Dynamic Optimal Redistributive Taxation with Endogenous Retirement

Jean-Baptiste Michau*

Centre for Economic Performance, London School of Economics

March 2009

Abstract

While the participation decision is discrete in a static context, i.e. to work or not to work, such is not the case in a dynamic context where workers choose the fraction of their lifetime that they spend working. In this paper, I therefore characterize the optimal redistributive policy in a dynamic environment with both an intensive and an extensive margin to labor supply. The government should optimally design a history-dependent social security system which induces higher productivity individuals to retire later. Redistribution should be done through the social security system rather than with a non-linear income tax.

Keywords: Extensive margin, Optimal social security, Redistribution, Retirement age

JEL Classification: E62, H21, H55, J26

1 Introduction

Labor supply indivisibilities, such as those caused by fixed costs of working, are pervasive. This creates an extensive margin to labor supply which forces individuals to make a participation decision. This choice is inherently discrete in a static context, i.e. to work or not to work, but not in a dynamic framework where agents choose the fraction of their lifetime that they spend working or, equivalently, their retirement age.

The implications of the extensive margin for optimal taxation have been analyzed rather extensively in a static environment (see, for instance, Diamond 1980, Saez 2002, Immervoll Kleven Kreiner Saez 2007, Chone Laroque 2005, 2008, Laroque 2005, Beaudry

*Contact: J.B.Michau@lse.ac.uk

Blackorby Szalay 2007, Blundell Shephard 2008). Importantly, this literature has provided some support for the implementation of tax credits, such as the Earned Income Tax Credit in the US. However, abstracting from the dynamic aspect of workers' labor supply problem seems to be more than a simplifying assumption. Indeed, it fundamentally changes the nature of the participation decision by, artificially, making it discrete. More generally, the importance of dynamic issues for the optimal design of taxes has long been recognized in economics, at least since Vickrey (1939).

Furthermore, the positive analysis of taxation has recently emphasized the relevance of the dynamic framework with an extensive margin (see Mulligan 2001, Ljungqvist Sargent 2006, 2008, Prescott Rogerson Wallenius 2009, Rogerson Wallenius 2008). In particular, it provides a natural explanation for the discrepancy between the well-documented small elasticity of labor supply along the intensive margin and the large effects of taxation needed to rationalize a number observed macroeconomic phenomena, such as the difference in the total amount of hours worked between Europe and the US[1]. As explained by Prescott (2006), micro elasticities along the intensive margin are small precisely because the adjustment occurs along the extensive margin.

The goal of this paper is therefore to determine the optimal redistributive policy in a dynamic environment where workers are heterogeneous in productivity. I allow for two dimensions to the labor supply decision: the number of hours of work conditional on participation, i.e. the intensive margin, and the retirement age, i.e. the extensive margin. I first rely on the revelation principle to determine the optimal incentive-feasible allocation of resources, in the spirit of Mirrlees (1971). I then turn to the implementation of the optimum in a decentralized economy. Finally, I perform a numerical calibration of the model to illustrate the main features of the optimal policy.

The main policy recommendations differ substantially from those of a static analysis. The career length of workers should be increasing in their productivity. Hence, the retirement age should be a key input of the fiscal system which, naturally, takes the form of a history-dependent social security system. Consequently, the shape of the period-by-period income tax schedule is indeterminate as anything can be undone by adjusting the history-dependent transfers received after retirement. This analysis therefore fails to provide some support for the implementation of tax credits. However, it justifies doing a large amount of redistribution within the pension system. While this is already the case in practice, there has, so far, been little theoretical justification for seeing social security as more than a savings device.

The issue of the optimal design of a social security system with heterogeneous agents

[1] Prescott (2002, 2004) implicitly invokes the existence of employment lotteries to justify a high elasticity of labor supply. But, as noted by Ljungqvist and Sargent (2006), it is more natural, and equally effective, to assume an extensive margin in a dynamic setup.

and endogenous retirement has, so far, been largely overlooked. Two important exceptions include the pioneering work of Diamond (2003, chapter 6) as well as Sheshinski (2008). In both cases, the source of heterogeneity is the disutility of labor, which could be interpreted as a fixed cost of working, rather than productivity. Their main finding is that agents with a low disutility of labor retire later than others and that some of the income generated by their extra activity is redistributed to those having a high disutility of labor. However, they restrict themselves to three period models and, hence, their conclusions should be seen as qualitative.

Cremer, Lozachmeur and Pestieau (2004) also look at optimal social security with endogenous retirement. Workers can only be of two or three types which differ in productivity and in disutility of labor. They show that the retirement age is distorted downward for everybody except for workers with the highest productivity and lowest disutility of labor. Again, their results should be seen as qualitative.

Of related interest, Gorry and Oberfield (2008) solve a dynamic optimal taxation problem in a life cycle framework with both an intensive and an extensive margin to labor supply. Their framework consists of a representative agent who must be taxed to finance an exogenous amount of government expenditure. Importantly, the only fiscal instrument allowed is a standard non-linear income tax. Hence, the policy which they drive is only constrained optimal. This explains why the "no distortion at the top" principle does not hold in their context.

There has recently been a growing literature on dynamic optimal taxation with heterogeneous agents. The main focus has been on the provision of insurance against skill risks. However, this literature has been unable to provide a general characterization of the optimal allocation of time between work and leisure, which seems paradoxical given the central importance of labor income taxes in the static optimal taxation literature. Hence, numerical simulations of optimal policies have only been possible in simplified setups. For instance, Golosov Tsyvinsky Werning (2006), Kocherlakota (2005) and Weinzierl (2008) restricted the number of time periods to two or three, Albanesi Sleet (2006) focused on independently and identically distributed shocks, Diamond Mirrlees (1978), Golosov Tsyvinski (2006) and Denk Michau (2008) had a permanent disability shock and Kapicka (2006) does not allow for savings. Also, Battaglini and Coate (2008) could characterize the optimal labor income tax in a dynamic redistribution problem with stochastically evolving skills; but they had to assume risk neutrality in order to kill any desire to provide insurance. This paper complements this literature by determining the optimal distortions to labor supply in a dynamic context without uncertainty.[2]

[2]Note that, with an intensive margin only and constant productivity throughout the lifetime of individuals, the dynamic optimal taxation problem is not particularly interesting as it is just a replication of the static optimal taxation problem.

I begin by describing, in section 2, the structure of the economy. The optimal incentive-feasible allocation is derived in section 3. I then show in section 4 that a history-dependent social security system can implement the optimum in a decentralized economy. Section 5 contains a numerical simulation of the optimal policy. This paper ends with a conclusion.

2 Model

Individuals face a deterministic life-span equal to H. Utility is additively separable between consumption and leisure. Agents derive an instantaneous utility $u(c_t)$ from consuming c_t at age t, where $u' > 0$, $u'' < 0$ and $\lim_{c \to 0} u(c) = -\infty$. They work until some retirement age R and get disutility $v(l_t)$ from supplying l_t units of labor at t, where $v(0) = 0$, $v'(0) = 0$, $v' \geq 0$ and $v'' > 0$. They also have to incur a fixed cost of working $b > 0$ which, for simplicity, is assumed to be constant over time. Lifetime utility V is time separable. Continuous time is assumed, which is convenient to derive the endogenous retirement age R. The future is discounted at rate ρ. Thus, individuals have the following preferences:

$$V = \int_0^H e^{-\rho t} u(c_t) dt - \int_0^R e^{-\rho t} \left[v(l_t) + b \right] dt. \tag{1}$$

Note that the value of leisure is normalized to zero when individuals are not working, i.e. from age R to T.

Resources can be transferred across time at an exogenous interest rate[3] which, for simplicity, is taken to be equal to the discount rate ρ. Each agent is characterized by a productivity index α and faces a deterministic productivity profile $\{\gamma_t(\alpha)\}_{t \in [0,H]}$. Thus an α-worker produces output $\gamma_t(\alpha)$ if he supplies one unit of labor at age t. As will become clear, I need to assume that productivity at each age is weakly increasing in the productivity parameter of the agent[4]. More formally, $\alpha > \alpha'$ implies $\gamma_t(\alpha) \geq \gamma_t(\alpha')$ for all t with a strict inequality for at least one t. Thus, the deterministic productivity profiles of two agents are not allowed to cross at any point in time. Although reasonable, this assumption rules out, for instance, football players who, contrary to the vast majority of the population, get their highest salary when young.

The specification of utility in (1) entails both an intensive and an extensive margin to labor supply. Clearly, conditional on working, agents need to choose a number of hours

[3] We therefore abstract from the way resources are shifted over time. In an overlapping generation framework, the model would therefore be compatible with a fully funded social security system, where the interest rate corresponds to the returns to capital, and with a pay-as-you-go system, where the interest rate is determined by the rates of growth of population and output.

[4] A natural candidate specification, which is used in the calibration of the model, is to have a baseline productivity profile γ_t, common to all workers, multiplied by the individual-specific productivity parameter α; thus $\gamma_t(\alpha) = \alpha \gamma_t$.

of work; this is the intensive margin. As the disutility cost of working v is increasing and convex, without fixed costs of working, agents would choose to work until their death, i.e. $R = H$. However, the fixed cost of working creates a labor supply indivisibility which induces agents to make a participation decision at each age; this is the extensive margin.

In a static context, this indivisibility generates a non-convexity in the workers' production possibility set which, as argued by Hansen (1985) and Rogerson (1988), could be overcome by resorting to employment lotteries together with a complete set of markets for consumption claims. A criticism to this theoretical argument is that lotteries are just not available to most household. However, in a dynamic context, agents can instead convexify their production possibility set by alternating spells of work and leisure while trading a risk-free asset to smooth their consumption over time. This directly applies to the framework of this paper and we therefore have a "time averaging" model of the labor supply *à la* Diamond Mirrlees (1978) or Mulligan (2001). Ljungqvist and Sargent (2006, 2009) have shown that, in continuous time, lotteries and time averaging models of indivisible labor are equivalent when productivity is constant and quantitatively very similar otherwise.

Thus, the key decision that agents have to make at the extensive margin is the fraction of their lifetime spent working. Our specification of utility in (1) implicitly assumes that agents prefer to work at the beginning of their life-span, from age 0 to R, and retire at the end, from R to H. This is the only possibility with a declining productivity profile as agents choose to work when their productivity is highest. This is one possibility among others with constant productivity as agents are indifferent about the timing of their work decision provided that the present value of their income[5] remains unchanged. However, the retained specification is more problematic with a quadratic productivity profile which should induce agents to also enjoy some leisure at the beginning of their life while their productivity is still low, as in Rogerson Wallenius (2008). It could nevertheless be objected that rising productivity at early ages reflects some on-the-job learning effects and, hence, postponing entry does not increase the starting productivity of a worker. In other words, age 0 is a normalization of the age at which work begins.[6]

The extensive margin is therefore associated to the determination of the retirement age R, which is a continuous choice variable that could be pinned down by a first-order condition. This stands in sharp contrast with the extensive margin of the static optimal taxation literature which, by forbidding employment lotteries, leads to a truly discrete participation decision.

While the above framework has recently been central in the macroeconomic literature

[5] Strictly speaking, it is only with no discounting, $\rho = 0$, that this present value is entirely determined by the fraction of time spent working.

[6] In general, the timing of the work decision is also influenced by the difference between the interest rate and the discount rate and by the time profile of the fixed cost of working.

dedicated to the positive analysis of the effects of taxation, the aim of this paper is to conduct the corresponding normative analysis. But, before specifying the optimal policy problem, I need to determine the informational structure of the economy.

The planner observes output y_t produced at each instant but does not observe the corresponding labor supply l_t; the two being related by $y_t = \gamma_t(\alpha)l_t$ for an α-worker. Instantaneous consumption c_t is also observable which is equivalent to assuming that savings could be monitored and, hence, taxed. Finally, the planner knows the retirement age R of each agent.[7] Full commitment is assumed.

3 Optimal allocation

This section relies on the revelation principle to determine the optimal allocation of resources, while the next section turns to the implementation of the optimal policy in a decentralized economy. Thus, for now, the planner's problem is to design a direct truthful mechanism where each agent is asked to report his type, α, and where telling the truth is the optimal strategy.

A worker claiming to be of type α receives a consumption stream $\{c_t(\alpha)\}_{t\in[0,T]}$, is required to work until age $R(\alpha)$ and needs to produce a flow of output $\{y_t(\alpha)\}_{t\in[0,R(\alpha))}$ while working. Hence, the welfare of an α-worker claiming to be of type α' is given by:

$$V(\alpha';\alpha) = \int_0^H e^{-\rho t}u(c_t(\alpha'))dt - \int_0^{R(\alpha')} e^{-\rho t}\left[v\left(\frac{y_t(\alpha')}{\gamma_t(\alpha)}\right) + b\right]dt, \tag{2}$$

where I have used the fact that an α-worker needs to supply $y_t(\alpha')/\gamma_t(\alpha)$ units of labor to produce output $y_t(\alpha')$. For the mechanism to be truthful, we need:

$$V(\alpha;\alpha) \geq V(\alpha';\alpha), \text{ for all } \alpha \text{ and } \alpha'. \tag{3}$$

An equivalent way of expressing this incentive compatibility condition is that, for any given α, $V(\alpha';\alpha)$ must be maximized when $\alpha' = \alpha$. I therefore impose the necessary first-order condition:

$$\frac{\partial V(\alpha;\alpha)}{\partial \alpha'} = 0, \text{ for all } \alpha. \tag{4}$$

Differentiating (2) with respect to α' and using the fact that[8] $V_1(\alpha';\alpha') = 0$, as implied

[7] With constant productivity, the actual timing of work is not determined, only the total amount of work done is. In this case, I assume that the government knows at any single point in time whether an agent is working or not. But, this might seem to be at odds with the assumption that l_t is not observable. To overcome this difficulty, we can consider that l_t stands for effort while working. Alternatively, I need to assume, following Mulligan (2001), that there is a maximum frequency at which agents can switch between work and leisure and that "the 'indivisibility' is at least as long as the tax accounting period".

[8] $V_i(\alpha';\alpha)$ denotes the derivative of V with respect to its ith argument.

by (4), I obtain:

$$\begin{aligned}\frac{\partial V(\alpha';\alpha)}{\partial\alpha'} &= \int_0^{R(\alpha')} e^{-\rho t}\left[\frac{1}{\gamma_t(\alpha')}v'\left(\frac{y_t(\alpha')}{\gamma_t(\alpha')}\right)-\frac{1}{\gamma_t(\alpha)}v'\left(\frac{y_t(\alpha')}{\gamma_t(\alpha)}\right)\right]\frac{dy_t(\alpha')}{d\alpha'}dt \\ &\quad +e^{-\rho R(\alpha')}\left[v\left(\frac{y_{R(\alpha')}(\alpha')}{\gamma_{R(\alpha')}(\alpha')}\right)-v\left(\frac{y_{R(\alpha')}(\alpha')}{\gamma_{R(\alpha')}(\alpha)}\right)\right]\frac{dR(\alpha')}{d\alpha'}. \end{aligned} \tag{5}$$

The first-order condition (4) characterizes a maximum if and only if $V_1(\alpha';\alpha) > 0$ for $\alpha' < \alpha$ and $V_1(\alpha';\alpha) < 0$ for $\alpha' > \alpha$. The disutility of labor being increasing and convex in the amount of labor supplied, $v(x)$ and $xv'(x)$ are both increasing in x.[9] Also, remember that $\alpha' > \alpha$ implies $\gamma_t(\alpha') \geq \gamma_t(\alpha)$. Hence, the two bracketed terms in (5) have the same sign as $(\alpha' - \alpha)$ whenever $\gamma_t(\alpha') \neq \gamma_t(\alpha)$ and are otherwise equal to zero. This leads to the following lemma.[10]

Lemma 1 *A sufficient condition for the first-order condition (4) to characterize a maximum is*

$$\frac{dy_t(\alpha)}{d\alpha} > 0 \textit{ and } \frac{dR(\alpha)}{d\alpha} \geq 0. \tag{6}$$

Thus, if (6) holds, the very complicated incentive compatibility condition (3) reduces to the much simpler first-order condition (4). Note that this simplification would not be possible if the productivity profiles of different workers were crossing over time. Indeed, the second-order condition (6) implicitly relies on the fact that, in (3), it is always the downward incentive compatibility constraint which is binding. The corresponding economic intuition is that redistribution is typically done from high to low productivity agents; but with crossing profiles it is nor clear who should benefit and who should lose from redistribution.

The lifetime utility of an α-worker who is telling the truth is:

$$V(\alpha) = \int_0^H e^{-\rho t}u(c_t(\alpha))dt - \int_0^{R(\alpha)} e^{-\rho t}\left[v\left(l_t(\alpha)\right)+b\right]dt, \tag{7}$$

where $l_t(\alpha) = y_t(\alpha)/\gamma_t(\alpha)$. Differentiating this function and using the first-order condition (4), the incentive compatibility constraint (3) could be expressed as:

$$V'(\alpha) = \int_0^{R(\alpha)} e^{-\rho t}l_t(\alpha)v'(l_t(\alpha))\frac{1}{\gamma_t(\alpha)}\frac{d\gamma_t(\alpha)}{d\alpha}dt. \tag{8}$$

[9]This implies that the Spence-Mirrlees condition is satisfied.

[10]This sufficient second-order condition could have alternatively been derived by imposing the positivity of the cross derivative of $V(\alpha';\alpha)$ at α, i.e. $V_{12}(\alpha;\alpha) > 0$. Indeed, totally differentiating the first-order condition (4) gives $V_{11}(\alpha;\alpha) + V_{12}(\alpha;\alpha) = 0$ and, hence, the standard second-order condition for a maximum $V_{11}(\alpha;\alpha) < 0$ is equivalent to $V_{12}(\alpha;\alpha) > 0$.

The economy-wide resource constraint is:

$$\int_0^{\bar{\alpha}} \left[\int_0^{R(\alpha)} e^{-\rho t} \gamma_t(\alpha) l_t(\alpha) dt - \int_0^H e^{-\rho t} c_t(\alpha) dt \right] f(\alpha) d\alpha \geq E, \tag{9}$$

where f is the density function of the distribution of the productivity index α across the population with support $[0, \bar{\alpha}]$ and E denotes an exogenous amount of government expenditures that must be financed. The bracketed term on the left-hand-side of (9) corresponds to the budgetary surplus generated by an α-worker. Finally, the planner's objective is to maximize social welfare, expressed as a Bergson-Samuelson functional:

$$\int_0^{\bar{\alpha}} \Psi \left(V(\alpha) \right) f(\alpha) d\alpha, \tag{10}$$

where Ψ is an increasing and concave function weighting the lifetime utility of individuals according to the redistributive objective. Ψ is typically specified as:

$$\Psi(V) = \frac{V^\kappa}{\kappa}, \tag{11}$$

with $\kappa \in (-\infty, 1]$ determining the social aversion to inequality. The two most common benchmark are the utilitarian preferences, $\kappa = 1$, where the planner only cares about the sum of individual utilities without any special concerns about their distribution across the population and the Rawlsian case, $\kappa = -\infty$, where the welfare of society is equal to the utility of the worst-off individual.

Note that, without an intensive margin, the incentive compatibility constraint (8) would boil down to imposing an equal lifetime utility for everyone. Indeed, as could be seen from (2) with v equal to 0, an individual's utility would not be affected by his type and, hence, it would not be necessary to give high productivity workers an informational rent to induce them to reveal their productivity parameter α. It follows that, with an extensive margin only, any second-best allocation is also the first-best allocation with Rawlsian social preferences.

The planner's problem is to maximize social welfare (10) subject to the resource constraint (9) and to the incentive compatibility constraint (8) holding for each α. This gives an optimal control problem with $c_t(\alpha)$ and $l_t(\alpha)$ as control variables and $V(\alpha)$ as the state variable and where $R(\alpha)$ is implicitly determined from (7). It could be solved using Pontryagin's maximum principle.

The first-order conditions to the problem imply that consumption should remain constant throughout the life of individuals, i.e. $c_t(\alpha) = c(\alpha)$ for all t. This is not surprising as, without uncertainty, the inverse Euler equation characterizing the optimal allocation of resources in a dynamic optimal taxation problem is identical to the standard Euler equation (Golosov Kocherlakota Tsyvinski 2003).[11] Thus, the interest rate being equal to the discount rate, there is nothing to be gained by distorting consumption over time. This implies that the optimal policy would not be affected if consumption or, equivalently, savings were not observable.

Let $\lambda > 0$ denote the multiplier associated to the resource constraint and $\mu(\alpha)$ the multiplier associated to the incentive compatibility constraint of the α-worker. The first-order condition to the problem corresponding to the state variable is:

$$-\mu'(\alpha) = \left[\Psi'(V(\alpha)) - \frac{\lambda}{u'(c(\alpha))}\right] f(\alpha). \tag{12}$$

We also have the two transversality conditions:

$$\mu(0) = \mu(\bar{\alpha}) = 0. \tag{13}$$

Note that it could easily be proved that μ is always non-positive (Werning 2000). The first-order condition associated to the intensive margin is:

$$\lambda \left[\gamma_t(\alpha) - \frac{v'(l_t(\alpha))}{u'(c(\alpha))}\right] f(\alpha) + \mu(\alpha) \frac{1}{\gamma_t(\alpha)} \frac{d\gamma_t(\alpha)}{d\alpha} \left[v'(l_t(\alpha)) + l_t(\alpha) v''(l_t(\alpha))\right] = 0. \tag{14}$$

Similarly, the corresponding condition associated to the extensive margin is:

$$\lambda \left[\gamma_{R(\alpha)}(\alpha) l_{R(\alpha)}(\alpha) - \frac{v(l_{R(\alpha)}(\alpha)) + b}{u'(c(\alpha))}\right] f(\alpha) + \mu(\alpha) \frac{1}{\gamma_{R(\alpha)}(\alpha)} \frac{d\gamma_{R(\alpha)}(\alpha)}{d\alpha} l_{R(\alpha)}(\alpha) v'(l_{R(\alpha)}(\alpha)) = 0. \tag{15}$$

We now have a complete characterization of the solution to the planner's problem.

Proposition 1 *The optimal allocation of resources* $\left\{R(\alpha), \{y_t(\alpha)\}_{t\in[0,R(\alpha))}, c(\alpha)\right\}_{\alpha\in[0,\bar{\alpha}]}$ *is characterized by the first-order conditions (12), (13), (14) and (15) together with the constraints of the planner's problem (8) and (9) and the lifetime utility function (7).*

Of course, if the sufficient second-order condition (6) of Lemma 1 is not satisfied, then the above first-order conditions might well be meaningless.

Let us define $\tau^i(\alpha, t)$ as the wedge along the intensive margin for an α-worker of age t as:

$$\gamma_t(\alpha)\left(1 - \tau^i(\alpha, t)\right) = \frac{v'(l_t(\alpha))}{u'(c(\alpha))}. \tag{16}$$

[11] At a deeper level, this is a consequence of the uniform commodity taxation theorem of Atkinson and Stiglitz (1976). Indeed, preferences are separable between consumption and leisure and consumption at different dates could be seen as different commodities which should, therefore, not be taxed differently.

Similarly, I define the extensive wedge $\tau^e(\alpha)$ for an α-worker as:

$$\gamma_{R(\alpha)}(\alpha)l_{R(\alpha)}(\alpha)\left(1-\tau^e(\alpha)\right)=\frac{v(l_{R(\alpha)}(\alpha))+b}{u'(c(\alpha))}. \tag{17}$$

These two equations state that, absent any distortions, i.e. $\tau^i(\bar{\alpha},t)=0$ and $\tau^e(\bar{\alpha})=0$, the marginal product of labor should be equal to the marginal rate of substitution between leisure and consumption where, for the extensive margin, the disutility from retiring marginally later is $v(l_{R(\alpha)}(\alpha))+b$ and the corresponding marginal product is $\gamma_{R(\alpha)}(\alpha)l_{R(\alpha)}(\alpha)$. Simple algebra using the first-order conditions for the intensive and extensive margins, (14) and (15), respectively, reveals that:

$$\tau^i(\alpha,t)=-\frac{\mu(\alpha)}{\lambda f(\alpha)}\frac{1}{[\gamma_t(\alpha)]^2}\frac{d\gamma_t(\alpha)}{d\alpha}\left[v'(l_t(\alpha))+l_t(\alpha)v''(l_t(\alpha))\right], \tag{18}$$

and:

$$\tau^e(\alpha)=-\frac{\mu(\alpha)}{\lambda f(\alpha)}\frac{1}{l_{R(\alpha)}(\alpha)\left[\gamma_{R(\alpha)}(\alpha)\right]^2}\frac{d\gamma_{R(\alpha)}(\alpha)}{d\alpha}l_{R(\alpha)}(\alpha)v'(l_{R(\alpha)}(\alpha)). \tag{19}$$

As $\mu(\bar{\alpha})=0$, the no distortion at the top principle holds along both margins. Similarly, as $\mu(0)=0$, the labor supply of the lowest productivity agent is not distorted provided that there is no bunching at the bottom of the income distribution[12]. Finally, wedges are strictly positive along both margins for any other value of α for which the first-order conditions hold.

Without an intensive margin, the utility of individuals would be independent of their productivity and no-one would get an informational rent. The optimal retirement age would therefore equalize the marginal rate of substitution to the marginal product of labor, i.e. equation (17) would hold with $\tau^e(\alpha)=0$ for all α. Thus, in the present context, even the wedge along the extensive margin is due to the existence of the intensive margin. Distortions are necessary to induce people to reveal their type and it is preferable to have two small distortions rather than a single large one.

4 Implementation in a decentralized economy

Now that I have characterized the optimal allocation, I turn to the description of how it could be implemented in a decentralized economy using realistic fiscal instruments.

Optimal consumption should be constant over life, which naturally occurs when agents can trade a risk-free asset. Capital taxes are therefore not needed, which considerably simplifies the problem. As shown by Weinzierl (2008), a history-independent income

[12] Bunching is likely to occur at low income levels as the optimal allocation for low productivity agents might be characterized by a corner solution imposing that they do not supply any labor.

tax cannot, in general, implement the optimal allocation, even if it is allowed to be age-dependent. The intuition for this is that a direct truthful mechanism implicitly has memory which reduces the amount of distortions needed to raise a given amount of resources; whereas a memory-less income tax is constrained to create distortions in every time period. To implement the optimum, we therefore need a fiscal instrument which is history-dependent until, at least, the retirement age. A natural candidate is a social security system which, in many countries, already takes the history of labor supply into account to determine the level of pensions.

Let us now solve the implementation problem.[13] I denote the optimal allocation by $\left\{R^*(\alpha), \{y_t^*(\alpha)\}_{t\in[0,R^*(\alpha))}, c^*(\alpha)\right\}_{\alpha\in[0,\bar{\alpha}]}$. To lighten notations, let y^R stand for a given history of labor supply, i.e. $y^R = \left\{R, \{y_t\}_{t\in[0,R)}\right\}$, and $y^{R*}(\alpha)$ stand for the optimal history of the α-worker, i.e. $y^{R*}(\alpha) = \left\{R^*(\alpha), \{y_t^*(\alpha)\}_{t\in[0,R^*(\alpha))}\right\}$. Let us define DOM as the set of labor supply histories compatible with a socially optimal allocation. More formally:

$$DOM = \left\{y^R : y^R = y^{R*}(\alpha) \text{ for some } \alpha \in [0,\bar{\alpha}]\right\} \tag{20}$$

We now define the function $\hat{c} : DOM \to \mathbb{R}$ such that:

$$\hat{c}(y^{R*}(\alpha)) = c^*(\alpha). \tag{21}$$

The second-order condition (6) of Lemma 1 implies that this function always exists. To make the implementation problem as simple as possible, I assume for now that agents get all their lifetime income when they retire. This social security payment received by workers at retirement is set equal to:

$$Q^*(y^R) = \begin{cases} e^{\rho R}\hat{c}(y^R)\frac{1-e^{-\rho H}}{\rho} & \text{if } y^R \in DOM \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

This solves the implementation problem.

Proposition 2 *The social security system Q^* implements the optimal allocation $\left\{y^{R*}(\alpha), c^*(\alpha)\right\}_{\alpha\in[0,\bar{\alpha}]}$.*

Proof. First, adopting a labor supply strategy y^R outside DOM cannot be individually rational as 0 consumption at any point in life generates a lifetime utility of $-\infty$. Let $y^{R*}(\alpha')$, with $\alpha' \in [0,\bar{\alpha}]$, be the labor supply strategy of an α-worker. By construction,

[13] The presentation is closely related to that of Grochulski and Kocherlakota (2008).

$y^{R*}(\alpha') \in DOM$. The α-worker will choose his consumption level by solving:

$$\max_{c_t} \int_0^H e^{-\rho t} u(c_t) dt - \int_0^{R(\alpha')} e^{-\rho t} \left[v\left(\frac{y_t(\alpha')}{\gamma_t(\alpha)} \right) + b \right] dt \tag{23}$$
$$\text{subject to } e^{-\rho R(\alpha')} Q^*(y^{R*}(\alpha')) \geq \int_0^H e^{-\rho t} c_t dt.$$

The solution to the problem implies a constant consumption level, which, from the budget constraint, must be equal to:

$$\frac{\rho e^{-\rho R}}{1 - e^{-\rho H}} Q^*(y^{R*}(\alpha')). \tag{24}$$

But, by definition of the social security system, (20), (21) and (22), this is just $\hat{c}(y^{R*}(\alpha')) = c^*(\alpha')$. It follows that choosing among $\{y^R, c\}$ given that $y^R \in DOM$ is equivalent to choosing among reporting strategies in a direct truthful mechanism. An α-worker therefore chooses $y^{R*}(\alpha)$ for his labor supply and consumes $c^*(\alpha)$. ■

Although it is commonly argued that redistribution should be one of the main objectives of a well designed pension system (Barr Diamond 2008), there is little theoretical justification for this. In particular, it is *a priori* not clear that an optimal income tax is not sufficient to achieve the desired level of redistribution. Proposition 2 contributes to this debate by implying that, indeed, equity concerns should be dealt with within an optimally designed social security system with endogenous retirement.

I shall now illustrate the fact that Q^* could be seen as a reduced form of a more realistic social security system. Current policies are typically designed such that individuals pay income taxes throughout their career and receive an annuitized pension after retirement.

Proposition 3 *For any income tax function T, the optimal policy can be implemented by giving retirees an annuitized pension P^*, where:*

$$P^*(y^R) = \begin{cases} \frac{\rho}{e^{-\rho R} - e^{-\rho H}} \left[\hat{c}(y^R) \frac{1 - e^{-\rho H}}{\rho} - \int_0^R e^{-\rho t} \left[y_t - T(y_t, t) \right] dt \right] & \text{if } y^R \in DOM \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

Proof. Choosing $y^R \notin DOM$ is still not desirable. For $y^R \in DOM$, the combination of the income taxes T and of the annuitized pensions P^* satisfies:

$$\int_0^R e^{-\rho t} \left[y_t - T(y_t, t) \right] dt + \int_R^H e^{-\rho t} P^*(y^R) dt = e^{-\rho R} Q^*(y^R). \tag{26}$$

So, the worker's budget constraint is not affected by the change from Q^* to (T, P^*) and, hence, (T, P^*) also implements the optimal allocation. ■

Clearly, the proposed policy is not fully identified. In particular, any income tax change could be offset within the social security system such as to leave the resulting allocation unchanged.

It has been extensively argued in the static optimal taxation literature, that the existence of an extensive margin to labor supply justifies the implementation of tax credits such as the Earned Income Tax Credit of the US or the Working Tax Credit of the UK. This conclusion no longer holds in a dynamic context.[14] Indeed, in the above framework, a tax credit is, at best, inconsequential, if its effects are undone after retirement, and, otherwise, unambiguously suboptimal. To understand this result, note that, as the participation decision ceases to be discrete, it is no longer possible, with an extensive margin only, to induce some to work more, thanks to a generous tax credit, without causing adverse effects on the labor supply decision of other harder working individuals.

I have so far assumed that agents can trade a risk-free asset at the exogenous interest rate ρ. If necessary, they can even use their future social security payment as a collateral to be able to borrow sufficiently to achieve perfect consumption smoothing. If, on the contrary, agents do not have a perfect access to the credit market, then the optimal policy can be fully identified.

Proposition 4 *If capital markets are dysfunctional and only the government can borrow and lend at the interest rate ρ, then the unique optimal policy is (T^*, P^*) with the optimal age-dependent income tax determined by:*

$$T^*(y_t^*(\alpha), t) = y_t^*(\alpha) - c^*(\alpha). \tag{27}$$

Proof. The optimal income tax function T^* is well defined whenever the condition $\frac{dy_t^*(\alpha)}{d\alpha} > 0$ of Lemma 1 holds. By construction, (T^*, P^*) is the only optimal policy which ensures perfect consumption smoothing without individuals trading any asset. ■

When thinking about the policy relevance of the proposed social security system, an important limitation is that we do not know what should be done if agents fail to be supply an optimal amount of labor supply, i.e. if their y^R fails to be in DOM. Clearly, to address this issue, the present framework would need to be enriched with features that could explain such outcomes. It could nevertheless be conjectured that, whether workers fail to choose $y^R \in DOM$ because of uncertainties such as skill risks or because of limited

[14] Strictly speaking, a tax credit is only justified in a static context if agents differ in their fixed cost of working (Chone Laroque 2008). Although solving such a multidimensional screening problem is beyond the scope of this paper, the resulting optimal allocation, which would be indexed by the productivity α and the fixed cost of working b of individuals, would still be implementable by a social security system similar to Q^* provided that the function $\hat{c}$ would still exist.

cognitive capacities, the unlikely labor supply histories would be penalized. Indeed, this would improve incentives to work at little cost in terms of welfare. Determining the robustness of optimal policies to modeling uncertainties remains an important issue for further research.

5 Simulation

I now simulate the optimal policy for a reasonable calibration of the model. Individuals can work from age 25 until they die on their 80th birthday. The annual discount rate is 2%; so $\rho = 0.02$. The disutility from supplying labor along the intensive margin is given by a standard power function:

$$v(l_t) = \frac{l_t^{1+\frac{1}{\delta}}}{1+\frac{1}{\delta}}, \tag{28}$$

where δ is the constant intertemporal elasticity of substitution. Following Kleven Kreiner Saez (2008) and Brewer Saez Shephard (2008), I take $\delta = 0.25$.[15] This reflects the low intensive elasticity of labor supply well documented in the empirical literature. The instantaneous utility derived from consumption is logarithmic:

$$u(c_t) = \log(c_t). \tag{29}$$

Note that, preferences being separable between consumption and leisure, this logarithmic specification is required to have the number of hours worked and the retirement age unaffected by the productivity level α when the government does not intervene.[16]

The productivity profile of an α-worker is proportional to a baseline productivity profile γ_t, the proportion being given by his productivity index α; thus $\gamma_t(\alpha) = \alpha\gamma_t$. The baseline profile is such that productivity is constant and normalized to 1 until age 60 and then declines smoothly and quadratically until it reaches 0 at 80. This is consistent with the fact that, under the current fiscal system, the number of hours worked by participating workers is almost constant until age 60 (see Prescott Rogerson Wallenius 2009, Figure 2). Furthermore, it also explains why some people, those who do not wish to retire after age 60, currently choose to alternate spells of employment and leisure rather than to enjoy all of their leisure after some early retirement age.[17] The distribution of the productivity

[15] This value also falls in the middle of the range of elasticities considered by Rogerson and Wallenius (2008).

[16] Similarly, in a Ramsey model with technological progress, logarithmic utility of consumption is needed to obtain a balanced growth path with constant labor supply.

[17] From the normative perspective of this paper, although I impose that individuals work continuously until retirement, it would be equally desirable to allow the low productivity workers who retire before 60 to alternate spells of employment and leisure provided that the present value of their production remains unchanged. To implement these alternative optimum allocations, the social security system (22) would

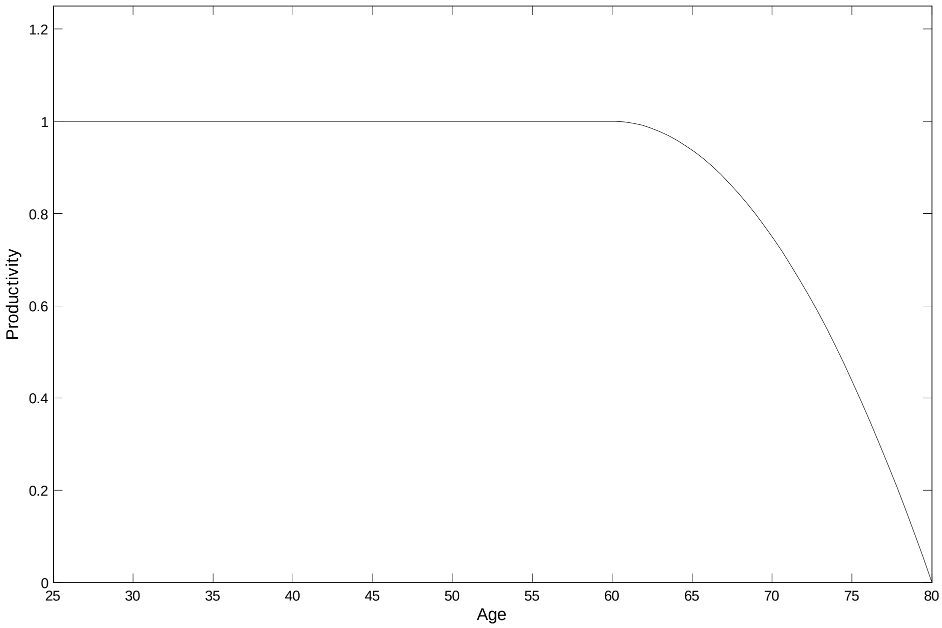

Figure 1: Baseline productivity profile

index, $f(\alpha)$, is lognormal. The mean is normalized to 1 and the standard deviation is set at 0.7, an empirically plausible value according to Kanbur and Tuomala (1994). The baseline productivity profile and the lognormal distribution of α are plotted in Figure 1 and 2, respectively.

The planner maximizes a utilitarian social welfare function; thus $\kappa = 1$ in (11). Finally, the fixed cost of working b is calibrated such that the average retirement age of participating workers is 62 and the level of government expenditures E is calibrated such that it amounts to a quarter of total output. To simulate the optimal policy, I just solve a discretized version of the first-order conditions characterizing the optimal allocation.

Let us now turn to the corresponding results. Figure 3 displays the lifetime production and consumption of workers as a function of their productivity index. The least productive individuals, those with $\alpha < 0.26$, never participate to the labor market. They only represent 3.3% of the population. Lifetime consumption exceeds production for about a third of workers, 35.9%; those whose productivity index α falls below 0.65. The most productive agents consume slightly more than 50% of their output. Figure 3 suggests that there is hardly any progressivity in the optimal fiscal system.

Figure 4 shows the budget surplus raised from each type of workers, i.e. the difference between the lifetime production and consumption of an α-worker multiplied by the

have to be changed slightly by setting, for instance, age 60 as the legal retirement age before which no history-dependent transfer could be made.

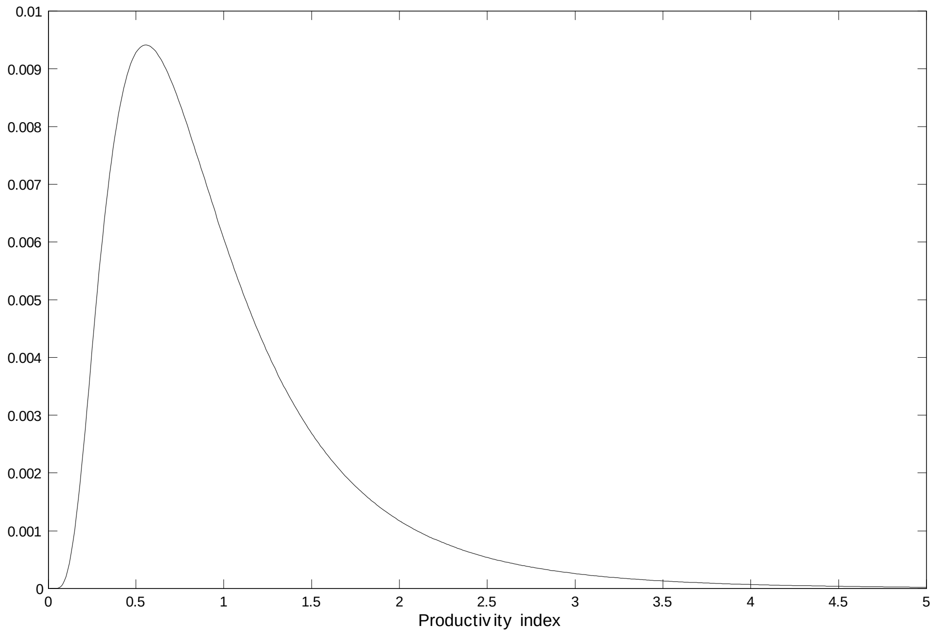

Figure 2: Lognormal distribution of the productivity index α

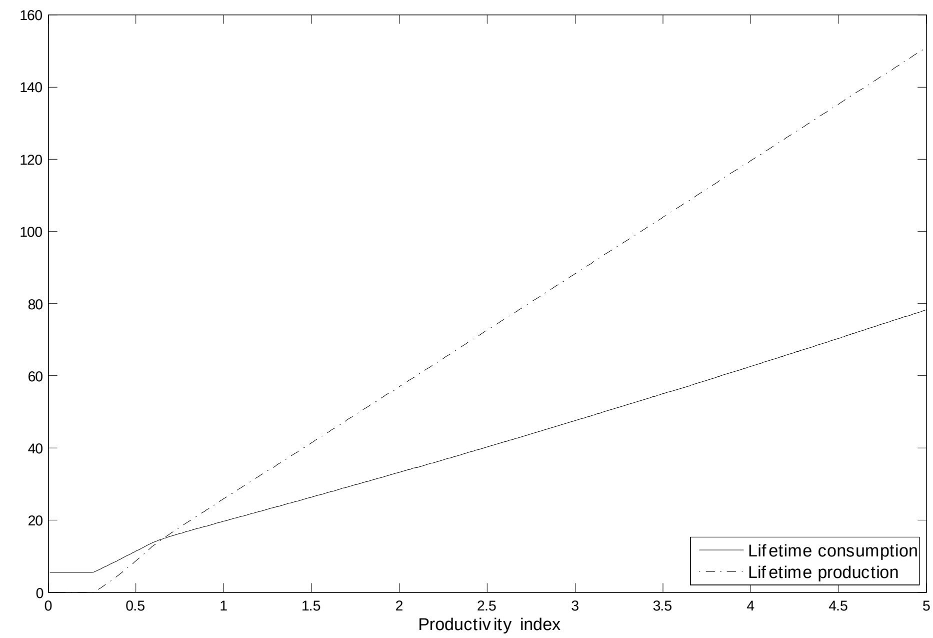

Figure 3: Lifetime production and consumption as a function of the productivity index α

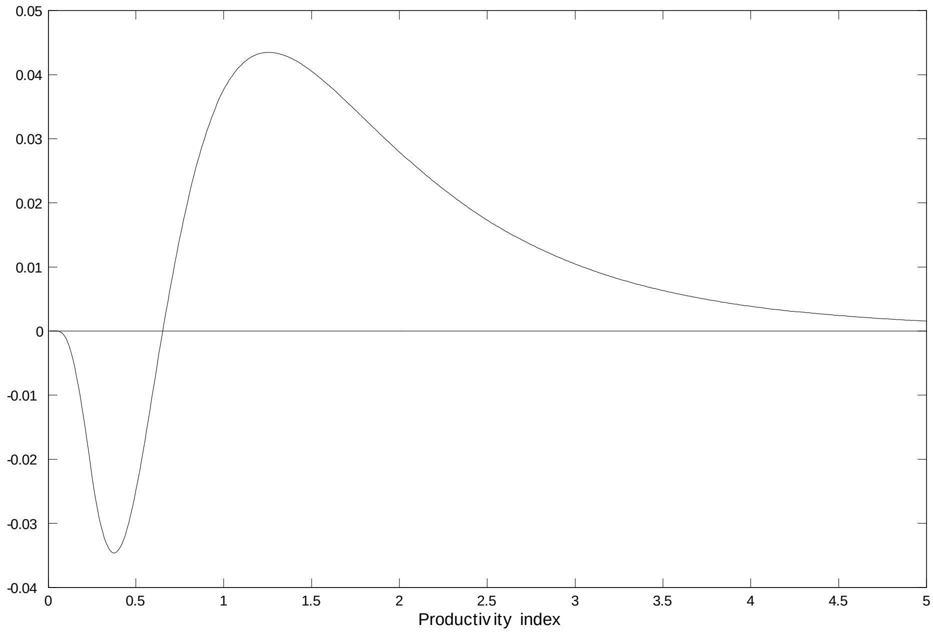

Figure 4: Budget surplus raised from each type of workers

number $f(\alpha)$ of such workers. This illustrates the well-known fact that the bulk of redistribution occurs from the upper-middle class to the lower-middle class. This is simply because the very rich and very poor are not very numerous. As could be seen from Figure 4, the total surplus across the whole population is positive. This is necessary to finance the government expenditures E which amount to a quarter of total output.

Unsurprisingly given the low intensive elasticity, the labor supply of participating workers is not very sensitive to productivity. It is equal to 0.79 for the least productive agent who participates, for whom $\alpha = 0.26$, a varies between 0.91 and 1.02 for an $\alpha = 5$ worker whose productivity while working fluctuates between 0.63 and 1. Hence, although pretty constant, labor supply is slightly increasing in the productivity index α as well as in the age-specific productivity of an α-worker.

A large part of the variation of the labor supply across agents is associated with the extensive margin. Figure 5 displays the retirement age of the different types of workers. As expected, it is desirable to have the career length of individuals increasing in their productivity. Indeed, the high productivity agents with $\alpha > 3.4$ retire after age 72, more than 10 years later than the average retirement age of participating workers which was set at[18] 62. Figure 6 shows the distribution of the retirement age across the population. Only 29.9% of individuals, of which 3.3% never work, retire before age 60, i.e. before

[18] The average retirement age for the whole population, including the 3.3% of agents who never work, is 60.8.

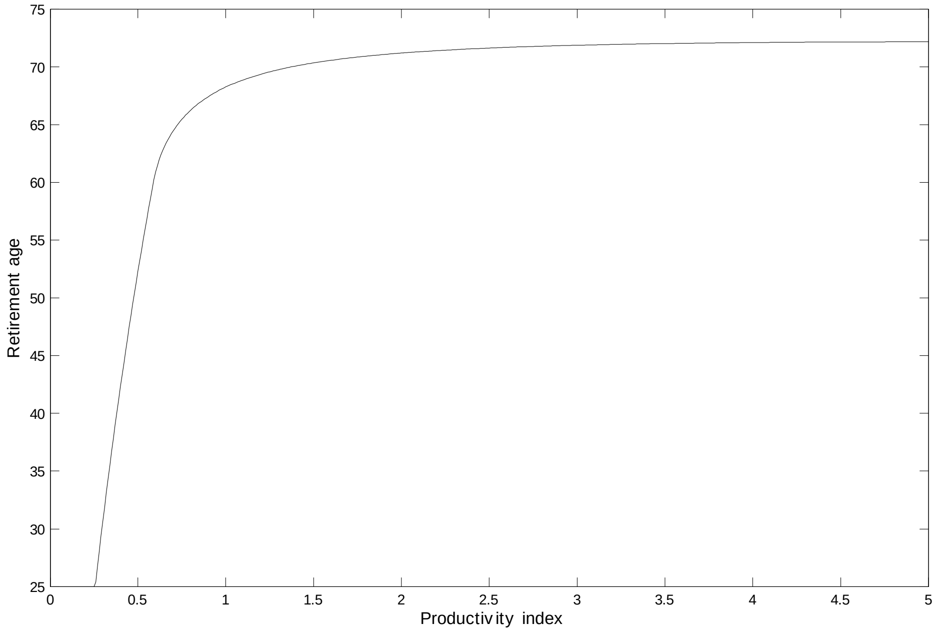

Figure 5: Retirement age as a function of the productivity index α

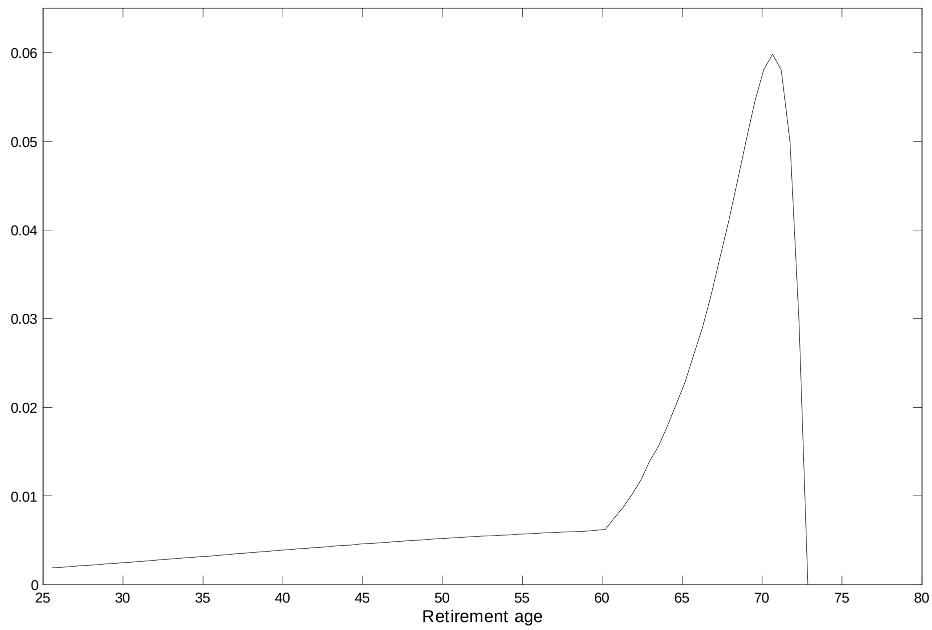

Figure 6: Distribution of the retirement age (a mass of 3.3% at age 25, corresponding to non-participating workers, is omitted from the graph)

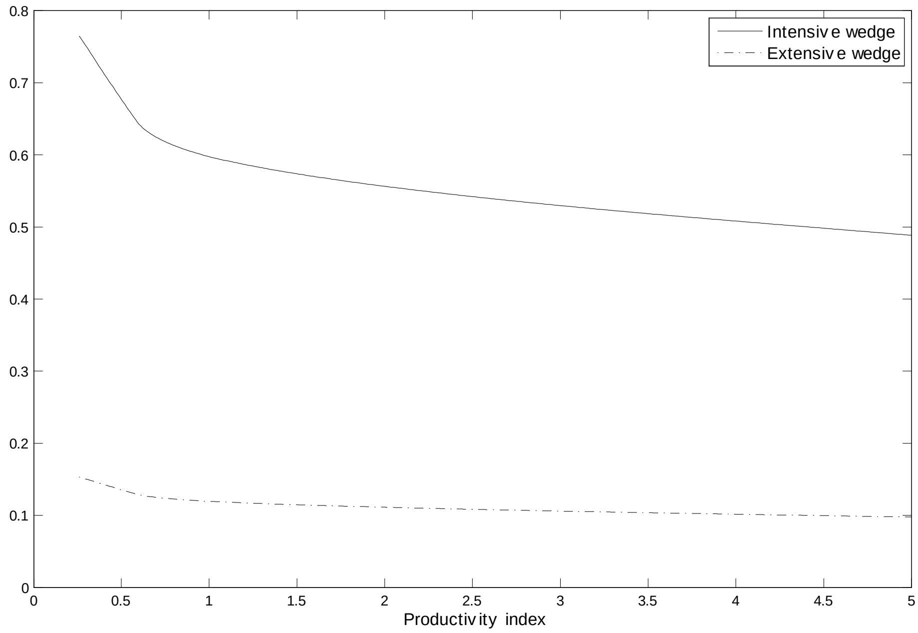

Figure 7: Intensive and extensive wedges as a function of the productivity index α

their productivity starts declining.

How do the wedges along the intensive and extensive margins compare? It turns out that, with a constant intertemporal elasticity substitution, as implied by (28), the relationship between the intensive, $\tau^i(\alpha, t)$, and the extensive, $\tau^e(\alpha)$, wedge satisfies:

$$\frac{\tau^i(\alpha, t)}{\tau^e(\alpha)} = 1 + \frac{1}{\delta}. \tag{30}$$

This is immediately obtained by dividing (18) by (19), after having plugged in (28). Hence, the lower is the elasticity of labor supply along the intensive margin, the higher should the intensive wedge be relative to the extensive wedge. The intuition is reminiscent of Ramsey's (1927) inverse elasticity rule: a low elasticity implies that a large wedge will only lead to a small behavioral response. In the extreme case where $\delta = 0$, all the burden falls on the intensive margin which, *de facto*, does not exist as participating workers always supply exactly one unit of labor.[19] Note that (30) implies that the intensive wedge faced by an α-worker is independent of his age. Figure 7 reports the wedges of participating workers. With $\delta = 0.25$, the intensive wedge is five times larger than the extensive wedge.

[19] In the opposite extreme where $\delta = +\infty$, both wedges are equal. With $v(l_t) + b = l_t + b$ the labor supply is equally responsive along both margins. Indeed, the worker can practically avoid paying the fixed cost of working by supplying all his labor at age 25.

6 Conclusion

In this paper, I have characterized the optimal redistributive policy in a dynamic framework with an intensive and an extensive margin to labor supply. My results advocate for the implementation of a history-dependent social security system which induces a positive correlation between the productivity of workers and their retirement age. Thus, rather than providing some support for tax credits, my analysis suggests that an important amount of redistribution should be done within the social security system. While I have not quantified the welfare gains to be expected from the implementation of the optimal policy, they must be at least as large as those generated by an optimal age-dependent income tax which were evaluated, by Weinzierl (2008), to be close to 2% of aggregate consumption in the US.

Due to the looming pension crisis, policy makers are starting to realize that, sooner or later, the retirement age will need to be raised. This creates a unique opportunity to reform social security systems and this work suggests that, rather than imposing an homogeneous increase in career length across the population, a well designed reform should encourage higher productivity people to retire later.

A number of issues remain for further research. It would be interesting to solve for the optimal policy when workers are heterogeneous in both productivity and fixed costs of working. This remains, however, a non-trivial multidimensional screening problem.[20] Also, I have abstracted from skill risks, which are at the heart of the recent dynamic optimal taxation literature. In particular, allowing for the random occurrence of a permanent disability shock, as in Diamond Mirrlees (1978) or Denk Michau (2008), seems particularly relevant for the optimal design of social security.

References

[1] Albanesi, S. and Sleet, C. (2006), 'Dynamic Optimal Taxation with Private Information', *Review of Economic Studies*, 73(1), 1-30.

[2] Atkinson, A.B. and Stiglitz, J.E. (1976), 'The Design of Tax Structure: Direct versus Indirect Taxation', *Journal of Public Economics*, 6, 55-75.

[3] Barr, N. and Diamond, P.A. (2008), *Reforming Pensions: Principles and Policy Choices*, New York and Oxford: Oxford University Press.

[20] Beaudry, Blackorby and Szalay (2007) manage to solve an optimal taxation problem with two dimensions of heterogeneity and observable labor supply. Formally, this is closely related to the dynamic problem of this paper with an extensive margin only. However, they hugely simplify their problem by assuming that individual utility is, *de facto*, linear in consumption and leisure.

[4] Battaglini, M. and Coate, S. (2008), 'Pareto Efficient Income Taxation with Stochastic Abilities', *Journal of Public Economics*, 92, 844-868.

[5] Beaudry, P., Blackorby, C. and Szalay, D. (2007), 'Taxes and Employment Subsidies in Optimal Redistribution Programs', *American Economic Review*, Forthcoming.

[6] Brewer, M., Saez, E. and Shephard, A. (2009), 'Means-Testing and Tax Rates on Earnings', Working Paper, Institute for Fiscal Studies.

[7] Blundell, R. and Shephard, A. (2008), 'Employment, Hours of Work and the Optimal Design of Earned Income Tax Credits', Working Paper, Institute for Fiscal Studies.

[8] Chone, P. and Laroque, G. (2005), 'Optimal Incentives for Labor Force Participation', *Journal of Public Economics*, 89, 395-425.

[9] Chone, P. and Laroque, G. (2008), 'Optimal Taxation in the Extensive Model', Working Paper, INSEE-CREST.

[10] Cremer, H., Lozachmeur, J.M. and Pestieau, P. (2004), 'Social Security, Retirement Age and Optimal Income Taxation', *Journal of Public Economics*, 88, 2259-2281.

[11] Denk, O. and Michau, J.B. (2008), 'Optimal Social Security with Imperfect Tagging', Working Paper, London School of Economics.

[12] Diamond, P.A. (1980), 'Income Taxation with Fixed Hours of Work', *Journal of Public Economics*, 13, 101-110.

[13] Diamond, P.A. (2003), *Taxation, Incomplete Markets, and Social Security*, Cambridge, MA: MIT Press.

[14] Diamond, P.A. and Mirrlees, J.A. (1978), 'A Model of Social Insurance with Variable Retirement', *Journal of Public Economics*, 10, 295-336.

[15] Golosov, M., Kocherlakota, N. and Tsyvinski, A. (2003), 'Optimal Indirect and Capital Taxation', *Review of Economic Studies*, 70(3), 569-587.

[16] Golosov, M. and Tsyvinski, A. (2006), 'Designing Optimal Disability Insurance: A Case for Asset Testing', *Journal of Political Economy*, 114(2), 257-279.

[17] Golosov, M., Tsyvinski, A. and Werning, I. (2006) 'New Dynamic Public Finance: A User's Guide', in *NBER Macroeconomics Annual 2006*, edited by Daron Acemoglu, Kenneth Rogoff and Michael Woodford, Cambridge, MA: MIT Press.

[18] Gorry, A. and Oberfield, E. (2008), 'Optimal Taxation over the Life Cycle', Working Paper, University of Chicago.

[19] Grochulski, B. and Kocherlakota, N.R. (2008), 'Nonseparable Preferences and Optimum Social Security Systems', Working Paper, University of Minnesota.

[20] Hansen, G.D. (1985), 'Indivisible Labor and the Business Cycle', *Journal of Monetary Economics*, 16, 309-327.

[21] Immervoll, H., Kleven, H.J., Kreiner, C.T. and Saez, E. (2007), 'Welfare Reforms in European Countries: A Microsimulation Analysis', *Economic Journal*, 117(516), 1-44.

[22] Kanbur, R. and Tuomala, M. (1994), 'Inherent Inequality and the Optimal Graduation of Marginal Tax Rates', *Scandinavian Journal of Economics*, 96(2), 275-282.

[23] Kapicka, M. (2006), 'Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping', *Review of Economic Dynamics*, 9(4), 612-639.

[24] Kleven, H.J., Kreiner, C.T. and Saez, E. (2008), 'The Optimal Income Taxation of Couples', *Econometrica*, 77(2), 537-560.

[25] Kocherlakota, N.R. (2005), 'Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation', *Econometrica*, 73(5), 1587-1621.

[26] Laroque, G. (2005), 'Income Maintenance and Labor Force Participation', *Econometrica*, 73(2), 341-376.

[27] Ljungqvist, L. and Sargent, T. (2006), 'Do Taxes Explain European Unemployment? Indivisible Labor, Human Capital, Lotteries, and Savings', in *NBER Macroeconomics Annual 2006*, edited by Daron Acemoglu, Kenneth Rogoff and Michael Woodford, Cambridge, MA: MIT Press.

[28] Ljungqvist, L. and Sargent, T. (2008), 'Taxes, Benefits, and Careers: Complete versus Incomplete Markets', *Journal of Monetary Economics*, 55, 98-125.

[29] Ljungqvist, L. and Sargent, T. (2009), 'Curvature of Earnings Profile and Careers Length', Working Paper, New York University.

[30] Mirrlees, J. (1971), 'An Exploration in the Theory of Optimal Income Taxation', *Review of Economic Studies*, 28(2), 175-208.

[31] Mulligan, C. (2001), 'Aggregate Implications of Indivisible Labor', *Advances in Macroeconomics*, 1(1).

[32] Prescott, E.C. (2002), 'Prosperity and Depression', *American Economic Review*, 92(2), 1-15.

[33] Prescott, E.C. (2004), 'Why Do Americans Work So Much More Than Europeans', *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(1), 2-13.

[34] Prescott, E.C. (2006), 'Nobel Lecture: The Transformation of Macroeconomic Policy and Research', *Journal of Political Economy*, 114(2), 203-235.

[35] Prescott, E.C., Rogerson, R. and Wallenius, J. (2009), 'Lifetime Aggregate Labor Supply with Endogenous Workweek Length', *Review of Economic Dynamics*, 12(1), 23-36.

[36] Ramsey, F.P. (1927), 'A Contribution to the Theory of Taxation', *Economic Journal*, 37(145), 47-61.

[37] Rogerson, R. (1988), 'Indivisible Labor, Lotteries and Equilibrium', *Journal of Monetary Economics*, 21, 3-16.

[38] Rogerson, R. and Wallenius, J. (2008), 'Micro and Macro Elasticities in a Life Cycle Model with Taxes', *Journal of Economic Theory*, Forthcoming.

[39] Saez, E. (2002), 'Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses', *Quarterly Journal of Economics*, 117(3), 1039-1073.

[40] Sheshinski, E. (2008), 'Optimum Delayed Retirement Credit', in *Pension Strategies in Europe and the United States*, edited by Robert Fenge, Georges de Menil and Pierre Pestieau, Cambridge, MA: MIT Press.

[41] Vickrey, W. (1939), 'Averaging of Income for Income-Tax Purposes', *Journal of Political Economy*, 47(3), 379-397.

[42] Weinzierl, M. (2008), 'The Surprising Power of Age-Dependent Taxes', Working Paper, Harvard.

[43] Werning, I. (2000), 'An Elementary Proof of Positive Optimal Marginal Taxes', Working Paper, MIT.