Optimal Redistribution with Intensive and Extensive Labor Supply Margins: A Life-Cycle Perspective

Jean-Baptiste Michau[*]

Ecole Polytechnique

July 2011

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 life-cycle context where workers choose the fraction of their lifetime that they spend working. In this paper, I therefore characterize the optimal redistribution policy in a life-cycle framework with both an intensive and an extensive margin of labor supply. The government should optimally design a history-dependent social security system which induces higher productivity individuals to retire later. Some redistribution therefore needs to be done through the pension system; a standard non-linear income tax is not enough.

Keywords: Extensive margin, Optimal redistribution, Retirement age, Social security

JEL Classification: E62, H21, H55, J26

1 Introduction

Any redistribution policy should provide resources to the poor while preserving incentives to work for higher productivity workers. Thus, to characterize the optimal trade-off between equity and efficiency, it is crucial to rely on a good model of the labor supply. In particular, it is now widely recognized that workers respond to incentives along two

margins: an intensive and an extensive margin. The intensive margin determines the number of hours, or the intensity, of work of participating workers. The extensive margin, due to the existence of fixed costs of working, determines whether individuals choose to participate or not.

While in a static setup, the participation choice is inherently discrete, i.e. to work or not to work, such is not the case in a life-cycle framework where agents choose the fraction of their lifetime that they spend working or, equivalently, their retirement age. This difference is due to the fact that fixed costs of working introduce a non-convexity into workers' labor supply problem which, in the absence of employment lotteries, can only be convexified over time by alternating spells of employment and leisure. Hence, acknowledging the life-cycle nature of workers' labor supply problem considerably strengthens Vickrey's (1939) case[1] for adopting a life-cycle perspective on the optimal redistribution problem.

In this paper, I therefore characterize the optimal redistribution policy in a life-cycle framework where workers are heterogeneous in productivity. I allow for two dimensions to labor supply: 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. I then turn to the implementation of the optimum in a decentralized economy. Finally, I calibrate the model in order to illustrate numerically the key features of the optimal policy.

The main policy recommendations are as follows. Higher productivity workers should have longer careers. 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. In fact, the optimal allocation cannot be implemented by a standard history-independent nonlinear income tax alone. This implies that some redistribution needs to be done 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.

1.1 Related Literature

Mirrlees (1971) solved the optimal redistribution problem in a static environment with an intensive margin only. More recently, the consequences of adding an extensive margin to that framework have been analyzed rather extensively (see, for instance, Diamond 1980, Saez 2002, Immervoll Kleven Kreiner Saez 2007, Chone Laroque 2005, 2011, Laroque 2005, Beaudry Blackorby Szalay 2009, Brewer Saez Shephard 2010, Blundell Shephard

---

[1] Vickrey's concern was that, for a given lifetime income, taxation should be neutral with respect to the point in time when income is realized. In an empirical investigation of this proposal, Liebmen (2002) showed that basing taxation on lifetime, rather than annual, income can reduce the deadweight loss of taxation by up to 11 percent.

2011). Importantly, this literature has provided some support for the implementation of a tax credit, such as the Earned Income Tax Credit in the US, which reduces the labor supply distortions induced by redistribution. However, abstracting from the life-cycle dimension of workers' labor supply problem is more than a simplifying assumption. Indeed, it fundamentally changes the nature of the participation decision by making it discrete.

The issue of the optimal design of a social security system with heterogeneous agents and endogenous retirement has, so far, been largely overlooked. Two important exceptions include the pioneering work of Diamond (2003, chapter 6) and of Sheshinski (2008). In both cases, agents are heterogeneous in their fixed disutility cost of working rather than in their productivity. The main finding is that agents with a low fixed cost retire later than others and some of the income generated by their extra activity is redistributed to those having a high fixed cost. However, in both cases, the result is derived within a three period model, not suitable for a quantitative exercise, and the authors do not describe how the optimal allocation could be implemented in a decentralized economy.

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.

Laroque (2009) determines the optimal taxation of income in a life-cycle model with an extensive margin only. He obtains the same labor income tax schedule as in a corresponding static analysis, except that the social weights should depend on lifetime, rather than current, income. However, a crucial difference with the approach of this paper is that he does not assume a fixed utility cost of working but, instead, a fixed productivity cost of working. This implies that, even in a life-cycle framework, the participation decision remains discrete, i.e. a worker participates at a given age if and only if his productivity net of the fixed cost at that age is positive. In other words, the fixed cost does not introduce a non-convexity into workers' labor supply problem which would have induced them to choose to work for a fraction of their lives.[2]

While I focus on redistribution, some work has been done on the optimal financing of an exogenous stream of government expenditures in a life-cycle context. Erosa and Gervais (2002) restrict their analysis to linear taxes and show that, if labor income taxes could not be decreasing with age, then taxing capital is a desirable, albeit imperfect, substitute. Gorry and Oberfield (2010) solve for the optimal taxation of a single agent who has both an intensive and an extensive labor supply margin (the latter induces him

---

[2]In the words of Ljungqvist and Sargent (2006), Laroque (2009) does not have a "time averaging" model of the labor supply.

to choose to participate for a fraction of his life). Importantly, the only fiscal instrument allowed is a standard history-independent non-linear income tax. Hence, the policy which they derive is only constrained optimal, which explains why the "no distortion at the top" principle does not hold in their context.

Finally, there has recently been major developments in dynamic optimal taxation with heterogeneous agents (see Kocherlakota 2010 for a comprehensive survey). While this literature builds on Mirrlees (1971), its main focus has not been on redistribution policies but, instead, on the optimal provision of insurance against skill risks. The main corresponding results are about savings distortions, not about 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. Quantitative analyses of labor supply distortions have nevertheless been performed under some special circumstances. For instance, Albanesi Sleet (2006) assumes independently and identically distributed productivity shocks, in Farhi Werning (2010) productivity follows an AR(1) process, Diamond Mirrlees (1978), Golosov Tsyvinski (2006) and Denk Michau (2010) only allow for permanent disability shocks, Golosov Troshkin Tsyvinski (2010) and Weinzierl (2011) focus on two or three period models and Kapicka (2008) does not allow for savings. My paper complements this literature by determining the optimal labor supply distortions in a life-cycle context without uncertainty.

Some of the most important results of this New Dynamic Public Finance literature are about the implementation of optimal allocations in decentralized economies. In particular, Grochulski and Kocherlakota (2010) have shown, in a very general context, that the implementation problem could be solved with a history-dependent social security system. My presentation of the optimal pension system builds on some of their insights.

I begin by presenting, in section 2, the structure of the economy and the corresponding labor supply model. The optimal incentive-feasible allocation of resources is derived in section 3. I then characterize in section 4 a history-dependent social security system which implements 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 $u'(.) > 0$, $u''(.) < 0$, $\lim_{c \to 0+} u(c) = -\infty$ and $\lim_{c \to 0+} u'(c) = +\infty$. They work from age 0 until a retirement age R and get disutility $v(l_t)$ from supplying l_t units of labor at age 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

independent of age. Lifetime utility V is time separable and the future is discounted at rate ρ. Individuals therefore 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 H. The continuous time specification makes it possible to rely on a first-order condition to determine the retirement age R.

The lifetime utility function (1) entails both an intensive and an extensive margin of labor supply. Clearly, conditional on working, agents need to choose a number of hours, or an intensity, l_t of work; this is the intensive margin. As the disutility cost of working $v(.)$ is increasing and convex, in the absence of a fixed cost $b > 0$ of working, agents would choose to work until the end of their lives, i.e. $R = H$. However, the fixed cost creates an indivisibility which induces agents to choose to work for only a fraction R/H of their lives; this is the extensive margin.

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. Productivity $\gamma_t(\alpha)$ is differentiable in both α and t. As will become clear, I need to assume that productivity $\gamma_t(\alpha)$ at each age t is weakly increasing in the productivity parameter α of the agent. More formally, $\alpha > \alpha'$ implies $\gamma_t(\alpha) \geq \gamma_t(\alpha')$ for all t with a strict inequality for at least one t.[3] Thus, the deterministic productivity profiles of two agents are not allowed to cross at any point in time. The distribution of the productivity index α across the population is given by the p.d.f. $f(.)$ with support $[0, \bar{\alpha}]$, where the lower bound of this support is normalized to 0.[4] Resources can be transferred across time at an exogenous interest rate which, for simplicity, is assumed to be equal to the discount rate ρ.

The above specification assumes that agents choose to work at the beginning of their lives, from age 0 to R, and to retire at the end, from R to H. In fact, with a constant fixed cost of working and an interest rate equal to the discount rate, the timing of participation is fully determined by workers' productivity profile: workers want to participate when their productivity is highest. Thus, if productivity profiles follow an inverted U-shape, then agents should choose to enjoy some leisure at the beginning of their lives while

---

[3] A natural candidate specification, which I subsequently use to calibrate the model, is to have a baseline productivity profile γ_t, common to all workers, multiplied by the individual-specific productivity index α, i.e. $\gamma_t(\alpha) = \alpha \gamma_t$.

[4] Alternatively, I could have assumed that all agents share the same productivity profile but have heterogeneous fixed costs of working. However, simultaneously allowing for both sources of heterogeneity would lead to a mutli-dimensional screening problem which would not simplify as much here as in a static context since, in a life-cycle framework, it is not possible to summarize workers' participation decision by comparing their fixed cost of working to a productivity-specific threshold.

their productivity is still low, as in Rogerson Wallenius (2009). However, note that rising productivity at early ages is likely to be due to on-the-job learning effects, which, if properly taken into account, would not lead agents to postpone their age of entry into the labor force. Thus, for simplicity, I shall consider throughout that agents have non-increasing productivity profiles, which is consistent with (1).[5] Interestingly, with constant productivity, the timing of participation is indeterminate provided that the present value of income[6] remains unchanged.

A recent literature in macro-labor has emphasized the relevance of the above life-cycle model of the labor supply (see Mulligan 2001, Ljungqvist Sargent 2006, 2008, 2010, Prescott Rogerson Wallenius 2009, Rogerson Wallenius 2009). Crucially, a fixed cost of working creates a non-convexity into workers' labor supply problem. According to Hansen (1985) and Rogerson (1988), workers convexify their problem by relying on employment lotteries together with a complete set of financial markets which provides insurance against the outcome of the lotteries. However, in many applications with a participation margin, the non-convexity problem has been ignored on the basis of the fact that most workers do not have access to such lotteries. Recently, Mulligan (2001) and Ljungqvist Sargent (2006) have emphasized that an alternative way for workers to convexify their labor supply problem is to alternate spells of employment and leisure while relying on a risk-free asset to smooth their consumption over time.[7] More precisely, Ljungqvist and Sargent (2006) 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.

According to Ljungqvist and Sargent (2010), these developments have led to the emergence of a new paradigm according to which workers' labor supply should be analyzed within a life-cycle framework where the key object of inquiry is workers' choice of career length. Interestingly, in their analysis of extensive margin elasticities, Chetty, Guren, Manoli and Weber (2011) have shown that a time averaging model *à la* Rogerson Wallenius (2009) generates an empirically plausible Hicksian extensive elasticity of labor supply.

To complete the exposition of the economy, I need to specify its information struc-

---

[5]However, allowing for inverted U-shaped productivity profiles, which induce agents to enjoy some leisure at the beginning of their lives, would not fundamentally change my results. Indeed, if the productivity of all agents is proportional to a baseline productivity profile, i.e. $\gamma_t(\alpha) = \alpha \gamma_t$, then, in the absence of discounting, i.e. $\rho = r = 0$, assuming a non-increasing productivity profile is without loss of generality if the index t ranks time by productivities (in descending order) rather then by ages (in ascending order).

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

[7]Interestingly, even though they were not aware of these controversies, Diamond and Mirrlees (1978) already relied on such a time averaging model of the labor supply to analyze the optimal provision of social insurance against the disability risk.

ture. 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.[8] Full commitment is assumed.

Importantly, this setup could be seen as being embedded in an overlapping generations framework. However, throughout my analysis I exclusively focus on redistribution within, and not across, generations. Hence, as I only focus on a single cohort, I do not need to specify the full overlapping generations structure of the economy; all I need to know is the interest rate at which physical resources are transferred across time.[9] Recall that, for simplicity, I exogenously assume that this interest rate is equal to the discount rate ρ. Thus, the cohort under investigation throughout my analysis could be seen as living on an isolated island which has the possibility to borrow and lend to the rest of the world at rate ρ.

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 determine the best allocation implementable by a direct truthful mechanism whereby each agent is asked to report his type α and where telling the truth is individually rational.

A worker claiming to be of type α receives a consumption stream $\{c_t(\alpha)\}_{t \in [0,H]}$, 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. Note that, if workers truthfully reveal their type α, then these functions jointly characterize the allocation of resources. I shall assume throughout that, for any t, $c_t(\alpha)$, $y_t(\alpha)$ and $R(\alpha)$ are all continuously differentiable in α. The welfare of an α-worker claiming to be of type α' is given by:

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

---

[8] The assumption that labor supply is observable along the extensive margin but not along the intensive margin is problematic if agents have the possibility to alternate spells of employment and leisure at a very high frequency. Thus, following Mulligan (2001), we implicitly assume that there is a maximum frequency at which agents can switch between work and leisure and that "the [resulting] 'indivisibility' is at least as long as the tax accounting period".

[9] Note that, fully specifying the overlapping generations structure of the economy would make it possible to endogenize the interest rate. For instance, this would reveal that, under a fully-funded social security system, the interest rate is equal to the rate of return on physical capital while, under a pay-as-you-go system, the interest rate is determined by the rate of growth of the population and of output.

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 $y_t(\alpha')$ at age t. Let $V(\alpha)$ denote the lifetime utility of an α-worker who is telling the truth, i.e. $V(\alpha) \equiv V(\alpha; \alpha)$. We have:

$$V(\alpha) = \int_0^H e^{-\rho t} u(c_t(\alpha)) dt - \int_0^{R(\alpha)} e^{-\rho t} \left[v\left(l_t(\alpha)\right) + b \right] dt, \qquad (3)$$

where $l_t(\alpha) = y_t(\alpha)/\gamma_t(\alpha)$.

By the revelation principle, any incentive-feasible allocation of resources is implementable by a direct truthful mechanism. It must therefore satisfy the following incentive compatibility constraints:

$$V(\alpha; \alpha) \geq V(\alpha'; \alpha), \text{ for all } \alpha \text{ and } \alpha'. \qquad (4)$$

An incentive-feasible allocation must also satisfy the economy-wide resource constraint:

$$\int_0^{\bar{\alpha}} \left[\int_0^{R(\alpha)} e^{-\rho t} \gamma_t(\alpha) l_t(\alpha) dt - \int_0^H e^{-\rho t} c_t(\alpha) dt \right] f(\alpha) d\alpha \geq E, \qquad (5)$$

where E denotes an exogenous amount of government expenditures that must be financed. The bracketed term on the left-hand-side corresponds to the present value of the lifetime budgetary surplus generated by an α-worker.

Finally, the planner's objective is to maximize social welfare, expressed as a Bergson-Samuelson functional:

$$\int_0^{\bar{\alpha}} \Psi\left(V(\alpha)\right) f(\alpha) d\alpha, \qquad (6)$$

where $\Psi(.)$ is an increasing and weakly concave function weighting the lifetime utility of individuals according to the redistributive objective. Following Tuomala (1990) and Blundell Shephard (2011), a natural specification of $\Psi(.)$, given that $V(.)$ can be negative, is:

$$\Psi(V) = \frac{1 - e^{-\kappa V}}{\kappa}, \qquad (7)$$

where $\kappa \in [0, +\infty)$ is the coefficient of absolute inequality aversion, i.e. $\kappa = -\Psi''(.)/\Psi'(.)$. The two most common benchmarks are the utilitarian social preferences, $\kappa = 0$, 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 exclusively determined by the utility of the worst-off individual.

The planner's problem is to maximize social welfare (6) subject to the resource constraint (5) and to the incentive compatibility constraints (4).

To solve this problem, it is now necessary to express the incentive compatibility con-

straints in a more manageable form. In particular, note that these constraints require that, for any given α, $V(\alpha'; \alpha)$ must be maximized when $\alpha' = \alpha$. Thus, a necessary first-order condition for incentive compatibility is:[10]

$$V_1(\alpha, \alpha) = 0, \text{ for all } \alpha. \tag{8}$$

By definition, $V'(\alpha) = V_1(\alpha; \alpha) + V_2(\alpha; \alpha)$. Thus, this necessary first-order condition for incentive compatibility could be written as:

$$
\begin{aligned}
V'(\alpha) &= V_2(\alpha; \alpha) \\
&= \int_0^{R(\alpha)} e^{-\rho t} l_t(\alpha) v'(l_t(\alpha)) \frac{\gamma_t'(\alpha)}{\gamma_t(\alpha)} dt,
\end{aligned}
\tag{9}
$$

where the second line was obtained by differentiating $V(\alpha; \alpha)$, given by (2), with respect to its second argument.[11]

In order to be able to replace the doubly infinite number of incentive compatibility constraints in (4) by the first-order condition (9), it is essential that this first-order condition does characterize a global maximum.

Lemma 1 *A sufficient condition for the first-order condition (8) or (9) to characterize a global maximum is*

$$\frac{dy_t(\alpha)}{d\alpha} \geq 0 \text{ for all } t \in [0, R(\alpha)) \text{ and } \frac{dR(\alpha)}{d\alpha} \geq 0. \tag{10}$$

Proof. Using the fact that the first-order condition implies that $V_1(\alpha', \alpha') = 0$ for any α', we have:

$$
\begin{aligned}
V_1(\alpha', \alpha) &= V_1(\alpha', \alpha) - V_1(\alpha', \alpha') \\
&= \int_0^{R(\alpha')} e^{-\rho t} \left[\frac{1}{\gamma_t(\alpha')} v' \left(\frac{y_t(\alpha')}{\gamma_t(\alpha')} \right) - \frac{1}{\gamma_t(\alpha)} v' \left(\frac{y_t(\alpha')}{\gamma_t(\alpha)} \right) \right] \frac{dy_t(\alpha')}{d\alpha'} dt \\
&\quad + e^{-\rho R(\alpha')} \left[v \left(\frac{y_{R(\alpha')}(\alpha')}{\gamma_{R(\alpha')}(\alpha')} \right) - v \left(\frac{y_{R(\alpha')}(\alpha')}{\gamma_{R(\alpha')}(\alpha)} \right) \right] \frac{dR(\alpha')}{d\alpha'}.
\end{aligned}
\tag{11}
$$

The disutility $v(.)$ of labor being increasing and convex in the amount of labor supplied, $v(x)$ and $xv'(x)$ are both increasing in x.[12] Also, remember that $\alpha' > \alpha$ implies $\gamma_t(\alpha') \geq \gamma_t(\alpha)$. Hence, the two bracketed terms in (11) have the same sign as $(\alpha - \alpha')$ whenever

---

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

[11]Note that $\gamma_t'(\alpha) = d\gamma_t(\alpha)/d\alpha$.

[12]This implies that the Spence-Mirrlees condition is satisfied. In a static optimal taxation framework, this condition ensures that the first-order condition together with a requirement that output is a non-decreasing function of productivity are necessary and sufficient for incentive compatibility. Here, due

$\gamma_t(\alpha') \neq \gamma_t(\alpha)$ and are otherwise equal to zero. Thus, $dy_t(\alpha)/d\alpha \geq 0$ for all $t \in [0, R(\alpha))$ and $dR(\alpha)/d\alpha \geq 0$ implies that $V_1(\alpha'; \alpha) \geq 0$ for $\alpha' < \alpha$ and $V_1(\alpha'; \alpha) \leq 0$ for $\alpha' > \alpha$, which guarantees that the first order condition (8) does characterize a global maximum. ∎

Note that, to prove Lemma 1, I had to use the assumption that the productivity profiles of different workers never cross. Indeed, when replacing the multiple inequalities in (4) by a first-order condition, I am implicitly using the fact that it is only the downward incentive compatibility constraints which are binding, i.e. workers must be prevented from reporting a slightly lower productivity than they truly have. Fundamentally, this structure is due to redistribution going from high to low productivity agents; but with crossing profiles is not clear who should benefit and who should lose from redistribution and, hence, it is generically not possible to have a first-order approach to the incentive compatibility problem.

Lemma 1 implies that, in the planner's optimization problem, the daunting incentive compatibility constraints (4) could be replaced by the much simpler first-order condition (9) provided that the resulting allocation of resources does satisfy (10). This gives an optimal control problem with $c_t(\alpha)$, $l_t(\alpha)$ and $R(\alpha)$ as control variables and $V(\alpha)$ as the state variable, where we must impose that these variables are related by (3).

Let $\lambda > 0$ denote the multiplier for the resource constraint and $\mu(\alpha)$ the multiplier for the incentive compatibility constraint of the α-worker.

Lemma 2 *The solution to the planner's problem is characterized by the following opti- mality conditions for consumption:*

$$c_t(\alpha) = c(\alpha); \tag{12}$$

for labor supply along the intensive margin:

$$\lambda \left[\gamma_t(\alpha) - \frac{v'(l_t(\alpha))}{u'(c(\alpha))} \right] f(\alpha) + \mu(\alpha) \frac{\gamma'_t(\alpha)}{\gamma_t(\alpha)} \left[v'(l_t(\alpha)) + l_t(\alpha) v''(l_t(\alpha)) \right] = 0; \tag{13}$$

and for labor supply along the extensive margin:

$$\lambda \left[\gamma_{R(\alpha)}(\alpha) l_{R(\alpha)}(\alpha) - \frac{v(l_{R(\alpha)}(\alpha)) + b}{u'(c(\alpha))} \right] f(\alpha) + \mu(\alpha) \frac{\gamma'_{R(\alpha)}(\alpha)}{\gamma_{R(\alpha)}(\alpha)} l_{R(\alpha)}(\alpha) v'(l_{R(\alpha)}(\alpha)) \leq 0, \tag{14}$$

which is binding whenever $R(\alpha) > 0$;[13] *where the multipliers $\mu(\alpha)$ and λ are implicitly*

---

to the multidimensional nature of labor supply, the Spence-Mirrlees condition only makes it possible to find a sufficient condition for the first-order condition to characterize an incentive compatible allocation of resources.

[13]If the constraint $R(\alpha) \leq H$ is binding, then (14) must be satisfied with the reverse inequality, i.e.

determined by:

$$-\mu'(\alpha) = \left[\Psi'(V(\alpha)) - \frac{\lambda}{u'(c(\alpha))}\right] f(\alpha) \ \text{ with } \mu(0) = \mu(\bar{\alpha}) = 0. \tag{15}$$

Proof. Let $\eta(\alpha)$ denote the multiplier for equation (3) which relates the state variable $V(\alpha)$ to the control variables $c_t(\alpha)$, $l_t(\alpha)$ and $R(\alpha)$. The Hamiltonian is:

$$
\begin{aligned}
\mathcal{H} \ = \ & \Psi(V(\alpha))f(\alpha) + \lambda \left[\int_0^{R(\alpha)} e^{-\rho t}\gamma_t(\alpha)l_t(\alpha)dt - \int_0^H e^{-\rho t}c_t(\alpha)dt - E\right] f(\alpha) \\
& + \mu(\alpha)\int_0^{R(\alpha)} e^{-\rho t}l_t(\alpha)v'(l_t(\alpha))\frac{\gamma_t'(\alpha)}{\gamma_t(\alpha)}dt \\
& + \eta(\alpha)\left[\int_0^H e^{-\rho t}u(c_t(\alpha))dt - \int_0^{R(\alpha)} e^{-\rho t}\left[v\left(l_t(\alpha)\right) + b\right]dt - V(\alpha)\right].
\end{aligned} \tag{16}
$$

According to Pontryagin's maximum principle, the control variables must be chosen such as to maximize the Hamiltonian. Importantly, note that none of the control variables can be negative.

The first-order condition for consumption is:[14]

$$\frac{\partial \mathcal{H}}{\partial c_t(\alpha)} = -\lambda e^{-\rho t}f(\alpha) + \eta(\alpha)e^{-\rho t}u'(c_t(\alpha)) = 0. \tag{17}$$

This implies (12) together with:

$$\eta(\alpha) = \frac{\lambda f(\alpha)}{u'(c(\alpha))}. \tag{18}$$

Optimal labor supply along the intensive margin is determined by:

$$\frac{\partial \mathcal{H}}{\partial l_t(\alpha)} = \lambda e^{-\rho t}\gamma_t(\alpha)f(\alpha) \tag{19}$$

$$+\mu(\alpha)e^{-\rho t}\left[v'(l_t(\alpha)) + l_t(\alpha)v''(l_t(\alpha))\right]\frac{\gamma_t'(\alpha)}{\gamma_t(\alpha)} - \eta(\alpha)e^{-\rho t}v'\left(l_t(\alpha)\right) \le 0,$$

which is binding whenever $l_t(\alpha) > 0$. Note that, if the inequality is strict, then we must have $l_t(\alpha) = 0$. But, since $v(0) = v'(0) = 0$, this would imply $\lambda e^{-\rho t}\gamma_t(\alpha)f(\alpha) < 0$, which is not possible. Thus, (19) is always binding. Combining this expression with (18) yields

---

with the left hand side greater or equal to zero. However, assuming that b is sufficiently high or that $\gamma_H(\bar{\alpha})$ is sufficiently low guarantees that even the most productive worker chooses to retire before the end of his life.

[14]I ignore the non-negativity constraint for consumption which cannot be binding since, by assumption, $\lim_{c \to 0+} u'(c) = +\infty$.

(13). The first-order condition for the extensive margin is:

$$\frac{\partial \mathcal{H}}{\partial R(\alpha)} = \lambda e^{-\rho R(\alpha)} \gamma_{R(\alpha)}(\alpha) l_{R(\alpha)}(\alpha) f(\alpha) \tag{20}$$

$$+ \mu(\alpha) e^{-\rho R(\alpha)} l_{R(\alpha)}(\alpha) v'(l_{R(\alpha)}(\alpha)) \frac{\gamma'_{R(\alpha)}(\alpha)}{\gamma_{R(\alpha)}(\alpha)} - \eta(\alpha) e^{-\rho R(\alpha)} \left[v\left(l_{R(\alpha)}(\alpha) \right) + b \right] \leq 0,$$

which is binding whenever $R(\alpha) > 0$. Combining this expression with (18) yields (14). Finally, we must have:

$$\frac{\partial \mathcal{H}}{\partial V(\alpha)} = \Psi'(V(\alpha)) f(\alpha) - \eta(\alpha) = -\mu'(\alpha) \tag{21}$$

together with the transversality conditions:

$$\mu(0) = \mu(\bar{\alpha}) = 0. \tag{22}$$

Substituting (18) into (21) yields (15). ∎

A notable feature of the solution to the planner's problem is that consumption should remain constant throughout the life of an individual. In fact, in the absence of uncertainty and with the interest rate equal to the discount rate, this result is not very surprising since there is nothing to be gained by distorting an individual's lifetime allocation of consumption.

It is now possible to solve for the optimal allocation of resources.

Proposition 1 *The optimal incentive-feasible allocation of resources* $\left\{ R^*(\alpha), \{y_t^*(\alpha)\}_{t \in [0, R^*(\alpha))} , \right.$ $\left. \{c_t^*(\alpha)\}_{t \in [0, H]} \right\}_{\alpha \in [0, \bar{\alpha}]}$ *is characterized by the planner's optimality conditions (12), (13), (14) and (15) together with the constraints of the planner's problem (5), (9) and (3), provided that this allocation[15] satisfies the sufficient condition (10).*

It is important to emphasize that the above resolution of the problem allows for bunching due to binding non-negativity constraints on the control variables, i.e. bunching at the bottom of the skill distribution.[16] However, by considering that the sufficient condition (10) holds, I am ruling out the possibility of bunching due to the failure of the first-order approach to the planner's problem. This type of bunching would be much more problematic here than in a static optimal taxation problem given that the sufficient

---

[15] The planner's optimality conditions are only necessary. Hence, if the system of equations yields more than one solution, then it is the solution that generates the highest level of social welfare which needs to satisfy the sufficient condition (10).

[16] As shown in the proof of Lemma 2, it is only the non-negativity constraint on the retirement age $R(\alpha)$ which could ever be binding.

condition (10) of Lemma 1 is not necessary for the first-order condition (9) to characterize a global maximum. In other words, imposing this sufficient condition together with the first-order condition would be more restrictive than imposing incentive compatibility.

Let $\tau^i(\alpha, t)$ denote the wedge along the intensive margin for an α-worker of age t, which is implicitly defined by:

$$\gamma_t(\alpha)\left(1 - \tau^i(\alpha, t)\right) = \frac{v'(l_t(\alpha))}{u'(c(\alpha))}. \tag{23}$$

Similarly, the extensive wedge $\tau^e(\alpha)$ for an α-worker is defined by:

$$\gamma_{R(\alpha)}(\alpha)l_{R(\alpha)}(\alpha)\left(1 - \tau^e(\alpha)\right) = \frac{v(l_{R(\alpha)}(\alpha)) + b}{u'(c(\alpha))}. \tag{24}$$

These two equations state that, absent any distortions, i.e. $\tau^i(\alpha, t) = 0$ and $\tau^e(\alpha) = 0$, 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 optimality conditions for the intensive (13) and extensive (14) margins, respectively, reveals that:

$$\tau^i(\alpha, t) = -\frac{\mu(\alpha)}{\lambda f(\alpha)}\frac{\gamma_t'(\alpha)}{\left[\gamma_t(\alpha)\right]^2}\left[v'(l_t(\alpha)) + l_t(\alpha)v''(l_t(\alpha))\right], \tag{25}$$

and:

$$\tau^e(\alpha) \leq -\frac{\mu(\alpha)}{\lambda f(\alpha)}\frac{\gamma_{R(\alpha)}'(\alpha)}{\left[\gamma_{R(\alpha)}(\alpha)\right]^2}v'(l_{R(\alpha)}(\alpha)), \tag{26}$$

with an equality whenever $R(\alpha) > 0$. As $\mu(\bar{\alpha}) = 0$, the no distortion at the top principle holds along both margins. The following Lemma is proved in Appendix A.

Lemma 3 *If the sufficient condition (10) of Lemma 1 holds and if there exists a $t \in [0, R(\bar{\alpha}))$ such that $\gamma_t(\alpha) < \gamma_t(\bar{\alpha})$ for all $\alpha < \bar{\alpha}$[17], then $\mu(\alpha) < 0$ for all $\alpha \in (0, \bar{\alpha})$.*

It follows from this Lemma together with (25) and (26) that the wedge is strictly positive along both margins for any $\alpha \in (0, \bar{\alpha})$ provided that the corresponding α-workers do participate.

Before turning to the following section, note that in the absence of an intensive margin, i.e. with $v(.) = 0$, the current model could almost be seen as a life-cycle interpretation of the static Mirrlees (1971) optimal taxation problem where $R(\alpha)$ is the labor supply of the α-worker. There is, however, one crucial difference which is that here the retirement age is

---

[17]This very mild assumption ensures that an α-worker with $\alpha < \bar{\alpha}$ is not *de facto* identical to an $\bar{\alpha}$-worker (as would be the case if they only differed in productivity while not working).

observable, while in Mirrlees (1971) labor supply is not directly observable. This implies that, in the current framework with $v(.) = 0$, the incentive compatibility constraint (9) reduces to $V'(\alpha) = 0$, i.e. all agents end up with identical welfare regardless of the social preferences captured by $\Psi(.)$.[18]

Thus, with an extensive margin only, all agents end up with an identical level of welfare and, hence, high productivity agents do not need to be provided with an informational rent in order to produce a high level of output. This explains why, with $v(.) = 0$, the choice of the retirement age is undistorted, i.e. $\tau^e(\alpha) = 0$ for any value of α. It follows that, in the general model with both margins, the strictly positive extensive wedge is due to the existence of the intensive margin. Indeed, when $V'(\alpha) > 0$, 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 of resources, I turn to the description of how the government could implement this allocation in a decentralized economy by relying on realistic fiscal instruments (rather than on a direct truthful mechanism).[19]

Recall that consumption should optimally be constant throughout the life of an individual. This can be achieved by letting agents trade a risk-free asset over time. This implies that capital taxes are not needed to implement the optimal allocation, which considerably simplifies the problem.

Is it possible to rely exclusively on history-independent income taxes to solve the implementation problem? It turns out that the answer to this question is no.

Lemma 4 *It is, in general, not possible to implement the optimal allocation of resources with a history-independent, but potentially age-dependent, income tax.*

Proof. To prove this statement, it is sufficient to find one example of an allocation that cannot be implemented with a history-independent income tax. Let us assume that

---

[18]Note that, even though the government observes individuals' output and labor supply, i.e. their retirement ages, it does not directly observe their productivity index α. Thus, workers can always claim to have a lower productivity than they truly have by producing inefficiently. Hence, incentive compatibility trivially requires $V'(\alpha) \geq 0$. By contrast, in a (non-Rawlsian) first-best allocation of resources, where the government can observe α, higher productivity workers are forced to retire later, without being compensated by higher consumption, and therefore end up with a lower level of welfare. (With Rawlsian social preferences, the second-best, i.e. unobservable α, and first-best, i.e. full information, allocations are identical.)

[19]Note that, in a static context, once the optimal allocation has been found, it is trivial to determine the optimal non-linear income tax schedule that implements this allocation in a decentralized economy. Indeed, if $y(\alpha)$ and $c(\alpha)$ denote the output and consumption of an α-worker, respectively, then the income tax $T(\alpha)$ paid by this α-worker is implicitly determined by: $T(\alpha) = y(\alpha) - c(\alpha)$. As this section shows, much more work needs to be done in a life-cycle context.

agents only face an extensive margin of labor supply, i.e. $l = 1$ and $v(1) = 0$,[20] and that workers have constant productivity profiles with $\gamma_t(\alpha) = \alpha$. For simplicity, I also impose that there is no time discounting, i.e. $\rho = 0$, and no exogenous amount of government expenditures, i.e. $E = 0$.

By Proposition 1, the optimal allocation of resources is characterized by the optimality condition:

$$\alpha = \frac{b}{u'(c^*(\alpha))}; \tag{27}$$

the resource constraint:

$$\int_0^{\bar{\alpha}} \left[\alpha R^*(\alpha) - c^*(\alpha)H \right] f(\alpha)d\alpha = 0; \tag{28}$$

and the incentive compatibility constraint $V'(\alpha) = 0$.

Let $T(\alpha, t)$ be the income tax paid by a worker who produces output α at age t. Importantly, the resulting income tax schedule is allowed to be age-dependent but not history-dependent. An α-worker therefore faces the following problem:[21]

$$\max_{\left\{ R, \{c_t\}_{t \in [0,H]} \right\}} \int_0^H u(c_t)dt - \int_0^R bdt \tag{29}$$

$$\text{subject to} \quad \int_0^R \left[\alpha - T(\alpha, t) \right] dt \geq \int_0^H c_t dt \tag{30}$$

This α-worker therefore considers that his effective productivity at age t is $\alpha - T(\alpha, t)$. But, note that each worker chooses to supply labor when his productivity is highest. We can therefore impose, without loss of generality, that $T(\alpha, t)$ is an non-decreasing function of t, which guarantees that the worker chooses to work when $t \in [0, R)$ and to enjoy leisure when $t \in [R, H]$.[22] The solution to the α-worker's problem is characterized by $c_t = c$ for all $t \in [0, H]$, by the optimality condition:

$$\alpha - T(\alpha, R) = \frac{b}{u'(c)}, \tag{31}$$

and by his binding budget constraint (30). If $T(\alpha, t)$ is discontinuous at $t = R$, then the

---

[20]This specification could be seen as resulting from a standard constant intertemporal elasticity of substitution utility function with zero elasticity, i.e. $v(l) = \lim_{\delta \to 0+} l^{1+1/\delta}/(1+1/\delta)$, where δ is the constant elasticity parameter.

[21]I am using the fact that, if the income tax schedule $T(.,.)$ implements the optimal allocation, then it must induce α-workers to produce output α throughout their careers. In other words, the marginal tax rate cannot exceed 100% at any age.

[22]Following Rogerson Wallenius (2007), this decreasing productivity profile could be seen as resulting from a change of variable such as to re-order time from the highest productivity instants to the lowest productivity instants.

optimality condition (31) becomes:

$$\lim_{\varepsilon \to 0+} \alpha - T(\alpha, R - \varepsilon) \geq \frac{b}{u'(c)} \geq \lim_{\varepsilon \to 0+} \alpha - T(\alpha, R + \varepsilon). \tag{32}$$

If the income tax schedule $T(.,.)$ implements the optimal allocation, then we must have, by (27) and (31), $T(\alpha, R) = 0$ or, if $T(\alpha, t)$ is discontinuous at $t = R$, by (27) and (32), $\lim_{\varepsilon \to 0+} T(\alpha, R - \varepsilon) \leq 0$. But, as $T(\alpha, t)$ is an non-decreasing function of t, this implies that we must also have $T(\alpha, t) \leq 0$ for all $t \in [0, R)$. In other words, if $T(\alpha, t) = 0$ when workers are indifferent between work and leisure at $t = R$, then we cannot have $T(\alpha, t) > 0$ when they strictly prefer to work at any $t \in [0, R)$.

If $T(.,.)$ implements the optimal allocation, then we can substitute the consumer's budget constraint (30) into the resource constraint (28), which reveals that we must have:

$$\int_0^{\bar{\alpha}} \left[\alpha R^*(\alpha) - c^*(\alpha) H \right] f(\alpha) d\alpha = \int_0^{\bar{\alpha}} \left[\int_0^{R^*(\alpha)} T(\alpha, t) dt \right] f(\alpha) d\alpha = 0. \tag{33}$$

This, together with the requirement that $T(\alpha, t) \leq 0$ for any $t \in [0, R)$, implies that $T(\alpha, t) = 0$ for any values of α and t. But, if the government does not intervene, then it is clear from (29) that higher productivity workers will be better off. This violates the incentive compatibility constraint $V'(\alpha) = 0$. ∎

The intuition for this result is that, when solving for the optimal incentive-feasible allocation of resources, the planner's direct truthful mechanism considers the life-cycle problem as a whole. It can therefore implicitly rely on its memory to reduce the amount of distortions needed to raise a given amount of revenue. By contrast, a history-independent income tax is constrained to create distortions at every single point in time.

To implement the optimal allocation, it is therefore necessary to rely on 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 into account the history of participation and of labor income in order to determine the level of pensions.[23]

Let us now characterize the optimal social security system.[24] To lighten notations, I denote by y^R a given history of participation and (gross) labor income, i.e. $y^R = \left\{ R, \{y_t\}_{t \in [0,R)} \right\}$, and by $y^{R*}(\alpha)$ the optimal incentive-feasible history of the α-worker, i.e. $y^{R*}(\alpha) = \left\{ R^*(\alpha), \{y_t^*(\alpha)\}_{t \in [0, R^*(\alpha))} \right\}$. I define DOM as the set of participation and labor

---

[23]It should nevertheless be emphasized that, while the optimal allocation of resources is typically unique, there is usually several ways to implement this allocation in a decentralized economy (the direct truthful mechanism itself being one way, albeit not very realistic).

[24]The presentation is closely related to that of Grochulski and Kocherlakota (2010).

income histories compatible with a socially optimal allocation. More formally:

$$DOM = \left\{ y^R : y^R = y^{R*}(\alpha) \text{ for some } \alpha \in [0, \bar{\alpha}] \right\}. \tag{34}$$

Finally, I define the function $\hat{c} : DOM \to \mathbb{R}$ such that:

$$\hat{c}(y^{R*}(\alpha)) = c^*(\alpha). \tag{35}$$

This function $\hat{c}(.)$ must exist. Indeed, if it did not exist, then two agents with the same history would end up with different levels of consumption. But, then both agents would claim to be of the type which yields the highest level of consumption, which would violate the incentive compatibility constraint (4).[25]

To make the implementation problem as simple as possible, I now focus on a highly stylized social security system whereby agents get their lifetime income when they retire. They do not receive any income at any other point of their lives. Of course, agents can borrow and lend against this lumpy income such as to smooth their consumption over time. The social security payment received by workers at retirement is set equal to:

$$Q^*(y^R) = \begin{cases} e^{\rho R} \frac{1 - e^{-\rho H}}{\rho} \hat{c}(y^R) & \text{if } y^R \in DOM \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

Proposition 2 *The stylized 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 that generates a history y^R outside DOM cannot be individually rational as the agent would end up with 0 consumption as soon as he deviates from DOM, which would provide him with a lifetime utility of $-\infty$.

Thus, let $y^{R*}(\alpha')$ for some $\alpha' \in [0, \bar{\alpha}]$ be the history of participation and (gross) labor income chosen by an α-worker. By construction, $y^{R*}(\alpha') \in DOM$. The α-worker will determine his consumption level by solving:

$$\max_{\{c_t\}_{t \in [0,H]}} \int_0^H e^{-\rho t} u(c_t) dt - \int_0^{R^*(\alpha')} e^{-\rho t} \left[v\left(\frac{y_t^*(\alpha')}{\gamma_t(\alpha)} \right) + b \right] dt \tag{37}$$

$$\text{subject to } e^{-\rho R^*(\alpha')} Q^*(y^{R*}(\alpha')) \geq \int_0^H e^{-\rho t} c_t dt$$

The agent optimally chooses a constant consumption level $c = c_t$ for all $t \in [0, H]$. The

---
[25]More formally, let $\tilde{\alpha} \neq \hat{\alpha}$ with $R^*(\tilde{\alpha}) = R^*(\hat{\alpha})$ and $y_t^*(\tilde{\alpha}) = y_t^*(\hat{\alpha})$ for all $t \in [0, R^*(\tilde{\alpha}))$ but with $c^*(\tilde{\alpha}) \neq c^*(\hat{\alpha})$. If $c^*(\tilde{\alpha}) > c^*(\hat{\alpha})$, then $V(\tilde{\alpha}; \hat{\alpha}) > V(\hat{\alpha}; \hat{\alpha})$; and, if $c^*(\tilde{\alpha}) < c^*(\hat{\alpha})$, then $V(\hat{\alpha}; \tilde{\alpha}) > V(\tilde{\alpha}; \tilde{\alpha})$.

budget constraint therefore simplifies to:

$$
\begin{aligned}
c &= \frac{\rho e^{-\rho R^*(\alpha')}}{1 - e^{-\rho H}} Q^*(y^{R*}(\alpha')), \\
&= \hat{c}(y^{R*}(\alpha')), \\
&= c^*(\alpha'),
\end{aligned}
\tag{38}
$$

where the second line follows from the definition of the social security system $Q^*(.)$, (36), and the third line from the definition of $\hat{c}(.)$, (35). Thus, if an agent chooses a history $y^{R*}(\alpha')$, then he ends up with a consumption level $c^*(\alpha')$. 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 the history $y^{R*}(\alpha)$ and ends up with consumption level $c^*(\alpha)$. $\blacksquare$

The stylized social security system $Q^*(.)$ relies on the full history of participation and (gross) labor income. It turns out that, in some special cases, it is possible to implement the optimal allocation by relying exclusively on some key summary statistics, which greatly simplifies the proposed policy. For instance, if workers' productivity profile is flat, then it is possible to have a system that only relies on two variables: the present value of lifetime (gross) labor income and the retirement age.[26] Also, if there is no intensive margin of labor supply, then the social security payment $Q^*(.)$ could exclusively depend on the history of participation, i.e. on the retirement age. However, in the general case, these simplifications are not possible since the time profile of (gross) labor income does provide some useful information about the worker's underlying productivity profile and, hence, about his productivity index α.

I 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 history-dependent pension after retirement.

Proposition 3 *For any income tax function $T(.,.)$, potentially age-dependent, the optimal allocation can be implemented by providing retirees with an annuitized pension $P^*(.)$, where:*

$$
P^*(y^R) = \begin{cases} \frac{\rho}{e^{-\rho R} - e^{-\rho H}} \left[\frac{1 - e^{-\rho H}}{\rho} \hat{c}(y^R) - \int_0^R e^{-\rho t} \left[y_t - T(y_t, t) \right] dt \right] & \text{if } y^R \in DOM \\ 0 & \text{otherwise} \end{cases}
\tag{39}
$$

---

[26]With constant productivity we trivially have, from (13), $y_t^*(\alpha) = y^*(\alpha)$ for all $t \in [0, R^*(\alpha))$. Also, workers spontaneously choose a flat labor supply profile in order to reach a desired present value of lifetime labor income. This implies that the two summary statistics pin down the entire history of participation and of (gross) labor income.

Proof. Choosing $y^R \notin DOM$ is still not desirable. For $y^R \in DOM$, the combination of the income tax schedule $T(.,.)$ and of the annuitized pension payments $P^*(.)$ satisfies:

$$\int_0^R e^{-\rho t} \left[y_t - T(y_t, t) \right] dt + \int_R^H e^{-\rho t} P^*(y^R) dt = e^{-\rho R} Q^*(y^R). \tag{40}$$

So, the worker's budget constraint is not affected by the switch 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 follows that the optimal policy is compatible with any specific recommendations about the shape of the income tax schedule since the effects of any given tax schedule can be undone after retirement by adjusting the pension payments.

Although it is commonly argued that redistribution should be one of the main objectives of a well designed pension system (cf., for instance, 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. Lemma 4 together with Proposition 2 and 3 contribute to this debate by showing that, indeed, a standard non-linear income tax needs to be supplemented with an optimally designed social security system which must be sensitive to equity concerns.

I have so far assumed that agents can trade a risk-free asset at the exogenous interest rate ρ. This implies that, if necessary, they can use their future social security payments as a collateral in order to be able to borrow enough to smooth their consumption perfectly over time. If, on the contrary, agents do not have perfect access to the credit market, then some restrictions on the shape of the income tax $T(.,.)$ must be imposed in order to implement the optimal allocation with the above social security system.

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 implicitly determined by:*

$$T^*(y_t^*(\alpha), t) = y_t^*(\alpha) - c^*(\alpha). \tag{41}$$

Proof. By construction, $\{T^*(.,.), P^*(.)\}$ is the only optimal policy such that an α-worker receives net income $c^*(\alpha)$ at any point in time. This implies that, even if it was possible, agents would never want to trade any assets. Note that the optimal income tax function $T^*(.,.)$ is well defined provided that $\frac{dy_t^*(\alpha)}{d\alpha} > 0$, which makes it possible to identify $c^*(\alpha)$ from $y_t^*(\alpha)$.[27] ∎

---

[27] If this condition does not hold and individuals can save but not borrow, then the highest value of

Thus, even if agents cannot borrow and lend, the government allows them to convexify their labor supply problem by implementing a policy which still induces them to work for only a fraction $R^*(.)$ of their lives.

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 supply an optimal amount of labor throughout their career, i.e. if their y^R fails to belong to DOM. Clearly, to address this issue, the present framework would need to be augmented with features that could explain such outcomes. It could nevertheless be conjectured that, whether workers fail to choose $y^R \in DOM$ because of skill risks (such as the occurrence of disability shocks) or because of limited cognitive capacities, the unlikely scenarios and their corresponding histories should be penalized. Indeed, this would improve incentives to work at little cost in terms of welfare (since the corresponding scenarios are unlikely to occur). 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. It should be emphasized that this section relies on the characterization of the optimal allocation summarized by Proposition 1, not on its implementation in a decentralized economy. In this section, I first rely on a simple no-redistribution benchmark to calibrate the parameters of the model. I then simulate the optimal redistribution policy. Finally, I decompose the social welfare gains generated by a switch from the no-redistribution benchmark to the optimal policy.

5.1   A No-Redistribution Benchmark

In this subsection, I focus on the life-cycle problem of a single α-worker in the absence of redistribution. To finance its expenditures, the government imposes on all workers a proportional labor income tax of rate π. Let $\left\{ \tilde{R}(\alpha), \left\{ \tilde{l}_t(\alpha) \right\}_{t \in [0, \tilde{R}(\alpha))}, \left\{ \tilde{c}_t(\alpha) \right\}_{t \in [0, H]} \right\}$ denote the allocation chosen by an α-worker in the no-redistribution benchmark. This

---

$c^*(\alpha)$ compatible with $y_t^*(\alpha)$ should be used to compute $T^*(.,.)$ from (41). If agents can neither save nor borrow, then the government could ask them to choose a value of $c^*(\alpha)$ compatible with $y_t^*(\alpha)$ and punish them with zero consumption after retirement if $c^*(\alpha)$ is not compatible with their entire history of participation and labor income.

allocation solves the following problem:

$$\max_{\left\{\tilde{R}(\alpha),\left\{\tilde{l}_t(\alpha)\right\}_{t\in[0,R(\alpha))},\left\{\tilde{c}_t(\alpha)\right\}_{t\in[0,H]}\right\}} \int_0^H e^{-\rho t}u(\tilde{c}_t(\alpha))dt - \int_0^{\tilde{R}(\alpha)} e^{-\rho t}\left[v(\tilde{l}_t(\alpha)) + b\right]dt \quad (42)$$

$$\text{subject to} \quad (1-\pi)\int_0^{\tilde{R}(\alpha)} e^{-\rho t}\gamma_t(\alpha)\tilde{l}_t(\alpha)dt \geq \int_0^H e^{-\rho t}\tilde{c}_t(\alpha)dt \quad (43)$$

It is straightforward to check that the first-order conditions to the problem are $\tilde{c}_t(\alpha) = \tilde{c}(\alpha)$ for all $t \in [0, H]$ together with:

$$\gamma_t(\alpha) = \frac{v'(\tilde{l}_t(\alpha))}{u'(\tilde{c}(\alpha))} \quad \text{and} \quad \gamma_{\tilde{R}(\alpha)}(\alpha)\tilde{l}_{\tilde{R}(\alpha)}(\alpha) = \frac{v(\tilde{l}_{\tilde{R}(\alpha)}(\alpha)) + b}{u'(\tilde{c}(\alpha))}. \quad (44)$$

The tax rate π is set such as to satisfy the government budget constraint:

$$\pi \int_0^{\bar{\alpha}} \left[\int_0^{\tilde{R}(\alpha)} e^{-\rho t}\gamma_t(\alpha)\tilde{l}_t(\alpha)dt\right] f(\alpha)d\alpha = E. \quad (45)$$

Individuals can work from age 23 until they die on their 80th birthday, i.e. $H = 80 - 23$ (since workers are 23 years old when $t = 0$). 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:

$$\gamma_t(\alpha) = \alpha\gamma_t. \quad (46)$$

Figure 1 displays the baseline profile γ_t which is such that productivity is constant and normalized to 1 until age 60 and then declines linearly until it reaches 0 at 80. This specification together with the first-order condition for $\tilde{l}_t$, (44), is consistent with the fact that, empirically, the number of hours worked by participating workers is nearly constant until age 60 (cf. Blundell Bozio Laroque 2011, Figure 36B and 37B).

The annual discount rate is set equal to 2%, i.e. $\rho = 0.02$. The instantaneous utility derived from consumption is logarithmic:

$$u(c_t) = \log(c_t). \quad (47)$$

Importantly, given that preferences are additively separable between consumption and leisure, this logarithmic specification is necessary and sufficient to have the number of hours worked and the retirement age independent of the productivity index α. Indeed, substituting the specification of productivity (46) and of utility (47) into the first-order conditions (44) and into the budget constraint (43) yields $\tilde{l}_t(\alpha) = \tilde{l}_t$, $\tilde{R}(\alpha) = \tilde{R}$ and $\tilde{c}(\alpha) = \alpha\tilde{c}(1)$. By the same token, the logarithmic specification implies that labor supply would remain constant in a growing economy, which is empirically reasonable.
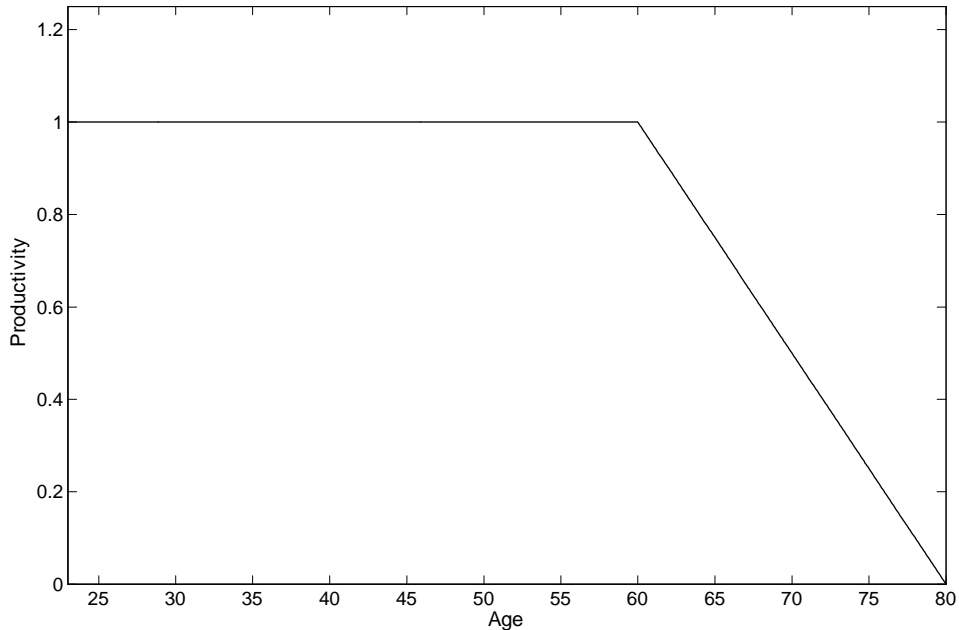
Figure 1: Baseline productivity profile

The disutility from supplying labor along the intensive margin is given by a standard power function:

$$v(l_t) = k \frac{l_t^{1+\frac{1}{\delta}}}{1 + \frac{1}{\delta}}, \tag{48}$$

where δ is the constant Frisch intensive elasticity of labor supply. Following Chetty Guren Manoli Weber (2011), I take $\delta = 0.5$. The fixed cost b of working and the scale parameter k in (48) are set such that workers choose to retire at age 65, i.e. $\tilde{R} = 65 - 23$, and such that the labor supply of prime age workers is normalized to one, i.e. $\tilde{l}_t = 1$ for $t \in [23 - 23, 60 - 23]$.[28] This yields $b = 0.526$ and $k = 1.215$.

Interestingly, it can be shown that, with this calibration of the model, the Frisch extensive elasticity of labor supply is equal to 0.36 while the Hicksian intensive and extensive elasticities are equal to 0.30 and 0.22, respectively. All these values are very close to the preferred empirical estimates that Chetty, Guren, Manoli and Weber (2011, Table 2) derive from their meta-analysis.[29] Of course, with logarithmic utility from consumption, the Marshallian intensive and extensive elasticities are both equal to zero.

The level of government expenditures E is calibrated to amount to a quarter of total

---

[28]It can be shown that, in the current context, the effect of k on the variables other than $\tilde{l}_t$ is identical to that of a proportional shift, i.e. a rescaling, of the productivity index α.

[29]Their preferred estimates are 0.54 and 0.29 for the Frisch intensive and extensive elasticities and 0.33 and 0.24 for the Hicksian intensive and extensive elasticities, respectively.
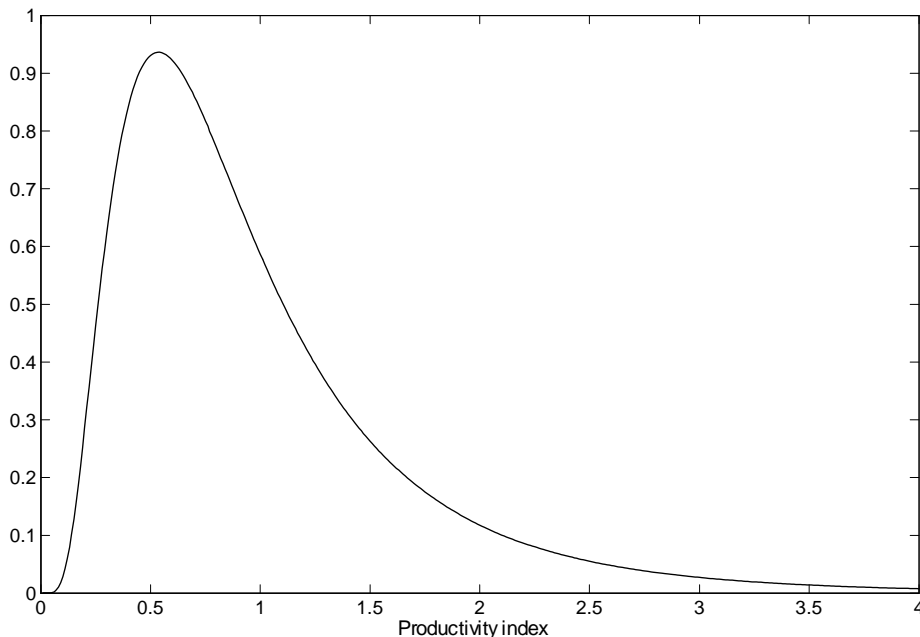
Figure 2: Lognormal distribution of the productivity index α

output. This trivially implies, from the government budget constraint (45), that $\pi = 0.25$. Normalizing the average productivity index α to 1, i.e. $E[\alpha] = 1$, the calibrated model gives $E = 6.998$ and $E[\tilde{c}(\alpha)] = E[\alpha\tilde{c}(1)] = \tilde{c}(1) = 0.617$. By construction, the average present value of lifetime consumption, $\int_0^H e^{-\rho t}\tilde{c}(1)dt$, is equal to $[(1-\pi)/\pi]E = 3E = 20.995$.

Note that, for workers less than 60 years old, the gross wage and the productivity index α are identically distributed since $\gamma_t(\alpha) = \alpha\gamma_t = \alpha$ when $t \leq 60 - 23$. I assume that both follow a lognormal distribution. While this distribution does not exactly match the wage distribution, especially for the tails, it seems to be a decent approximation. Relying on CPS data, Eckstein Nagypal (2004, Table 3) and Heathcote Perri Violante (2010, Table 4) both find a standard deviation of log wages in the U.S. in the early 2000s equal to about 0.65. This implies a coefficient of variation of lognormally distributed wages equal to $\sqrt{e^{0.65^2} - 1} = 0.725$. Hence, to simulate the optimal policy, I assume that the productivity index α is lognormally distributed with a mean of 1 and a standard deviation of 0.725. The density of this distribution is plotted in Figure 2.[30]

Now that the theoretical model is fully calibrated, I turn to the simulation of the

---

[30] Of course, to the extent that a high wages could reflect high efforts along the intensive margin, it would be preferable to rely on the structure of the model and on the current fiscal system to infer the skill distribution from the earnings distribution, as in Saez (2001). However, doing so in a life-cycle context is not straightforward and would probably not be very sensible given the stylized nature of my model.
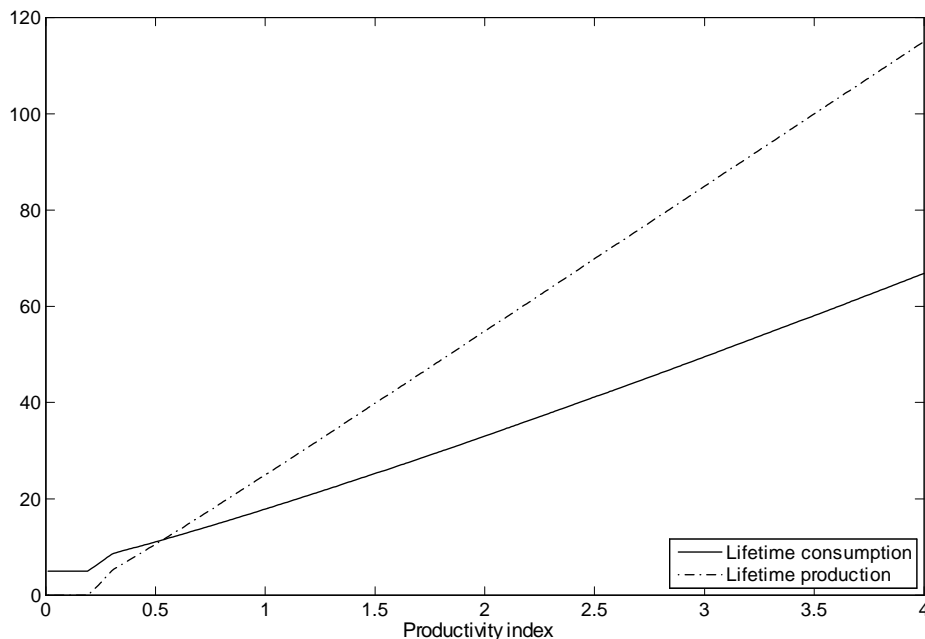
Figure 3: Lifetime production and consumption as a function of the productivity index α

optimal redistribution policy.

5.2 Optimal Redistribution Policy

I now derive the optimal allocation of resources chosen by a utilitarian planner who wants to maximize the sum of individual lifetime utilities; thus $\kappa = 0$ in (7). To perform the simulation, I have relied on a discretized version of the first-order conditions of Lemma 2 which I have obtained by solving the planner's problem under a discrete ability distribution.

Under the optimal redistribution policy, total output is 9.5% lower than under the no-redistribution benchmark. This implies that public expenditures E now absorb to 27.6% of total output. Figure 3 displays the lifetime production and consumption of workers as a function of their productivity index. The least productive 1.3% of the population, those with $\alpha < 0.19$, never participate to the labor market. Thanks to redistribution, they can nevertheless sustain a consumption level equal to 27% of the average consumption level in the economy. Lifetime consumption exceeds production for 26.0% of workers; those whose productivity index α falls below 0.53. On the other side of the distribution, the most productive agents consume at least 58% of their output.

Figure 4 shows the budget surplus raised from each type of worker, i.e. the difference between the lifetime production and consumption of an α-worker multiplied by the
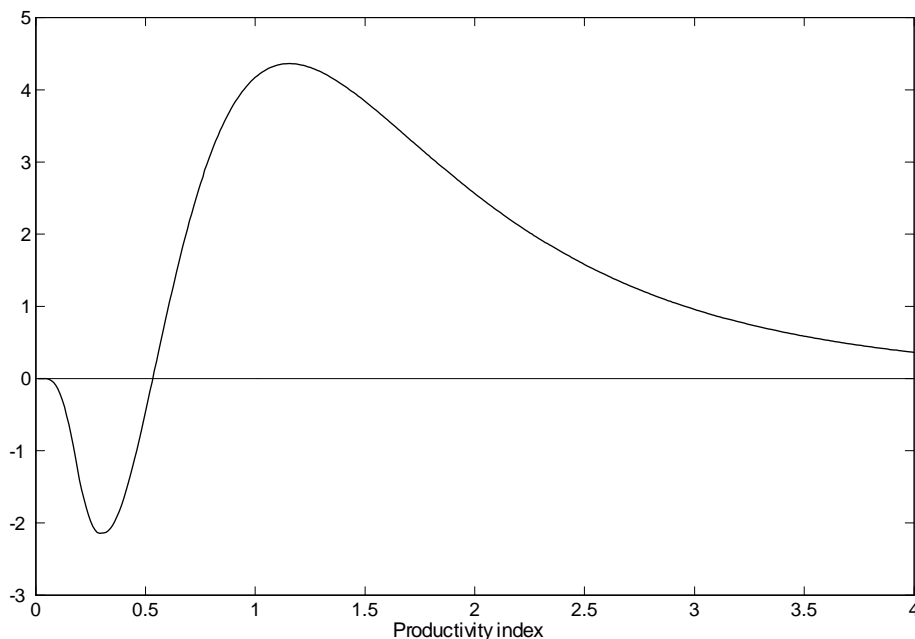
24

Figure 4: Budget surplus raised from each type of worker

number $f(\alpha)$ of such workers. This figure illustrates the well-known fact that the bulk of the financing of public expenditures and redistribution falls on the upper middle class. While the rich are, individually, among the largest contributors, they are not sufficiently numerous to be the main source of government revenue. Note that the excess of surpluses over deficits seen on Figure 4 is necessary to finance the government expenditures E.

Figure 5 displays the retirement age of workers as a function of their productivity index. Career length is increasing in productivity.[31] Recall that, in the no-redistribution benchmark, all workers choose to retire at age 65. It follows that the substantial variation in retirement ages across workers could be fully attributed to the planner's intervention. Figure 6 shows the distribution of retirement ages across the population. The average retirement age is 65.05, while its standard deviation across the population is 7.47 years. There is a mass 1.3% of workers whose productivity is so low that they never participate. Note that, with $v(0) = v'(0) = 0$, bunching at the bottom of the distribution would not occur in the absence of a fixed cost $b > 0$ of working. 5.2% of workers do participate but retire before age 60.[32] 74.5% of workers retire between 65 and 69.3, the highest retirement

---

[31]Under the current calibration, it turns out that, for very high levels of productivity, the retirement age is very slightly declining in the productivity index. However, this only affects a small mass of workers. Also, it could easily be checked numerically that all the incentive compatibility constraints in (4) are satisfied, implying that the first-order approach used to solve the planner's problem is indeed valid.

[32]These workers have constant productivity throughout their career. Hence, instead of working continuously until retirement, they could choose to alternate spells of employment and leisure provided that the present value of their production remains unchanged. To implement these alternative allocations,
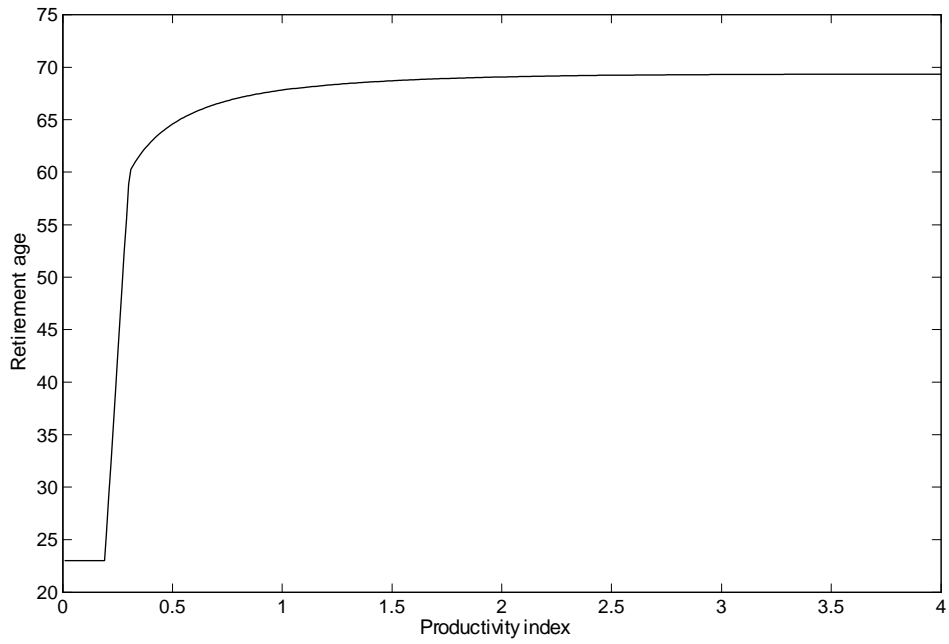
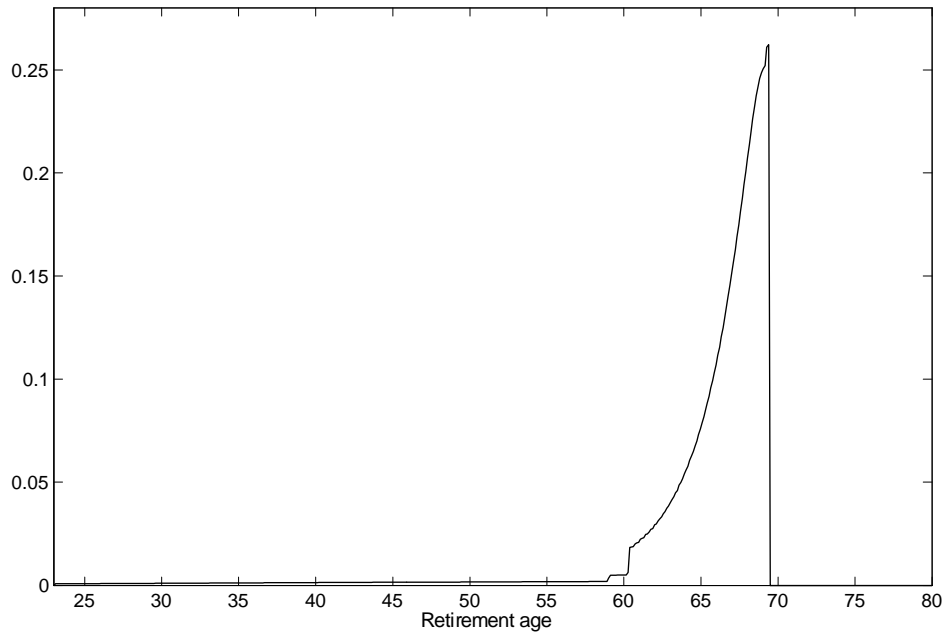Figure 5: Retirement age as a function of the productivity index α

Figure 6: Distribution of retirement ages (at age 23, a mass of 1.3% of non-participating workers is omitted from the graph)

age. This substantial heterogeneity in career length shows that the retirement age is a key margin that the government should exploit as part of an optimal redistribution policy.

There is also some variation in labor supply along the intensive margin since higher productivity agents work longer hours while participating. The least productive workers who participate, for whom $\alpha = 0.19$, supply 0.61 units of labor at each instant of their (infinitesimally short) career. The median worker, characterized by $\alpha = 0.81$, supplies 0.84 units of labor before his productivity starts declining at age 60. Only very high productivity workers, with $\alpha > 4.5$, supply more labor along the intensive margin than in the no-redistribution benchmark (where labor supply before 60 was normalized to 1). For those who retire after 60, labor supply falls as productivity declines at the end of their career.[33]

How do the wedges along the intensive and extensive margins compare? It turns out that, with a constant Frisch intensive elasticity, as implied by (48), the relationship between the intensive, $\tau^i(\alpha, t)$, and the extensive, $\tau^e(\alpha)$, wedge satisfies:[34]

$$\frac{\tau^i(\alpha, t)}{\tau^e(\alpha)} = 1 + \frac{1}{\delta}. \tag{49}$$

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. This is reminiscent of Ramsey's (1927) inverse elasticity rule. 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.[35] More generally, the intensive wedge is always at least as large as the extensive wedge. This is due to the convexity of the intensive disutility cost $v(.)$ of working, cf. (25) and (26), which raises the temptation for workers to underreport their true productivity such as to have to produce a smaller output at any age. Thus, the incentive compatibility constraint distorts the intensive margin more than the extensive margin. Figure 7 reports the wedges of participating workers. With $\delta = 0.5$, the intensive wedge is three times larger than the extensive wedge.

---

the social security system (36) would have to be changed slightly by setting, for instance, age 60 as the legal retirement age before which no history-dependent transfers could be made.

[33]It can be shown from (13) that, for any given value of α, the ratio $[l_t(\alpha)]^{1/\delta} / \gamma_t$ remains constant for all values of $t \in [0, R(\alpha))$.

[34]This immediately follows from the definition of the wedges, (25) and (26), and from the previous footnote which implies that the intensive wedge is independent of age.

[35]It can be shown from (13) that $v'(l_t(\alpha))$ tends to 0 as δ tends to 0 and, hence, from (23), $\tau^i(\alpha, t)$ tends to 1 as δ tends to 0.
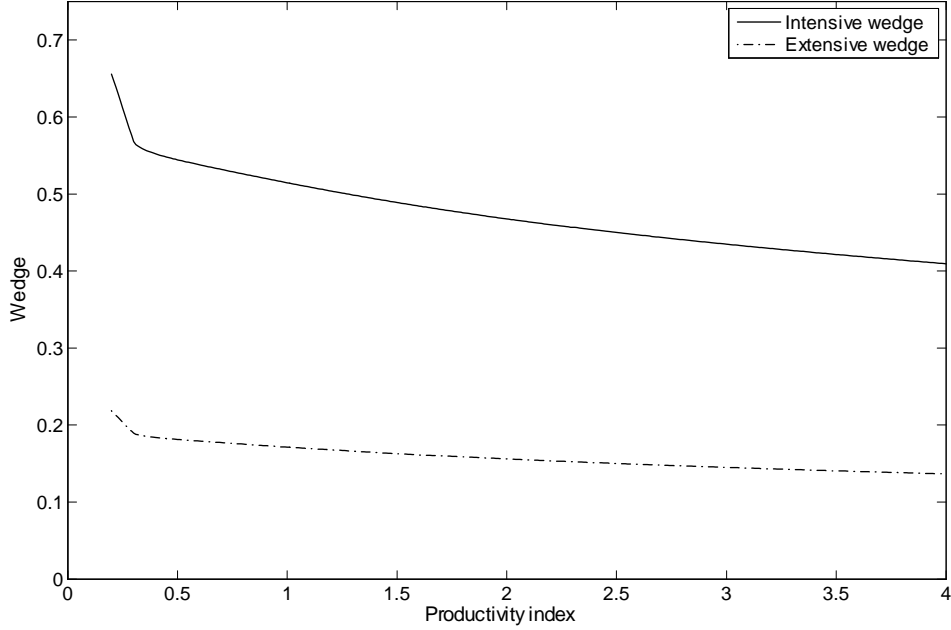
Figure 7: Intensive and extensive wedges as a function of the productivity index α

5.3   Decomposition of the Social Welfare Gains

By construction, the optimal redistribution policy improves social welfare compared to the no-redistribution benchmark. Although this redistribution policy decreases GDP by 9.5%, it generates a consumption equivalent social welfare gain of 7.7%, i.e. the level of social welfare is identical in the optimal redistribution policy as in the no-redistribution benchmark with the consumption of all agents increased by 7.7%.

The welfare of an individual depends on both his consumption level and his labor supply. Hence, it would be interesting to decompose the social welfare gain from the optimal policy into four components: a consumption component and three labor supply components corresponding, respectively, to the intensive margin, the extensive margin and an interaction between the two. This is what I do in this subsection.

The welfare of an α-worker under the optimal policy is given by:

$$V(\alpha) = \int_0^H e^{-\rho t} u(c(\alpha)) dt - \int_0^{R(\alpha)} e^{-\rho t} \left[v\left(l_t(\alpha)\right) + b \right] dt; \tag{50}$$

while, under the no-redistribution benchmark, it is equal to:

$$\tilde{V}(\alpha) = \int_0^H e^{-\rho t} u(\alpha \tilde{c}(1)) dt - \int_0^{\tilde{R}} e^{-\rho t} \left[v\left(\tilde{l}_t\right) + b \right] dt. \tag{51}$$

28

Hence, the welfare gain of an α-worker can be written as:

$$V(\alpha) - \tilde{V}(\alpha) = \int_0^H e^{-\rho t} \left[u(c(\alpha)) - u(\alpha \tilde{c}(1)) \right] dt \qquad (52)$$

$$- \int_0^{\tilde{R}} e^{-\rho t} \left[v(l_t(\alpha)) - v\left(\tilde{l}_t\right) \right] dt$$

$$- \int_{\tilde{R}}^{R(\alpha)} e^{-\rho t} \left[v\left(\tilde{l}_t\right) + b \right] dt$$

$$- \int_{\tilde{R}}^{R(\alpha)} e^{-\rho t} \left[v(l_t(\alpha)) - v\left(\tilde{l}_t\right) \right] dt,$$

which decomposes the welfare gain into a consumption component, an intensive component, an extensive component and an interaction component, respectively.[36] To obtain the social welfare gain from the optimal policy, it is necessary to aggregate these components over the whole population, i.e. over the whole distribution of α.

Figure 8 displays the welfare gain generated by each component and for each type α of worker. Thus, each curve corresponds to one of the four components of (52) multiplied by the mass $f(\alpha)$ of α-workers to whom it applies. It follows that the social welfare gain from the optimal policy is equal to the sum of the areas[37] under these four curves.

Figure 8 shows that most low productivity workers gain more from redistribution through an earlier retirement age than through a higher consumption level. For high productivity workers, lower consumption is a significant source of welfare loss. This is not surprising since GDP is lower and, moreover, the distribution of consumption across the population is more equal. Almost all workers benefit from a reduction of their labor supply along the intensive margin. Finally, to understand why the interaction component is negative for low productivity workers, note that they retire earlier thanks to the redistribution policy but, if they did participate until $\tilde{R}$, they would have supplied less labor during those years than under the no-redistribution benchmark. In other words, the interaction component corrects for the fact that the intensive component reports a large gain for these workers even though some them never participate.

Aggregating these four components reveals that 50.9% of the population, those with $\alpha < 0.82$, gain from redistribution; while the others prefer the no-redistribution benchmark. The social welfare gain from redistribution can be broken down into the four components by integrating (52) over the whole population. Equivalently, the contribution of each component is proportional to the area under the corresponding curve of Figure 8.

---

[36]Note that, to compute some of these components, it is necessary to rely on counterfactual values of labor supply along the intensive margin, i.e. values of $l_t(\alpha)$ for $t > R(\alpha)$ and $\tilde{l}_t$ for $t > \tilde{R}$, which can nevertheless be computed from the first-order conditions (13) and (44). Alternatively, to avoid relying on these values, it is possible to merge the interaction term with the intensive component when $R(\alpha) < \tilde{R}$ and with the extensive component when $R(\alpha) > \tilde{R}$.

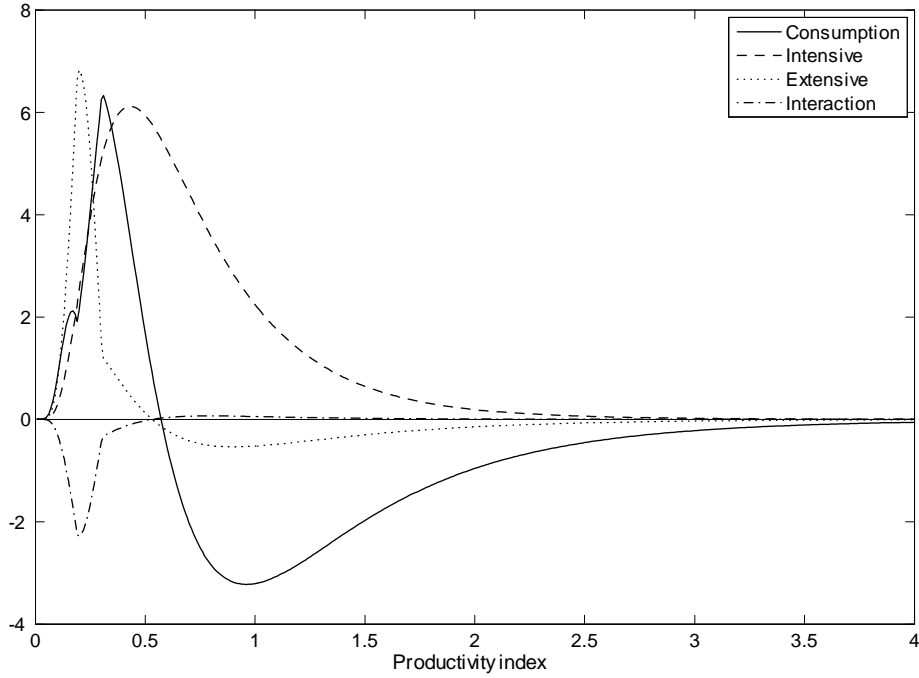[37]The area is negative when the curve is below the horizontal axis.

Figure 8: Decomposition of the Social Welfare Gains from the Optimal Policy

Table 1 reports the consumption equivalent social welfare gain from the optimal policy as well as its decomposition into the four components.

Table1: Decomposition of the social welfare gain from the optimal policy

	GDP	Welfare	Breakdown			
	loss	gain	Consumption	Intensive	Extensive	Interaction
$\delta = 0.5$	9.5%	7.7%	-86.7%	182.3%	15.1%	-10.7%
$\delta = 0.25$	8.2%	11.2%	-8.0%	95.1%	19.7%	-6.8
$\delta = 0$	1.1%	29.2%	64.1%	0%	35.9%	0%

Interestingly, when $\delta = 0.5$, the social welfare gain generated by the optimal policy is not due to a preferred allocation of consumption, but to a welfare enhancing reduction of labor supply, especially along the intensive margin. Not only does the consumption allocation not increase social welfare, in fact it lowers welfare significantly: the more equal distribution of consumption across the population does not compensate for the lower level of output.

Decreasing the intensive elasticity δ of labor supply relaxes the incentive compatibility constraint, which leads to a smaller drop of output and a larger social welfare gain

generated by the optimal policy. Recall that, in the extreme case where $\delta = 0$, the only dimension of labor supply, i.e. the retirement age, is observable by the government, which simplifies the incentive compatibility constraint to $V'(\alpha) = 0$. The smaller drop in output leads to a larger consumption component. Also, when δ is low, it is easier to induce high skilled agents to produce a high level of output,[38] which allows the government to further reduce the retirement age of low skilled workers such as to raise their welfare. This increases the extensive component while maintaining GDP nearly constant.

In the absence of public expenditures, i.e. $E = 0$, and with $\delta = 0.5$, the implementation of the optimal policy induces a much larger drop in output, equal to 19.9% of GDP, and it generates a social welfare gain of 14.0% of consumption. The magnitude of the extensive component increases to 97.1% while that of the intensive component reduces to 135.8%. The consumption component is equal to -100.9% and the interaction component to -32.1%. Thus, when there is no need to finance any government expenditures, the planner relies much more extensively on the retirement age in order to redistribute welfare to the low skilled. Indeed, in the corresponding optimal allocation, 26.5% of the workforce retires before age 60, which includes 8.1% of the population who never participate.

6   Conclusion

In this paper, I have characterized the optimal redistribution policy in a life-cycle framework with both an intensive and an extensive margin of 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, a substantial amount of redistribution should be done within the social security system.

In many industrialized countries, a looming pension crisis makes it necessary to increase the average retirement age. This creates a unique opportunity to reform social security systems and my 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.

For simplicity, I have assumed that the fixed cost of working remains constant throughout the life of an individual. However, my analysis implies that if a worker, such as a mother of young children, faces a high fixed cost of working over a few years, then she should take some time off during those years.[39] This policy recommendation is very different from that implied by a corresponding static analysis of optimal redistribution with an extensive margin which would advocate for the implementation of a tax credit to

---

[38]When $\delta = 0.5$, switching from the no-redistribution benchmark to the optimal policy increases the lifetime production of only 3.9% of workers, those with $\alpha > 2.55$; while, when $\delta = 0$, the optimal policy increases the lifetime production of 51.5% of workers, those with $\alpha > 0.79$.

[39]Of course, my analysis abstracts from human capital considerations.

induce that person to work. Hence, further research on the precise nature of the extensive margin of labor supply could have dramatic consequences for policy recommendations.

Another promising avenue for further research is the introduction of skill risks within the framework of this paper. In particular, some high productivity workers might become unable to have long careers. Thus, allowing for the random occurrence of permanent disability shocks, as in Diamond Mirrlees (1978), 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] Barr, N. and Diamond, P.A. (2008), *Reforming Pensions: Principles and Policy Choices*, Oxford and New York: Oxford University Press.

[3] Beaudry, P., Blackorby, C. and Szalay, D. (2009), 'Taxes and Employment Subsidies in Optimal Redistribution Programs', *American Economic Review*, 99(1), 216-242.

[4] Brewer, M., Saez, E. and Shephard, A. (2010), 'Means-Testing and Tax Rates on Earnings', in *The Mirrlees Review: Dimensions of Tax Design*, edited by the Institute for Fiscal Studies, Oxford and New York: Oxford University Press.

[5] Blundell, R., Bozio, A. and Laroque, G. (2011), 'Extensive and Intensive Margins of Labour Supply: Working Hours in the US, UK and France', IFS Working Paper 01/11.

[6] Blundell, R. and Shephard, A. (2011), 'Employment, Hours of Work and the Optimal Taxation of Low Income Families', Working Paper, UCL and Princeton.

[7] Chetty, R., Guren, A., Manoli, D. and Weber, A. (2011), 'Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities', Working Paper, Harvard University, UCLA and University of Mannheim.

[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. (2011), 'Optimal Taxation in the Extensive Model', *Journal of Economic Theory*, 146, 425-453.

[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. (2010), '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] Eckstein, Z. and Nagypal, E. (2004), 'The Evolution of U.S. Earnings Inequality: 1961-2002', *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(2), 10-29.

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

[17] Farhi, E. and Werning, I. (2010), 'Insurance and Taxation over the Life Cycle', Working Paper, Harvard and MIT.

[18] Golosov, M., Troshkin, M. and Tsyvinski, A. (2010), 'Optimal Dynamic Taxes', Working Paper, Yale and University of Minnesota.

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

[20] Gorry, A. and Oberfield, E. (2010), 'Optimal Taxation over the Life Cycle', Working Paper, UC Santa Cruz and Federal Reserve Bank of Chicago.

[21] Grochulski, B. and Kocherlakota, N. (2010), 'Nonseparable Preferences and Optimum Social Security Systems', *Journal of Economic Theory*, 145, 2055-2077.

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

[23] Heathcote, J., Perri, F. and Violante, G.L. (2010), 'Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006', *Review of Economic Dynamics*, 13, 15-51.

[24] 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.

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

[26] Kocherlakota, N.R. (2010), *The New Dynamic Public Finance*, Princeton, NJ: Princeton University Press.

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

[28] Laroque, G. (2009), 'On Income and Capital Taxation in a Life Cycle Model with Extensive Labor Supply', Working Paper, INSEE-CREST.

[29] Liebman, J.B. (2002), 'Should Taxes Be Based on Lifetime Income? Vickrey Taxation Revisited', Working Paper, Harvard University.

[30] 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.

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

[32] Ljungqvist, L. and Sargent, T. (2010), 'Career Length: Effects of Curvature of Earnings Profiles, Earning Shocks, and Social Security', Working Paper, Stockholm School of Economics and New York University.

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

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

[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. (2007), 'Micro and Macro Elasticities in a Life Cycle Model with Taxes', NBER Working Paper No. 13017.

[39] Rogerson, R. and Wallenius, J. (2009), 'Micro and Macro Elasticities in a Life Cycle Model with Taxes', *Journal of Economic Theory*, 144, 2277-2292.

[40] Saez, E. (2001), 'Using Elasticities to Derive Optimal Income Tax Rates', *Review of Economic Studies*, 68(1), 205-229.

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

[42] 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.

[43] Tuomala, M. (1990), *Optimal Income Tax and Redistribution*, Oxford and New York: Oxford University Press.

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

[45] Weinzierl, M. (2011), 'The Surprising Power of Age-Dependent Taxes', *Review of Economic Studies*, Forthcoming.

A   Proof of Lemma 3

Integrating the optimality conditions for the multipliers, given by (15), yields:

$$\mu(\alpha) = \int_\alpha^{\bar{\alpha}} \left[\Psi'(V(x)) - \frac{\lambda}{u'(c(x))} \right] f(x)dx. \tag{A1}$$

Let us define:

$$D(\alpha) = \frac{1}{1 - F(\alpha)} \int_\alpha^{\bar{\alpha}} \Psi'(V(x))f(x)dx, \tag{A2}$$

where $F(\alpha) = \int_0^\alpha f(x)dx$; and:

$$E(\alpha) = \frac{1}{1 - F(\alpha)} \int_\alpha^{\bar{\alpha}} \frac{f(x)}{u'(c(x))}dx. \tag{A3}$$

Thus, (A1) could be written as:

$$\mu(\alpha) = [1 - F(\alpha)] [D(\alpha) - \lambda E(\alpha)]. \tag{A4}$$

We know, by (15), that $\mu(0) = 0$. Hence:

$$\lambda = \frac{D(0)}{E(0)}. \tag{A5}$$

Substituting this value into (A4) yields:

$$\mu(\alpha) = [1 - F(\alpha)] E(\alpha) \left[\frac{D(\alpha)}{E(\alpha)} - \frac{D(0)}{E(0)}\right]. \tag{A6}$$

Let us now show that $D(\alpha)$ is non-increasing in α, while $E(\alpha)$ is non-decreasing in α. Differentiating (A2) gives:

$$
\begin{aligned}
D'(\alpha) &= \frac{-\Psi'(V(\alpha))f(\alpha)[1 - F(\alpha)] + f(\alpha)\int_\alpha^{\bar{\alpha}} \Psi'(V(x))f(x)dx}{[1 - F(\alpha)]^2}, \\
&= \frac{f(\alpha)}{[1 - F(\alpha)]^2} \int_\alpha^{\bar{\alpha}} [\Psi'(V(x)) - \Psi'(V(\alpha))] f(x)dx \leq 0.
\end{aligned}
\tag{A7}
$$

The main bracket of the integral cannot be positive since $\Psi''(.) \leq 0$ and, by the incentive compatibility constraint (9), $V(x) \geq V(\alpha)$ if $x > \alpha$. Differentiating (A3) yields:

$$
\begin{aligned}
E'(\alpha) &= \frac{1}{[1 - F(\alpha)]^2} \left[\frac{-f(\alpha)}{u'(c(\alpha))}[1 - F(\alpha)] + f(\alpha)\int_\alpha^{\bar{\alpha}} \frac{f(x)}{u'(c(x))}dx\right], \\
&= \frac{f(\alpha)}{[1 - F(\alpha)]^2} \int_\alpha^{\bar{\alpha}} \left[\frac{1}{u'(c(x))} - \frac{1}{u'(c(\alpha))}\right] f(x)dx \geq 0.
\end{aligned}
\tag{A8}
$$

This derivative is non-negative since $u''(.) < 0$ and $c(x) \geq c(\alpha)$ if $x > \alpha$, where this last inequality follows from the first-order condition $V_1(\alpha, \alpha) = 0$, which could be written explicitly by differentiating (2), together with the sufficient condition of Lemma 1. These results imply that the ratio $D(\alpha)/E(\alpha)$ is non-increasing in α. Hence, by (A6), we must have $\mu(\alpha) \leq 0$.

We can now use this result to prove the slightly stronger result that $\mu(\alpha) < 0$ for all $\alpha \in (0, \bar{\alpha})$. Let us assume for a contradiction that there exists an $\tilde{\alpha} < \bar{\alpha}$ such that $c(\tilde{\alpha}) = c(\bar{\alpha})$. As $c(.)$ cannot be decreasing, we must have $c(\alpha) = c(\bar{\alpha})$ for all $\alpha \in [\tilde{\alpha}, \hat{\alpha}]$. This implies, by $V_1(\alpha, \alpha) = 0$, $dy_t(\alpha)/d\alpha \geq 0$ and $dR(\alpha)/d\alpha \geq 0$, that we must also have $y_t(\alpha) = y_t(\bar{\alpha})$ and $R(\alpha) = R(\bar{\alpha})$ for all $\alpha \in [\tilde{\alpha}, \hat{\alpha}]$, i.e. there is bunching at the top. Note

that $\mu(\bar{\alpha}) = 0$ implies, by (13), that:

$$\gamma_t(\bar{\alpha}) = \frac{v'(l_t(\bar{\alpha}))}{u'(c(\bar{\alpha}))}. \tag{A9}$$

We have:

$$\begin{aligned}
\gamma_t(\tilde{\alpha}) - \frac{v'(l_t(\tilde{\alpha}))}{u'(c(\tilde{\alpha}))} &= \gamma_t(\tilde{\alpha}) - \frac{v'(l_t(\tilde{\alpha}))}{u'(c(\bar{\alpha}))}, \\
&= \gamma_t(\tilde{\alpha}) - \frac{v'(l_t(\tilde{\alpha}))}{v'(l_t(\bar{\alpha}))}\gamma_t(\bar{\alpha}), \tag{A10}
\end{aligned}$$

where the first line follows from $c(\tilde{\alpha}) = c(\bar{\alpha})$ and the second from (A9). Note that:

$$v'(l_t(\tilde{\alpha})) = v'\left(\frac{y_t(\tilde{\alpha})}{\gamma_t(\tilde{\alpha})}\right) = v'\left(\frac{y_t(\bar{\alpha})}{\gamma_t(\tilde{\alpha})}\right) \geq v'\left(\frac{y_t(\bar{\alpha})}{\gamma_t(\bar{\alpha})}\right) = v'(l_t(\bar{\alpha})), \tag{A11}$$

where the inequality is strict for the values of $t \in [0, R(\bar{\alpha}))$ such that $\gamma_t(\alpha) < \gamma_t(\bar{\alpha})$ for all $\alpha < \bar{\alpha}$. It follows from (A10) and (A11) that, for some values of $t \in [0, R(\bar{\alpha}))$, we have:

$$\gamma_t(\tilde{\alpha}) - \frac{v'(l_t(\tilde{\alpha}))}{u'(c(\tilde{\alpha}))} < \gamma_t(\tilde{\alpha}) - \gamma_t(\bar{\alpha}) < 0. \tag{A12}$$

But, this together with $\mu(\tilde{\alpha}) \leq 0$ is inconsistent with (13) evaluated at $\tilde{\alpha}$. Hence, there is no bunching at the top and $c(\alpha) < c(\bar{\alpha})$ for all $\alpha < \bar{\alpha}$. The consumption function $c(.)$ being continuous, this implies, by (A8), that $E'(\alpha) > 0$ for all $\alpha < \bar{\alpha}$. Thus, the ratio $D(\alpha)/E(\alpha)$ is strictly decreasing in α and, by (A6), we must have $\mu(\alpha) < 0$ for all $\alpha \in (0, \bar{\alpha})$.[40]

---

[40]Note that, even if the sufficient condition of Lemma 1 does not hold, Lemma 3 remains true provided that consumption is weakly increasing in the productivity index and that it is strictly increasing at the top of the distribution.