</td>
<td>33</td>
<td>2.73</td>
<td>-6.75</td>
<td>0.65</td>
<td>0.50</td>
<td>7.69</td>
<td>4.13</td>
<td>0.59</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>0.510</td>
<td>38</td>
<td>4.67</td>
<td>-5.26</td>
<td>0.68</td>
<td>0.50</td>
<td>7.69</td>
<td>4.13</td>
<td>0.57</td>
</tr>
<tr>
<td>(2, 3)</td>
<td>0.483</td>
<td>38</td>
<td>4.67</td>
<td>-5.26</td>
<td>0.63</td>
<td>0.52</td>
<td>11.21</td>
<td>6.44</td>
<td>0.60</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>0.462</td>
<td>56</td>
<td>11.08</td>
<td>-1.13</td>
<td>0.31</td>
<td>0.49</td>
<td>10.56</td>
<td>6.03</td>
<td>0.53</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>0.434</td>
<td>36</td>
<td>7.07</td>
<td>-3.45</td>
<td>0.64</td>
<td>0.71</td>
<td>24.57</td>
<td>15.26</td>
<td>0.23</td>
</tr>
</tbody>
</table>

$\omega = 1.00$, $r = 0.24$, $K = 8.18$, $L = 13.99$

<table>
<thead>
<tr>
<th>Type</th>
<th>$E(U)$</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>0.534</td>
<td>35</td>
<td>2.90</td>
<td>-3.44</td>
<td>0.66</td>
<td>0.30</td>
<td>2.90</td>
<td>-3.44</td>
<td>0.80</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>0.527</td>
<td>35</td>
<td>2.90</td>
<td>-3.44</td>
<td>0.60</td>
<td>0.54</td>
<td>8.40</td>
<td>0.76</td>
<td>0.50</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>0.516</td>
<td>27</td>
<td>3.40</td>
<td>-3.03</td>
<td>0.84</td>
<td>0.54</td>
<td>8.40</td>
<td>0.76</td>
<td>0.50</td>
</tr>
<tr>
<td>(2, 3)</td>
<td>0.482</td>
<td>27</td>
<td>3.40</td>
<td>-3.03</td>
<td>0.80</td>
<td>0.58</td>
<td>12.53</td>
<td>3.28</td>
<td>0.49</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>0.467</td>
<td>43</td>
<td>8.40</td>
<td>0.76</td>
<td>0.64</td>
<td>0.58</td>
<td>12.53</td>
<td>3.28</td>
<td>0.40</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>0.427</td>
<td>43</td>
<td>8.40</td>
<td>0.76</td>
<td>0.50</td>
<td>0.65</td>
<td>22.57</td>
<td>9.95</td>
<td>0.36</td>
</tr>
</tbody>
</table>

$\omega = 1.00$, $r = 0.24$, $K = 8.09$, $L = 13.84$

<table>
<thead>
<tr>
<th>Type</th>
<th>$E(U)$</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>0.553</td>
<td>36</td>
<td>2.43</td>
<td>-3.76</td>
<td>0.60</td>
<td>0.30</td>
<td>2.43</td>
<td>-3.76</td>
<td>0.76</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>0.545</td>
<td>36</td>
<td>2.43</td>
<td>-3.76</td>
<td>0.52</td>
<td>0.55</td>
<td>7.04</td>
<td>-0.41</td>
<td>0.42</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>0.534</td>
<td>28</td>
<td>2.83</td>
<td>-3.43</td>
<td>0.81</td>
<td>0.55</td>
<td>7.04</td>
<td>-0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>(2, 3)</td>
<td>0.499</td>
<td>28</td>
<td>2.83</td>
<td>-3.43</td>
<td>0.77</td>
<td>0.58</td>
<td>10.34</td>
<td>1.43</td>
<td>0.43</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>0.484</td>
<td>43</td>
<td>7.04</td>
<td>-0.41</td>
<td>0.58</td>
<td>0.58</td>
<td>10.34</td>
<td>1.43</td>
<td>0.33</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>0.444</td>
<td>43</td>
<td>7.04</td>
<td>-0.41</td>
<td>0.42</td>
<td>0.65</td>
<td>18.64</td>
<td>6.64</td>
<td>0.30</td>
</tr>
</tbody>
</table>

$\omega = 0.83$, $r = 0.74$, $K = 4.58$, $L = 13.95$
The wage process

To check the sensitivity of the results with respect to the structure of the skill process, we consider a different specification of it, and re-run our model for the baseline case with uncertainty and an ex-ante max-min SWF. For this purpose, following a procedure similar to the one described in section 4 for the case of US, we have constructed a wage process based on Swedish data. The welfare gain results for the nonlinear tax case are summarized in the second column of table 7.

Table 7: Wage process sensitivity analysis.

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>SWE</th>
<th>Mixed Economies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(w = w_{US})</td>
<td>(w = w_{SWE})</td>
<td>(w = w_{US})</td>
</tr>
<tr>
<td>(p = p_{US})</td>
<td>6.65%</td>
<td>5.48%</td>
<td>6.42%</td>
</tr>
<tr>
<td>(p = p_{SWE})</td>
<td>4.54%</td>
<td>4.49%</td>
<td>4.38%</td>
</tr>
<tr>
<td>Al, Restr. Debt</td>
<td>2.60%</td>
<td>2.50%</td>
<td>2.51%</td>
</tr>
<tr>
<td>Total gain</td>
<td>4.05%</td>
<td>2.98%</td>
<td>3.91%</td>
</tr>
<tr>
<td>IC share(^1)</td>
<td>2.11%</td>
<td>0.99%</td>
<td>2.04%</td>
</tr>
</tbody>
</table>

\(^1\) Welfare gains from age-dependency due to slackening of incentive-constraints.

As we can see, using the wage process obtained from Swedish data delivers a smaller welfare gain from age-dependency. The total welfare gain falls from about 4% of GDP, the value obtained under the US wage process, to about 3% of GDP. Looking at the decomposition of the welfare gain, this reduction is entirely due to smaller gains from mimicking-deterring effects. Capital-accumulation effects generate in both cases a welfare gain of about 2% of GDP; however, mimicking-deterring effects produce a welfare gain of about 2% of GDP for the US wage process and of about 1% of GDP for the Swedish wage process. This result can be explained by the fact that the tightness of the incentive-compatibility constraints faced by the government depends on the pre-tax wage inequality, with higher inequality implying tighter incentive-compatibility constraints thwarting the government’s pursuit of redistributive objectives. With the wage structure being more compressed in Sweden than in US, the role of age-dependency as an instrument to relax the incentive-compatibility constraints is weaker in Sweden.

numbers are lowered to respectively 858 and 210. In this case tagging by age eliminates 648 IC-constraints (corresponding to about 75% of the total IC-constraints under the AI tax).

\(^{30}\) Appendix B provides a detailed description of the procedure and a table summarizing the Swedish wage process.
than in US.\textsuperscript{31,32}

To get a better understanding of which features of the wage process are primarily responsible for the difference between the welfare gain for US and Sweden, we perform two additional sets of calculations based on mixtures of the US and Swedish wage processes. First, we run the model for a wage process where the wage rates for young and middle-aged are taken from the US wage process and the transition probabilities are taken from the Swedish process. Second, we do the reverse and run the model for a wage process where the wage rates are taken from the Swedish wage process and the transition probabilities are taken from the US process. The welfare gain results for these two cases are displayed in the third and fourth column of table 7. As we can see, the results in the third column are approximately the same as those obtained under the US wage process, whereas the results in the fourth column are approximately the same as those obtained under the Swedish wage process. This suggests that the difference between the welfare gains obtained for the US and Swedish wage processes is mainly due to a difference in the wage distributions.

The coefficient of relative risk aversion and the Frisch elasticity

To check the sensitivity of the results with respect to changes in the parameters entering the individuals’ utility function, we consider how the welfare gains are affected when using alternative values for $\gamma$ and $\kappa$. In particular, we re-run the model, for the baseline case with uncertainty and an ex-ante max-min SWF, using for $\gamma$ the values 0.9 and 2.5, and using for $\kappa$ the values 1.5 and 4. The results are presented in table 8.

One way to look at the relative wage compression in US versus Sweden is to compare how close is the median wage rate to the average wage rate in the two countries. As one can easily calculate from tables 1 (on page 21) and 10 (in the appendix), the median wage rate is closer to the average wage rate in Sweden, both when calculated within age-groups and when calculated on an expected lifetime basis. The estimated Gini coefficient for the US data is 0.378 for young agents, 0.499 for middle-aged agents. For the Swedish data it is 0.321 for young agents and 0.396 for middle-aged agents.

The fact that the age-dependency becomes less important as a mimicking-deterring device can also be seen by looking at tables 3 (page 25) and 11 (in the appendix) and comparing how marginal tax rates change when moving from an AI tax to an AD tax. 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, especially for the top-skilled workers among the middle-aged individuals.
Table 8: Sensitivity analysis: the coefficient of relative risk aversion and the Frisch elasticity.

<table>
<thead>
<tr>
<th>Variation in $\gamma$</th>
<th>$\gamma = 0.9; \kappa = 3$</th>
<th>$\gamma = 2; \kappa = 3$</th>
<th>$\gamma = 2.5; \kappa = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>6.12%</td>
<td>6.65%</td>
<td>6.45%</td>
</tr>
<tr>
<td>AI</td>
<td>4.23%</td>
<td>4.54%</td>
<td>5.21%</td>
</tr>
<tr>
<td>AI Restr. Debt</td>
<td>3.21%</td>
<td>2.60%</td>
<td>3.85%</td>
</tr>
<tr>
<td><strong>Total gain</strong></td>
<td>2.91%</td>
<td>4.05%</td>
<td>2.60%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variation in $\kappa$</th>
<th>$\gamma = 2; \kappa = 1.5$</th>
<th>$\gamma = 2; \kappa = 3$</th>
<th>$\gamma = 2; \kappa = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>3.49%</td>
<td>6.65%</td>
<td>6.94%</td>
</tr>
<tr>
<td>AI</td>
<td>2.23%</td>
<td>4.54%</td>
<td>5.03%</td>
</tr>
<tr>
<td>AI Restr. Debt</td>
<td>0.57%</td>
<td>2.60%</td>
<td>2.97%</td>
</tr>
<tr>
<td><strong>Total gain</strong></td>
<td>2.92%</td>
<td>4.05%</td>
<td>3.97%</td>
</tr>
</tbody>
</table>

As we can see, the results are sensitive to the choice of the parameters but the welfare gains from age-dependency are always substantial, ranging from about 3% to 4% of GDP.

### 5.4 Linear taxation

In this section we present the simulation results for the case when age-dependency is nested upon a linear taxation system. In table 9 we present the results for the different linear tax optima. The top panel refers to the ex-ante max-min SWF and the bottom panel to the utilitarian case. Both tables are calculated under the baseline choice for the parameters $\gamma$ and $\kappa$, and assuming that agents face uncertainty regarding their skill in the second period.
Table 9: Linear taxation

<table>
<thead>
<tr>
<th>Max-min</th>
<th>Debt</th>
<th>$\tau_K$</th>
<th>$\tau_1$</th>
<th>$\tau_2$</th>
<th>$G_1$</th>
<th>$G_2$</th>
<th>$d$</th>
<th>Welfare Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Yes</td>
<td>0.16</td>
<td>0.63</td>
<td>0.63</td>
<td>4.68</td>
<td>4.68</td>
<td>-2.41</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0.72</td>
<td>0.57</td>
<td>0.57</td>
<td>4.59</td>
<td>4.59</td>
<td>-</td>
<td>-2.07%</td>
</tr>
<tr>
<td>AD$^1$</td>
<td>-</td>
<td>0.10</td>
<td>0.64</td>
<td>0.62</td>
<td>4.66</td>
<td>4.66</td>
<td>-2.47</td>
<td>$\approx$0.01%</td>
</tr>
</tbody>
</table>

| AI              | Yes  | 0.12     | 0.34     | 0.34     | 2.90  | 2.90  | 3.22    | -            |
|                 | No   | 0.74     | 0.24     | 0.24     | 2.88  | 2.88  | -       | -2.99%       |

<table>
<thead>
<tr>
<th>Utilitarian</th>
<th>Debt</th>
<th>$\tau_K$</th>
<th>$\tau_1$</th>
<th>$\tau_2$</th>
<th>$G_1$</th>
<th>$G_2$</th>
<th>$d$</th>
<th>Welfare Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Yes</td>
<td>0.18</td>
<td>0.31</td>
<td>0.35</td>
<td>2.91</td>
<td>2.91</td>
<td>-3.08</td>
<td>$\approx$0.01%</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0.48</td>
<td>0.27</td>
<td>0.34</td>
<td>5.57</td>
<td>0</td>
<td>-</td>
<td>$-0.16%$</td>
</tr>
</tbody>
</table>

$^1$ The golden rule is achievable in the age-dependent with non-negative lump-sum transfers. Thus age-dependent lump-sum transfers can be replaced by a uniform transfer and debt.

$^2$ The golden rule is not achievable in the age-dependent with non-negative lump-sum transfers. Non-negative age-dependent lump-sum transfers are not equivalent to a uniform transfer and debt.

Under an unrestricted debt policy, a linear AD tax offers only a negligible welfare gain as compared to the optimal linear AI tax. 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 solution to the government’s problem when debt policy is unconstrained. In this case, an optimal linear AD tax delivers welfare gains ranging from 2% (under an ex-ante max-min SWF) to 3% (under a utilitarian SWF). 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 a linear taxation framework.

Having completed the presentation of our results pertaining to the effects of age-dependent taxes, in the next section we provide some further comments on the role played in our model by interest income taxation.

### 5.5 Interest income taxation

There is in the economic literature a long standing interest in the question of whether there should be a positive tax on the return on savings. Our simulations contribute to shed light on this issue too. We have considered several different specifications of our model and in almost all specifications we have obtained a positive tax on interest income.$^{33}$ However, its numerical

---

$^{33}$This is in line with many recent studies like, for example, Conesa et al. (2009) who find that the optimal tax rate on the return on savings should be around 30%. Early studies,
magnitude varies quite a lot: from a minimum of -4% for the nonlinear AI tax under unrestricted debt policy and no uncertainty regarding second-period skills (see table 6), up to 84% for the nonlinear AI tax with public debt restricted to be non-negative and uncertainty about second-period skills (see table 3). 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 mitigating incentive-compatibility 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 where the set of policy instrument is so rich that the golden-rule capital-labor ratio can be obtained without any help from interest income taxation. These are the nonlinear AI tax under unrestricted public debt and the nonlinear AD tax systems. For these cases a non-zero tax on interest income is only due to its role as a mimicking-deterring device. Since interest income taxation can be viewed as a surtax on future consumption, its role as a redistributive device in nonlinear tax models can be analyzed within the framework of the Atkinson-Stiglitz (1976) theorem. This theorem states conditions that should be satisfied for the redundancy of (differentiated) commodity taxation as an additional instrument in models of nonlinear labor income taxation. One such condition is that labor be weakly separable from commodities in the utility function. In our model this condition is satisfied. However, the Atkinson-Stiglitz result is also based on the implicit assumption that a mimicker’s disposable income coincides with that of the agent being mimicked. This condition, which holds in most optimal taxation models, is violated in our setting. More precisely, in our model the second-period expected disposable labor income of a mimicker differs in general from that of the agent being mimicked. Thus, (differentiated) commodity taxation is not a redundant instrument and, according to the prescriptions of the optimal commodity tax literature, one should tax more heavily the commodities for which the mimicker’s consumption is larger than that of the mimicked agent. Hence, a positive capital income tax rate becomes desirable if, at the binding incentive-compatibility 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 derived in Appendix A.4. From our simulations we can see that the optimal value of the interest income tax rate jumps around quite a lot. However, this is hardly surprising given that $\tau_K$ depends on the difference between the amount of savings of the mimickers, and that of the agents being mimicked, like Judd (1985) and Chamley (1986), using an infinitely-lived representative agent model, obtained the result that the long-run tax on the return on savings should be zero. The opinions in the academic community about the desirability of taxing capital income are still mixed. Two recent contributions expressing diverging views are provided by Mankiw et al. (2009) and Diamond and Saez (2011).
at the various incentive-compatibility constraints that are binding at the solution to the government’s problem. Since a change in the wage process or a switch from a nonlinear AI labor income tax to a nonlinear AD 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 $\tau_K$ is affected.

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 under the non-negativity restriction on public debt. For an AD tax with no restrictions on the demogrants, private savings (and hence the capital stock) can be perfectly controlled by setting the demogrants in a proper way. For the linear AI tax the demagrant must be the same in both periods. However, a tax on interest income can serve as an imperfect substitute for AD demogrants. Increasing the tax on interest income and using the proceeds to raise the demagrant implies a redistribution of resources to the young, hence it will tend to increase savings. The point that the effect of interest income taxation on the savings behavior of agents depends crucially on how the proceeds from interest income taxation are rebated back to agents is discussed at length in Sandmo (1985) and Kotlikoff and Summers (1987). In general, if an increase in the interest income tax rate is accompanied by a reduction in the average tax rate levied on labor income earned at the time when the return on savings accrue, savings will be discouraged. However, if the proceeds from interest income taxation are rebated to agents through a reduction in the average tax rate levied on labor income earned prior to the time when the return on savings accrue, savings tend to be boosted. Under a linear AI tax both the marginal tax rates and the demogrants are age-invariant, which means that a compensated increase in the interest income tax rate will necessarily lower to some extent also the average tax rate on labor income earned when young. In this way it will tend to have stimulating effect on savings.

Of course, an increase in the tax on interest income will also 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 income tax rate, from about 10% to about 70% (see table 9), that is obtained for the linear AI tax as we impose the non-negativity restriction on public debt can best be understood as a way to increase individual savings and hence the capital stock.

For the nonlinear AI tax with the non-negativity constraint on public debt, 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 instance, for our baseline scenario, we see that, when the non-negativity constraint on public debt is imposed, the optimal interest
tax rate increases from 40% to 84% for the ex-ante max-min SWF under a nonlinear AI labor income tax (see table 3) and from 13% to 73% for the utilitarian SWF (see table 4).\footnote{However, it is not clear that one should attribute all of this increase to the capital accumulation mechanism as the set of binding self-selection constraints will in general also change as we introduce the non-negativity constraint on debt.}

We have seen that the optimal rate of the interest income tax varies to a large extent with the specification of the model. 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 economies. However, this does not necessarily imply large welfare losses. Whether it does or not, it crucially depends on the sophistication of the other instruments at the government’s disposal. We have done simulations looking at how the welfare gains (with respect to the benchmark linear AI tax scenario) are lowered when interest income is tax-exempt.\footnote{These results are available upon request.} The results suggest that setting the proportional interest income tax at the optimal rate is of second order importance under an optimal age-dependent labor income tax. Under an age-independent labor income tax, on the other hand, the welfare loss due to tax-exempting interest income can be substantial, especially when the labor income tax is linear and public debt is assumed to be non-negative. For our benchmark specification, we find for the ex-ante max-min SWF a welfare loss of about 2% of GDP under a nonlinear age-independent labor income tax, and of 3% under a linear age-independent tax, in both cases assuming that public debt is restricted to be non-negative. For the utilitarian SWF, the corresponding figures are 3.5% and 4%.

6 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 the government faces restrictions in the use of debt policy, which for our purposes means that the government is prevented from running a negative debt. For US, under our baselines scenario and with a max-min social welfare function, the welfare gain of switching from an optimal nonlinear age-independent income tax to an optimal non-
linear age-dependent income tax is equivalent to about 4% of total output. This gain can be decomposed into a share due to incentive effects arising from alleviating self-selection constraints and a share due to capital-accumulation effects, making easier for the government to meet the golden-rule level of capital. Each source of welfare gain accounts for approximately half of the total welfare gain. Lowering the degree of social aversion to inequality and taking a utilitarian social welfare function reduces the welfare gain to about 3.3% of GDP and slightly increases the share of the total welfare gain due to capital-accumulation effects.

Using a wage process calibrated on Swedish data lowers the welfare gain under a max-min social welfare function from 4% to 3% of GDP, with the reduction being triggered by a diminished importance of age-dependency as a mimicking-deterring device. This result is consistent with the idea that a higher wage inequality tightens the incentive-compatibility constraints faced by the government in the design of a nonlinear income tax. As wage inequality is lower in Sweden as compared to US, the gain from age-dependency due to relaxing incentive constraints becomes smaller.

Assuming that agents face no uncertainty about their future wage rate does not affect the magnitude of the overall welfare gain obtained by making the nonlinear income tax age-dependent. However, the share of the total gain due to capital-accumulation effects becomes more important, accounting for almost three-fourths of the welfare gain. The intuition for the result is twofold. On one hand, in the absence of uncertainty agents no longer save for precautionary motives, and this makes it harder for the economy to reach the golden-rule relying solely on private savings. On the other hand, removing uncertainty reduces the number of incentive-compatibility constraints faced by the government in the design of the income tax, both under an age-dependent and under an age-independent tax. This tends to mechanically weaken the power of age-dependency as a mimicking-deterring device.

As compared to the gains which can be obtained by making the nonlinear income tax age-dependent, those generated when tagging by age is adopted under a linear income tax are smaller. Essentially, this happens because the only source of gain which is left under a linear income tax is given by the gains due to capital-accumulation effects.

We also find that the gains of using a nonlinear labor income tax instead of a linear tax are roughly of the same order of magnitude as those due to tagging by age (at least in the baseline scenario where agents face uncertainty). However, our results also suggest that reducing the degree of social aversion to inequality lowers the magnitude of the welfare gains that can be achieved by using a nonlinear tax more than it lowers the welfare gains that can be reaped by making the nonlinear tax age-dependent.

Finally, we find that setting the proportional interest income tax at the optimal rate is of second order importance under an optimal age-dependent labor income tax. Under an age-independent labor income tax, on the other
hand, the welfare loss due to tax-exempting interest income can be substantial, especially when the labor income tax is linear and public debt is constrained to be non-negative. For our benchmark specification, we find a welfare loss of about 2% of GDP under a nonlinear age-independent labor income tax, and of 3% under a linear age-independent tax, in both cases assuming that public debt is constrained to be non-negative.
References


Fukushima, K. (2010). Quantifying the welfare gains from simple tax systems. mimeo.


A Proofs and derivations.

A.1 Calculation of the total number of strategies (nonlinear taxation).

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 under a nonlinear AD tax can be derived as follows. Suppose that there is a function $\phi(i)$ that, for every ability type $i \in I$ when young, gives the corresponding number of possible ability types when old. For instance, $\phi(i) = j$ would mean that, if an agent is of ability type $i$ when young, his ability type when old 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 number of bundles offered by the government is $\Psi = \sum_{i=1}^{m^y} \phi(i)$. Consider an agent with a specific skill type when young. In the first period he can choose among $m^y$ different income points. 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 $\phi(i)$ possible skill realizations there are $\Psi \phi(i)$ 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, one gets that the total number of strategies, including the strategies entailing truthful revelation, is given by $|\Gamma| = m^y \left[ \sum_{i=1}^{m^y} \Psi \phi(i) \right]$. Under the special assumption that each agent has the same number of different possible skill realizations in the second period, i.e. $\phi(i) = \phi$ for all $i$, one gets $|\Gamma| = m^y [m^y \Psi \phi] = (m^y)^2 \Psi \phi$.

A similar reasoning allows calculating the total number of strategies under a nonlinear AI tax. In this case there are $\Psi + m^y$ bundles to choose from both when young and when old. Consider an agent with a specific skill type when young. In the first period he can choose among $(m^y + \Psi)$ different bundles. For any given choice in the first period, a strategy must specify a bundle to be chosen for each possible skill realization in the second period; with $\phi(i)$ possible skill realizations there are $(\Psi + m^y) \phi(i)$ ways to choose a bundle in the second period. Taking into account that this applies to each of the $m^y$ possible types of young agents, the total number of strategies, including the strategies entailing truthful revelation, is given by $|\hat{\Gamma}| = (m^y + \Psi) \left[ \sum_{i=1}^{m^y} (m^y + \Psi) \phi(i) \right]$. Under the special assumption that each agent has the same number of different possible skill realizations in the second period, i.e. $\phi(i) = \phi$ for all $i$, one gets $|\hat{\Gamma}| = (m^y + \Psi) \left[ \sum_{i=1}^{m^y} (m^y + \Psi) \phi \right]$. 

45
A.2 Proof that the nonlinear AD tax always achieves the golden rule.

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

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

and the capital market equilibrium condition is:

\[ k_{t+1} - (1-\delta) \frac{k_t}{1+n} \frac{\tilde{L}_t}{L_{t+1}} = \frac{s_t}{(1+n)\tilde{L}_{t+1}} + \frac{s_{t-1}}{(1+n)^2 \tilde{L}_{t+1}} \]

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

\[ k = \frac{1}{n+\delta} \left( \frac{s}{L} + \frac{s-1}{(1+n)L} \right) \]

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

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

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

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

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

\[ \max_{\{b_i,y_i\} \in \Gamma, \{b_{ij},y_{ij}\} \in \Gamma_{ij \in T \land K}} W(V_1(\tilde{\sigma}^1), ..., V_m(\tilde{\sigma}^m)) \]

subject to:

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

46
and

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

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 \(-[1 + r (1 - \tau_K)]\) 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 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

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

and denoting by \( \mu \) the Lagrange multiplier associated with the production constraint, the effect of the reform is given by:
\[-\mu \left[ \sum_i p_i - \frac{1+r(1-\tau_K)}{1+n} \sum_{ij} p_i p_{ij} \right] \]
\[-\mu \left[ -\frac{f'}{n+\delta} - \sum_i p_i + \frac{1+r(1-\tau_K)}{1+n} \sum_{ij} p_i p_{ij} + 1 \right] \left\{ \sum_i \frac{\partial s}{\partial b_i} - [1 + r \ (1 - \tau_K)] \sum_{ij} \frac{\partial s}{\partial b_{ij}} \right\} = 0,\]

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 \( \sum_i \frac{\partial s}{\partial b_i} - [1 + r \ (1 - \tau_K)] \sum_{ij} \frac{\partial s}{\partial b_{ij}} = 1, \)

the equation above can be simplified to:

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

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.

A.3 The nonlinear AI tax and the conditions for golden rule.

In Appendix A.2 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 AI tax, it should not be surprising that the first order conditions of the government’s problem do not allow recovering 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 A.2. 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 AI 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.\(^\text{36}\) 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:

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

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.\(^\text{37}\) 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 A.2) cannot be a solution to the equation.\(^\text{38}\)

\(^{36}\) 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.

\(^{37}\) 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.

\(^{38}\) The self-selection term would only vanish if, at the solution to the government’s problem under a nonlinear AI 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 AI to a nonlinear AD tax would be nil.
A.4 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 \( r \left[ s_i + \frac{s_i(j)}{1+r(1-\tau)} \right] \) 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 \( \tau_K \). Notice that the policy experiment of marginally increasing \( \tau_K \), while at the same time raising all the various \( b_{ij} \) by \( r \left[ s_i + \frac{s_i(j)}{1+r(1-\tau)} \right] \), 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. Denote 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 \( \sigma^i \), and denote by \( \left( \frac{dV_i(\sigma^i)}{d\tau K} \right) dV_1(\tilde{\sigma}^1)=0,\ldots,dV_{my}(\tilde{\sigma}^{my})=0 \) the effect of a compensated (for non-deviating agents) marginal increase in \( \tau_K \) on the expected utility of a type-\( i \) agent who follows the deviating strategy \( \sigma^i \). Then, after some manipulations, one obtains the following formula, where a superscript \( h \) has been used to denote Hicksian demands:  

\[
\mu \left[ \frac{f}{n+\delta} - \frac{1+r(1-\tau_K)}{1+n} \right] \left( \frac{\partial s^h}{\partial \tau K} + \frac{1}{1+n} \frac{\partial s^{-h}}{\partial \tau K} \right) = \sum_i \sum_{\sigma^i} \lambda_i(\sigma^i) \left( \frac{dV_i(\sigma^i)}{d\tau K} \right) dV_1(\tilde{\sigma}^1)=0,\ldots,dV_{my}(\tilde{\sigma}^{my})=0 . \tag{7}
\]

\[39\] With the expression “compensated marginal increase in \( \tau_K \)” we refer to a reform which raises marginally \( \tau_K \) and at the same time adjust the various after-tax labor incomes \( b_{ij} \) offered to the middle-aged workers in such a way to leave unchanged the expected utility for all non-deviating agents. As explained in the text this requires raising each \( b_{ij} \) by \( r \left[ s_i + \frac{s_i(j)}{1+r(1-\tau)} \right] \).

\[40\] The formula can be viewed as a generalization of eq. (9) in Blomquist and Micheletto (2008).
Given that under a nonlinear AD tax we have $r = n \Rightarrow f' = n + \delta$ and that the same condition holds under a nonlinear AI tax when debt policy is available, in these cases the condition implicitly characterizing the optimal interest income tax rate can be simplified to:

$$\mu \frac{r\tau K}{1 + n} \left( \frac{\partial s^h}{\partial \tau K} + \frac{1}{1 + n} \frac{\partial s^h_{-1}}{\partial \tau K} \right) = \sum_i \sum_{\sigma^i} \lambda_{i(\sigma^i)} \left( \frac{dV_i(\sigma^i)}{d\tau_k} \right) dV_i(\tilde{\sigma}^1)=0, \ldots, dV_{m^n}(\tilde{\sigma}^{m^n})=0. \tag{8}$$

The important thing to notice about (7) and (8) is that the value of each of the terms which are summed up on the right-hand side is ultimately related to 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 AI to a nonlinear AD tax is likely to produce significant changes in the pattern of binding self-selection constraints, one should not be surprised that the level of the optimal interest income tax rate varies in an a priori unpredictable way.

**B The wage process and optimal allocation for 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 a procedure and sample selection which resembles as closely as possible that which was applied to the US data. However, due to the high quality of this data-set we use the same data-set to estimate the transition probabilities and the wage distributions for the young and middle-aged. The wage distribution for the young is represented by the wages of those individuals who were aged 25 during the period 1981-1985. The wage distribution for the middle-aged is represented by the wages of these same individuals when they were 45, during the period 2001-2005.\footnote{Individuals with an hourly wage rate of less than SEK 15 or greater than SEK 350 (1980 prices) were dropped from the sample. In Sweden payroll taxes are much higher than in the US (in Sweden the payroll tax rate was 32.82% in 2003). Because the nonlinear income taxes are meant to represent the complete tax-transfer system, we have inflated.
Table 10: Transition matrix and hourly wages for the Swedish Economy.\textsuperscript{1}

\begin{table}[h]
\centering
\begin{tabular}{cccccc}
\hline
 & \multicolumn{4}{c}{Middle-aged} & \\
\hline
Young & 1 & 2 & 3 & 4 & Wages \\
\hline
1 & 0.38 & 0.62 & 0 & 0 & 10.55 \\
2 & 0 & 0.54 & 0.46 & 0 & 17.52 \\
3 & 0 & 0 & 0.65 & 0.35 & 22.81 \\
\hline
Wages & 13.72 & 19.73 & 24.45 & 37.30 & \\
\hline
\end{tabular}
\end{table}

\textsuperscript{1} For comparison with the wages for the US economy, the wages above have been transformed from 2003 Swedish Kronor into 2003 US Dollars using the average exchange rate over the period 2001-2005.
Table 11: Swedish wage process, ex-ante max-min

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.44</td>
<td>0.29</td>
<td>3.08</td>
<td>-7.96</td>
<td>0.71</td>
<td>0.37</td>
<td>5.03</td>
<td>1.57</td>
<td>0.72</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.42</td>
<td>0.44</td>
<td>7.67</td>
<td>-4.50</td>
<td>0.49</td>
<td>0.51</td>
<td>10.16</td>
<td>5.21</td>
<td>0.48</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>-0.42</td>
<td>0.44</td>
<td>7.67</td>
<td>-4.50</td>
<td>0.49</td>
<td>0.51</td>
<td>10.16</td>
<td>5.21</td>
<td>0.48</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.38</td>
<td>0.50</td>
<td>11.48</td>
<td>-2.48</td>
<td>0.27</td>
<td>0.52</td>
<td>12.63</td>
<td>6.56</td>
<td>0.41</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.38</td>
<td>0.50</td>
<td>11.48</td>
<td>-2.48</td>
<td>0.27</td>
<td>0.52</td>
<td>12.63</td>
<td>6.56</td>
<td>0.41</td>
</tr>
</tbody>
</table>

$\omega = 1.00$  $r = 0.24$  $K = 9.88$  $L = 16.87$

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.45</td>
<td>0.31</td>
<td>3.27</td>
<td>-4.17</td>
<td>0.67</td>
<td>0.35</td>
<td>4.75</td>
<td>-2.94</td>
<td>0.77</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.42</td>
<td>0.38</td>
<td>6.69</td>
<td>-1.53</td>
<td>0.61</td>
<td>0.56</td>
<td>11.05</td>
<td>1.16</td>
<td>0.35</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>-0.42</td>
<td>0.38</td>
<td>6.69</td>
<td>-1.53</td>
<td>0.61</td>
<td>0.56</td>
<td>11.05</td>
<td>1.16</td>
<td>0.35</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.39</td>
<td>0.48</td>
<td>11.05</td>
<td>1.16</td>
<td>0.36</td>
<td>0.52</td>
<td>12.71</td>
<td>2.02</td>
<td>0.48</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.39</td>
<td>0.48</td>
<td>11.05</td>
<td>1.16</td>
<td>0.36</td>
<td>0.52</td>
<td>12.71</td>
<td>2.02</td>
<td>0.48</td>
</tr>
</tbody>
</table>

$\omega = 1.00$  $r = 0.24$  $K = 9.67$  $L = 16.52$

<table>
<thead>
<tr>
<th>Type</th>
<th>E(U)</th>
<th>$\ell_i$</th>
<th>$y_i$</th>
<th>$T(y_i)$</th>
<th>$T'(y_i)$</th>
<th>$\ell_{ij}$</th>
<th>$y_{ij}$</th>
<th>$T(y_{ij})$</th>
<th>$T'(y_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>-0.46</td>
<td>0.31</td>
<td>2.80</td>
<td>-4.46</td>
<td>0.62</td>
<td>0.35</td>
<td>4.06</td>
<td>-3.44</td>
<td>0.73</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>-0.43</td>
<td>0.38</td>
<td>5.68</td>
<td>-2.33</td>
<td>0.56</td>
<td>0.57</td>
<td>9.45</td>
<td>-0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>-0.43</td>
<td>0.38</td>
<td>5.68</td>
<td>-2.33</td>
<td>0.56</td>
<td>0.57</td>
<td>9.45</td>
<td>-0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>-0.40</td>
<td>0.49</td>
<td>9.45</td>
<td>-0.22</td>
<td>0.26</td>
<td>0.52</td>
<td>10.73</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-0.40</td>
<td>0.49</td>
<td>9.45</td>
<td>-0.22</td>
<td>0.26</td>
<td>0.52</td>
<td>10.73</td>
<td>0.35</td>
<td>0.35</td>
</tr>
</tbody>
</table>

$\omega = 0.85$  $r = 0.67$  $K = 5.87$  $L = 16.67$