CSEF

Centre for Studies in Economics and Finance

WORKING PAPER NO. 266

Mirrlees meets Laibson: Optimal Income Taxation with Bounded Rationality

Matteo Bassi

December 2010

University of Naples Federico II | University of Salerno | Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS – UNIVERSITY OF NAPLES
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unisa.it

WORKING PAPER NO. 266

Mirrlees meets Laibson: Optimal Income Taxation with Bounded Rationality

Matteo Bassi[*]

Abstract
This paper studies an optimal taxation problem in a dynamic economy inhabited by individuals who differ in productivity (as in Kocherlakota, 2006) and in the short-term discount factor. We determine incentive compatible Pareto optimal allocations in a multidimensional screening model where individuals have to report truthfully their types. Moreover, we characterize the optimal non linear tax on capital and labor income that implements such allocations in a competitive equilibrium. Two forms of bounded rationality are considered: in the first one, some individuals discount future payoffs at a higher rate than others (myopia). In this application, the planner respects consumers' sovereignty, and maximizes a Paretian social welfare function. In the second application, some individuals are time inconsistent: they systematically change future plans and regret ex-post for the lack of commitment power. We show that the marginal tax on capital income implementing the optimal allocation consists of several elements, which combine incentive compatibility and bounded rationality considerations. The resulting optimal tax is a decreasing function of both the fraction of short-sighted individuals and the intensity of myopia/hyperbolic discounting. Our results are not driven by the paternalistic behavior of the planner, but by the incentive/self control problem and the necessity of providing the right incentive to high productive, far-sighted individuals. However, when the planner would like to push hyperbolic individuals toward the right consumption/saving path, we show that the optimal marginal tax includes also a paternalistic component that further decrease the optimal tax compared to the case with only exponential agents.

Keywords: Dynamic Optimal taxation, Saving, Capital accumulation, Time Inconsistency, Myopia, Minimal Paternalism, Multidimensional Screening.

JEL Classification: A21, H21.

Acknowledgments: I thank Helmuth Cremer, Georges Casamatta and Pierre Pestieau for helpful comments.

[*] Università di Salerno and CSEF

Address: CSEF, Dipartimento di Scienze Economiche e Statistiche, Università di Salerno, Via Ponte Don Melillo 84084 Fisciano (SA), Italy. E-mail: mbassi@unisa.it, matteo.bassi@gmail.com

Table of contents

1 Introduction

The optimal design of capital and labor income taxes represents a classical issue in public economics.

There is a presumption among economists that capital taxes raise revenues in a less efficient way than wage or consumption taxes (see, for instance, Diamond, 1970, Atkinson and Sandmo, 1980, Judd, 1985 and 1999, Chamley, 1986, Chari, Christiano and Kehoe, 1994). The celebrated Atkinson-Stiglitz (1976) result shows that, under the assumption that individuals' utility function is additively separable between leisure and consumption, labor income taxes does not need to be integrated by taxes on capital income. The Atkinson-Stiglitz result has been challenged by a series of paper by Cremer et al. (2001 and 2003) showing that, even with the separability assumption, if both ability levels and inherited wealth are unobservable, an interest tax represents an optimal instrument to screen individuals' wealth that, being private information, may escape taxation. The optimal design of labor taxes has been deeply explored since the seminal work of Mirrlees (1971). Since then, many contributions have followed (for a review, see Toumala, 1990). Most of thsee contributions are, however, static: individuals' skill levels are randomly assigned at birth, and do not change over time. In reality, individuals' abilities evolve over time: a talented person may awake one day in bad shape, lowering momentarily its ability; workers may have low productivity in a certain job, but higher in another one. Some people learn faster of forget slower than others. Finally, workers' productivity may evolve throughout their career not only exogenously (as a consequence of stochastic shocks), but also through mechanisms that are endogenous to individual decisions (education, learning by doing etc.). These considerations motivate a recent strand of literature, the "The New Dynamic Public Finance" (Kochelakota, 2005a, and Golosov et al. 2006), interested in extending to a dynamic framework Mirrlees' results.

Once that a dynamic framework is considered, it is necessary to look not only at the evolution of skills, but also at the consistency over time of individuals' plans. concerning savings.

Motivated by the observation that saving rates have fallen over last 20 years in most developed countries[1], behavioral economists have studied how self-control problems may play an important in eliciting suboptimal saving rates. Laibson (1997) shows that individuals often report (and regret) that actual saving are lower than planned.

To justify these behaviors, in apparent contrast with the perfect rationality assumption, behavioral economists have begun to devote a growing attention to the "preference reversal/hyperbolic discounting" observed in laboratory experiments[2]. Therefore, if the possibility that individuals' behavior is better

[1] In U.S, as documented by the National Income and Product Accounts (NIPE), saving rates fell from 8.3% of net national product in 1980 to 1.8% in 2003. Such low saving rates have detrimental effects on investments, growth, balance of payments and the financial security of households. Consequently, governments are currently debating the possible public reforms that can affect private savings.

[2] The evidence supporting hyperbolic preferences in any domain is limited (see Ainslie, 1992 and Liabson, 1998), but it is much more stronger when consumption-saving decisions are considered (Liabson D, Repetto and J. Tobacman, 1998). Moreover, as showed by Laibson and Harris (2001), hyperbolic calibrated simulations are able to reproduce the observed high comovement between consumption and income and the drop in post retirement consumption better than exponential calibrated simulations. Moreover, there is no evidence, psychological or other, supporting time consistent preferences over

described by a models of bounded rationality is seriously taken into account, not only a new model of individual decision making is needed, but also a new model of intertemporal taxation of saving appears to be justified.

In this spirit, our paper considers a dynamic model of optimal nonlinear capital and labor income taxes, with individuals differing in two unobservable characteristics: productivity level and degree of self-control, who can be either myopia or time inconsistency. In other words, we assume that not only individuals have different, privately observed and stochastic abilities, but also, although with different degrees, self-control issues. Our aim is to answer several questions: does bounded rationality change the classical consumption-savings problem? If yes, should the governments subsidize saving to correct the bias toward the present? By how much? Which instruments implement the optimal allocation? Does the structure of a nonlinear income taxation should change with rationally bounded consumers?

To our knowledge, there are no papers trying to extend the optimal labor and wealth income tax problem to a dynamic setting *à la* Mirrlees with bounded rationality. However, within this setting, at least thrtwoee difficulties arise. First, bounded rationality makes the informational problem between tax authority and agents more stringent: it is reasonable to assume that both abilities and behavioral type are individuals' private information and therefore not observable by the government. If equity considerations require that high ability individuals should pay a larger share of government spending, those with higher incomes must pay more taxes. As usual, we need to prevent mimicking by making optimal taxes incentive compatible. However, it is not clear which direction redistribution between far-sighted and bounded rational individuals should go. This ultimately depends on the planner's evaluation of individuals' utilities through his social welfare function. This leads to the second possible difficulty; once the planner recognizes individuals' self-control problems, how welfare should be evaluated?

According to Kahneman (1994), a corrective intervention by the government is required whenever the "decision utility", the utility function that each individual maximizes and reflects his choices, and the "experience utility", the utility function that reflects agent's welfare, differ. Other approaches have been followed in the literature: Rabin and O'Donoughe (2001), Gruber and Koszegi (2004) and Cremer et al. (2008) consider a paternalistic planner that seeks to overcome individuals' self-control issues by maximizing a far-sighted social welfare function and taking into account, at the same time, that individuals respond to taxes according to their true preferences. Other approaches include Krusell et al. (2001), where the planner displays itself self-control problem, or Blomquist and Micheletto (2007), where the planner maximizes a utility function that differs from the individual one. Finally, Caplin and Leahy (2000) allow the planner to give more weight to future individuals' utility. Since agents discount the future too much, this social welfare function allows the government to promote more future-oriented policies.

We generalize these approaches by allowing the planner to adapt his preferences to the kind of behav-

hyperbolic ones in any domain (Gruber and Koszegi, 2004).

ioral anomaly displayed by the population. More precisely, when individuals are myopic discount future utilities at a higher rate than far-sighted, the planner fully respects consumers' sovereignty and maximizes the sum of the true individuals' preferences. In this case, the government interprets myopia as a simple stronger preference for the present and corrective interventions seems not to be justified. However, when individuals are time inconsistent and regret ex-post their lack of commitment, planner's paternalism and corrective policies appear to be appropriate. The rationale for government's intervention is subtle: with time inconsistency, individuals plan to save a certain amount, but their hyperbolic preferences lead them to change later those decisions and to regret about this lack commitment. In this case, the tax system may represent an efficient way to correct these behaviors.

Anticipating the results, our model shows that, when individuals differ along two dimensions, myopia and productivity, and the social planner is interested in redistribution, the optimal capital tax rate is lower than the tax that would emerge with only heterogeneity in productivity levels. The lower capital tax rates is not due to the fact that the planner would like to induce the right amount of savings, but to the incentive problem. If the incentive compatibility constraint is binding from time consistent to myopic and to prevent the former mimicking the latter, the optimal capital tax includes a rent for far-sighted, in the form of a higher after-tax return from saving.

When, instead, time inconsistency is considered, previous results are further reinforced: the incentive component still characterizes the expression for the optimal tax, and hyperbolic discounting adds also a self-control element, that further reduces the tax. Intuitively, hyperbolic individuals, looking for a commitment device that force them to save optimally, need a further reduction of the tax with respect to the myopic case. In fact, even if the optimal tax is globally incentive compatible (from the perspective of the initial self there are not incentive to mimic other types), it is possible that, due to the lack of self-control that characterize hyperbolic discounters, future self may have the incentive to untruthfully report their future shocks. To account for this possibility, the optimal tax is further reduced compared to the myopic case, and this additional reduction acts as a commitment device offered to hyperbolics' future selves. Moreover, since hyperbolic individuals regret their lack of commitment power in intertemporal decisions, the planner has a motive for a corrective intervention. If this is effectively the case, the optimal tax includes also a paternalistic component which, depending on the pattern of binding incentive constraints, may increase or decrease the optimal tax.

The paper proceeds as follows: in section 2, we review the main findings of the the three strands of literature we benefit from: the New Dynamic Public Finance Literature (NDPF), Behavioral Economics and Multidimensional Screening, highlighting our main contributions and departures from them. In section 3, we present our general model. Section 4 studies the first application of the model, myopia. Section 5 analyses the second application, time inconsistency. Section 5 briefly consider the case with correlated shocks. Section 6 concludes.

2 Literature Review

2.1 The NDPF Literature

Models of dynamic optimal taxation are based on the Ramsey approach (in which agents are treated symmetrically and the linear tax instruments available to the government are exogenously specified[3]) show that capital should go untaxed in the long run (see, for instance, Chamley and Judd, 1985): it is optimal, in the lung run, to eliminate the wedge between expected marginal utility of investing in capital and the marginal utility of current consumption.

These results have been recently challenged by the "The New Dynamic Public Finance" (NDPF) literature, whose objective is to extend the static Mirrlees (1971) model to a dynamic framework, in which agents' skills are subject, every period, to stochastic shocks. In this richer framework, the zero wedge result does not hold anymore and that it is optimal to discourage savings through a tax on wealth (Golosov et al. 2003).

Our paper is closely related to Golosov, Kocherlakota and Tsyvinski (2003). In a dynamic model where individuals' productiivity is subject, every period, to idiosyncratic shocks, the paper shows the optimality of a positive wedge between the marginal rate of substitution and the individual marginal rate of substitution. Kocherlakota (2005) extends these findings to the case of individual and aggregate shocks. The optimal wedge is still positive, and the optimal allocation includes distortion on saving accumulation. Moreover, the tax system implementig this allocation is fully characterized: the optimal wedge does not translate directly into a positive tax on capital; instead, each individual faces an idiosyncratic wealth tax risk that deters savings: in expected terms, wealth taxes are zero, but may be positive at the individual level. In fact, it exists a negative correlation between each period's wealth tax and consumption level: low-skilled agents at time t face a positive capital tax, to refrain them from saving too much today and lower work effort in the following period.

Albanesi and Sleet (2006) does not consider aggregate shocks and assumes that the idyosincratic skill shocks are *i.i.d.* The tax system that implements the optimal allocation includes a positive tax on capital income conditioned only on current wealth and current skill level. The negative covariance between consumption and wealth continues to hold.

Werning (2006) studies the optimal taxation problem in a two-periods economy with government expenditure and aggregate productivity shocks. Taxation is non-linear and unrestricted as in Mirrlees model. The main result is the characterization of the necessary conditions for perfect tax smoothing: the marginal tax on labor income should remain constant over time and remain invariant in the face of government expenditure and aggregate technology shocks. In addition, the tax on capital income should be zero. However, in this paper, all the idiosyncratic incertitude about individual abilities is resolved after the first period. In fact, as shown by Golosov et al. (2006), the optimal tax smoothing result does

[3] Lump sum trasfers are ruled out by appealing to incentive or administrative contraints.

not hold anymore when incertitude is maintained.

Less attention has been devoted to optimal dynamic wage taxation: an exception is Battaglini and Coates (2005) where, with persistent shock to individuals' abilities and linear utility of consumption, it is shown that optimal marginal labor taxes converge to zero, not only for the top of the income distribution. Balle and Spadaro (2006) develop a model in which agents' productivity evolves over time according to two different factors: an exogenous component and a learning-by-doing process endogenous to fiscal policy. Within this framework, Mirrlees' results change: if the social planner maximize a social welfare function reflecting aversion to inequality, it may be optimal to set a negative marginal labor income tax for high productivity workers.

2.2 Behavioral Public Economics

Merging Public Economics and Psychology is a recent challenge. The literature has shown a growing interest for the experimental and empirical evidence showing that various forms of bounded rationality better describe individuals' behavior. The aim of this new strand of literature is to verify whether policy recommendations are robust to the new behavioral assumption or should be accordingly modified.

Individuals' bounded rationality often justifies government's paternalism: the policy-maker should not fully respect individuals' preferences and consumer sovereignty but should try, through opportune policy instruments, to correct their behavior. A paternalistic intervention would give agents the right incentives for internalizing the cost of their bounded rationality (Herrnstein 1993). However, there is no agreement about paternalism being the appropriate answer to all consumers' mistakes. O'Donoughe and Rabin (2003) stresses that heavy-handed policy interventions, such as banning purchases of goods that some people mistakenly consume, are often inferior to minimal interventions, they can significantly harm those for whom the behavior is rational. Concerns about heavy-handed paternalism led researchers and policymakers to focus on minimally interventionist policies: "cautious paternalism" (O'Donoghue and Rabin, 1999a); "asymmetric paternalism" (Camerer, Issacharoff, Loewenstein, O'Donoghue, and Rabin, 2003), "libertarian paternalism" (Sunstein and Thaler, 2003). All these approaches promote policies that help people who make errors while having little effect on those who are fully rational. Examples include education, minimal modifications of short-term incentives, or setting of easy-to-change default options.

Our paper follows this recent view: the government, through the tax system, helps present-biased consumers to make the right consumption-savings decision. This may be seen as a "minimal "paternalism, as opposed to a command policy, in which the government chooses for the consumer the optimal path of savings.

The literature on optimal taxation with bounded rationality can be classified into two groups: a first one focusing on the effect of present-biased preferences on consumption-savings decisions, and the second one is concerned with the structure of optimal commodity taxation of addictive goods overconsumed by hyperbolic consumers.

A series of papers by Krusell, Kurescu and Smith (2001 and 2005) belong to the first group. The first one, in a Ramsey fremework with symmetric individuals and linear tax on capital and labor income, characterizes the optimal tax policy in an economy inhabited by time inconsistent agents and a time inconsistent government, unable to commit to future taxes. They show that the competitive equilibrium provides higher welfare than the planner's allocation (although lower than the full commitment allocation). The result is driven by the planner's time-inconsistency: whereas consumers take prices as given (and their marginal propensity to save is constant), the planner understands that its decisions affect the return of savings (its marginal propensity to save is decreasing). Then, the planner save less than agents since, from its point of view, increasing savings yields a lower return. The second paper, in a Gul and Pesendorfer's (2001) "temptation" framework, shows that the optimal tax on capital includes a Piguovian subsidy on investments, whose size depends on the degree of temptation/self control problem of the population.

The second group of papers (Rabin and O'Donoghue, 1997 and 2006, and Gruber and Koszegi, 2000) studies the optimal taxation of addictive goods (cigarettes, fatty foods, alcohol etc.). Here, taxes play the role of commitment devices that help agents to behave correctly. Government's intervention is justified on the basis of the double externality imposed by the consumption of addictive goods: a *social* negative externality, in the form of an increase in public medical expenditures due to bad health habits, and also a *private* negative externality, since the addicted individual is imposing now a cost on his future health.

Gruber and Köszegi (2000) study the determination of optimal excises on cigarettes, in a framework in which agents differ for their degree of time inconsistency. They show that the presence of hyperbolic discounters increases optimal excises on harmful goods compared to a model with perfectly rational consumers; in fact, the optimal tax includes two components: the first one that takes into account the social externality imposed by smoking, while the second one, called "self-control adjustment"), provides the consumer with a self-control tool that help him to reduce consumption levels. Finally, they show that the excess burden imposed by the taxation of addictive goods is lower than the burden imposed by taxes of commodities that are exactly as harmful but not addictive.

Rabin and O'Donoghue (2001, 2003 and 2006), in a model with myopic individuals (who discount future payoffs at higher rates) show that excises on addictive goods may increase total surplus, if the tax proceeds are returned lump sum to consumers: sin taxes redistribute income from individuals with self-control problems (wo are not hurt by the tax, since it helps them to reduce the overconsumption of the harmful good) to fully rational consumers.

Blomquist and Micheletto (2006) study the optimal design of income and commodity taxes when government's and individuals preferences differ[4]. They show that both the "no distortion at the top" (Mirrlees, 1971), and "irrelevance of indirect taxation" (Atkinson and Stiglitz, 1976) break down: when-

[4]Their model is very general: no conditions are imposed, except for the fact that the planner does not share consumer preferences about streams of consumptions, but no restrictions upon the difference between individual and social utility function.

ever individuals' MRS between post and pre-tax income differs from the planner's one, skilled workers face a positive marginal tax on income. Moreover, consumption of merit (*resp.* demerit) goods should be encouraged (*resp.* discouraged) through a subsidy (*resp.* tax).

Cremer et al. (2008) analyze the optimal design of a nonlinear social security system where individuals are heterogeneous with respect productivity and degree of myopia. They show that the paternalistic, second best, solution does not necessarily involves forced saving for myopic, since the paternalistic motive is mitigated by the incentive problem: not subsidizing savings allows the planner to relax the incentive constraints.

The empirical robustness of traditional model to these new behavioral assumptions has been widely tested. Petersen (2001) compares the excess burden of different taxes in a standard exponential model and in one with hyperbolic preferences. The aim is to show whether similar policy experiments (changes in labor income tax, capital income tax and consumption tax) give identical results within the two frameworks: if yes, standard time-consistent models are robust to the new behavioral assumptions; if not, and provided that consumers display hyperbolic discounting, the government has to modify accordingly its policies. Simulations show that, if the degree of hyperbolicness is high enough[5], the conventional approach does not hold anymore and recommendations may change: more precisely, capital income taxation is the most detrimental to welfare, with losses that increase significantly the more hyperbolic are the consumers. The reason is intuitive: in the hyperbolic economies, higher capital income taxation decreases the already very low incentive to save. These simulations,therefore, justify the necessity of a normative analysis of capital taxes in presence of various forms of bounded rationality.

2.3 Multidimensional Screening

Our paper follows the optimal taxation literature with multidimensional heterogeneity in individuals' characteristics. Although the models of the one-dimensional population, starting from Mirrlees (1976), have been useful for computations and examinations of optimal income tax problem, they are not accurate representation of reality (see Tuomala, 1990 for a survey of this literature). To analyze redistributional policies more in depth, it is more appropriate to consider situations where individuals are characterized by more than just one parameter of heterogeneity.

The screening problem of multidimensional types has been first studied in the context of nonlinear pricing (Armstrong, 1996, Rochet and Choné, 1998, Armstrong and Rochet, 1999) and later in optimal income taxation theory. Cremer et al. (2001 and 2003) study the properties of capital and labor taxes in an economy populated by individuals with different skills and inherited wealth levels: they show that the informational problem can be overcome through a tax on capital income, since it represents an indirect way to tax wealth, which is unobservable and may escape taxation.

[5]In line with the estimates of Ainslie (1992) and Laibson (2000): "most of experimental evidence suggest that the one-year discount factor is at least 30%-40%, and thus the implied short term discount factor $\beta = 0,7$".

Shapiro (2001) explores the optimal characteristics of poverty assistance programs with individuals differing both in their income generating ability and the disutility of effort. The optimal program resemble a negative income tax with a benefit reduction rate that depends on the distribution of individual characteristics.

Beaudry et al. (2007) examine the optimal tax policy when individuals have private information about their productivities in market and non market activities. The optimal scheme involves upward distortion in market for low skilled individuals (through decreasing wage-contingent employment subsidies), and a downward distortion the market for high skill individuals (through positive and increasing marginal income tax rates).

Tarkiainen and Tuomala (1999 and 2004) consider optimal tax problem when individuals differ with respect to their skill levels and preferences for leisure. The optimal tax system in the two-dimensional case entails more redistribution relative to the one-dimensional one.

3 Setup of the Model

Our framework is similar to Golosov, Kocherlakota and Tsivinsky (2004), Kocherlakota (2006) and Cremer et al. (2008). The economy lasts for T periods, with T finite, and it is inhabited by a continuum of agents of measure one. There exists a single consumption good produced by capital and labor.

Preferences Agents share the same intertemporal utility function:

$$U(\tilde{u}(c_t,l_t),...,\tilde{u}(c_T,l_T)) = \tilde{u}(c_t,l_t) + \sum_{\tau=t}^{T} \beta\delta^{\tau+1-t}\tilde{u}(c_{\tau+1},l_{\tau+1}) \tag{1}$$

where $c_t \in \mathbb{R}$ denotes agent's consumption at time t and $l_t \in \mathbb{R}$ period t labor supply. Parameters β and δ represent, respectively, the subjective short-term discount factor and the long-run one. We make the following assumptions on $\tilde{u}(.)$:

- **A1**: $\tilde{u} : A \subset \mathbb{R}^2_+ \to \mathbb{R}_+$, is continuos and differentiable.
- **A2**: $\tilde{u}(\cdot,\cdot)$ is strictly increasing and strictly concave in c and increasing and strictly convex in l; Inada conditions are satisfied: $\tilde{u}(0,l) = \tilde{u}(c,0) = 0$ and, $\forall l \geq 0$, $\lim_{c\to 0}\tilde{u}(c,l) = +\infty$ and $\lim_{l\to 0}\tilde{u}(c,l) = +\infty$. Finally, $\tilde{u}$ is continuously differentiable at least three times.
- **A3**: $\tilde{u}(\cdot,\cdot) = u(c_t) - v(l_t)$.

Assumptions 1 and 2 are quite standard. Assumption 3 (separability between consumption and leisure in the utility function) is in line both with the static Mirrlees (1971) model and the NPDF literature. Moreover, this specification guarantees that the objective function is strictly increasing (Lemma 1 in Golosov, Kocherlakota and Tsyvinski, 2004).

Idiosyncratic Shocks There are two idiosyncratic shocks in our economy: the first one is on individuals' productivity level: an agent may be more or less productive in different periods of his life. The innovation of our paper is to introduce a second idiosyncratic private shock that affects the short-term discount factor β in (1). The idea behind our assumption is simple: as illustrated in the introduction, intertemporal preferences of some individuals can be better represented through a model with bounded rationality. Modeling bounded rationality as a shock on β allow us to develop a general framework that accounts for the two more common forms of lack of self-control: from one side, some agents simply attach less value to future utilities (*myopia*). From the other side, the value agents attach to sequences of consumption levels may depend on the agents' vantage point. Some actions may be valuated differently ex-post than at the time those actions are taken, and individual may later regret their lack of commitment (*time inconsistency*). We assume that productivity and behavioral shocks are uncorrelated: there are no reasons to assume that a high productive individuals are also time consistent or vice-versa. With perfect correlation, we would return to the one dimensional case already developed in the literature.

Behavioral Shocks The shock on β works as follows: let B a Borel set in $\mathbb{R}$ and let μ_B be a probability measure over the Borel subset of B^T. At the beginning of period 1, an element β^T is drawn for each agent according to μ_β. The draws are independent across agents. The shock on the discount factor β influences individuals' decision about capital accumulation as follows: define a function $b : \Theta^T \times B^T \to (0, \infty)$ that determines *effective savings* at time t: $s_t(\theta^T, \beta^T) = b_t(\theta^T, \beta^T)s_t(\theta^T, \beta^T)$. The function b_t represents agent's time inconsistency in history (θ^t, β^t). The law of large number applies: the fraction of individuals of type β^T in the Borel set B is given, by μ_β. Every agent knows the realization of the shocks β_t at the beginning of period t: it follows that he knows the history of his own shocks: $\beta^t = (\beta_1, ..., \beta_t)$.

Productivity Shocks We model the productivity shock as follows: let Θ be a Borel set in $\mathbb{R}$ and let μ_θ be a probability measure over the Borel subset of θ^T. At $t = 1$, an element θ^T is drawn from μ_θ. The law of large number applies: the fraction of individuals of type θ^T in the Borel sets Θ is given by μ_θ. The shock on productivity impacts on individual's skills in the following way: define a function $\psi_t : \Theta^T \to (0, \infty)$ that determines workers' *effective labor* at time t: $I_t(\theta^T) = \psi_t(\theta^T)l_t(\theta^T)$. The function ψ_t and l_t represents the agent's skill in history (θ^t). Every agent knows the realization of the shock θ_t at the beginning of period t: it follows that he knows the history of his own shocks $\theta^t = (\theta_1, ..., \theta_t)$.

Commitment We assume that private markets are incomplete: insurance against productivity and behavioral shocks are not available, and therefore agents insure themselves by smoothing consumption over time through private saving. In addition, private commitment devices that help time inconsistent agents to follow[6] their long-term plans are not available.This assumption can appear in contrast with

[6] Commitment devices are helpful only for sophisticated consumers aware of their self control problem, and not for naive consumers who believe to be time consistent.

reality: private insurance are well developed and there is a widespread use of savings accounts, such as 401(k) plans for retirement in the U.S. However, firstly, despite the diffusion of private insurances, governments still play an important role in insuring productivity risks. Second, insurance contracts would be quite elaborate to avoid renegotiation. Third, if a self-control device is provided by the private market, the same markets would be also interested in providing also"counter commitment devices" to exploit the inconsistency of consumers (credit cards with unlimited ceilings etc.). Therefore, assuming no commitment for agents is a good approximation of reality.

Allocation An *allocation* in the economy is a three-dimensional vector (c, I, s) such that:

$$\begin{aligned} s &: \Theta^T \times B^T \to \mathbb{R}_+^{T+1} \\ c &: \Theta^T \times B^T \to \mathbb{R}_+^T \\ I &: \Theta^T \times B^T \to \mathbb{R}_+^T \end{aligned}$$

where $c_t(\theta^T, \beta^T)$ and $I_t(\theta^T, \beta^T)$ denote the amount of consumption and effective labor that is assigned to an individual of type (θ^T, β^T). K_{t+1} is the amount of savings carried over from period t into $t+1$. Define K_0 the initial endowment of the economy, and G the exogenous public expenditure at time t. The allocation (c, I, s) is *feasible* if, $\forall t$,

$$C_t + K_{t+1} + G_t \leq F(K_t, L_t) + (1-d)S_t \tag{2}$$

$$L_t = \int_{\beta^T \in B^T} \int_{\theta^T \in \Theta^T} I_t(\theta^T, \beta^T) d\mu_B d\mu_\Theta$$

$$C_t = \int_{\beta^T \in B^T} \int_{\theta^T \in \Theta^T} c_t(\theta^T, \beta^T) d\mu_B d\mu_\Theta$$

$$K_0 \geq K_1 \tag{3}$$

where c_t and I_t represent per capita consumption and per capita effective labor. We write the production function at time t as $F_t(K_t, L_t)$, where $F_t : \mathbb{R}_+^2 \to \mathbb{R}_+$. The function is assumed to have the usual properties: strictly increasing, weakly concave, homogeneous of degree 1 and continuously differentiable at least two times. Notice that, differently from Kocherlakota (2004), the production function itself does not depend directly on the shocks, but only indirectly through K and Y. Factor prices (wage rate for efficiency unit and interest rate) are given by, respectively: $w = F_Y$ and $1 - r = F_K$, where subscripts denote partial derivatives. The wage rate of an individual of productivity is therefore θw. We assume that the accumulated capital does not depreciate from one period to another.

Government's Preferences and Information Structure The social welfare function maximized by the planner depends upon the behavioral anomaly displayed by individuals.

When agents are myopic, differently from Cremer et al. (2008), we consider a *laissez-faire* social planner that fully respects consumers' sovereignty and maximizes the sum of individuals' utilities. Myopic individuals simply have different preferences for the timing of their consumption levels and, as long as they do not regret for their bias towards the present, the planner's intervention is hard to justify. The laissez-faire planner fully is only interested in redistribution and in providing the right incentives to high productivity agents, in order to maximize the total resources available in the economy, taking however into account that myopic may react to incentives differently from rational agents. Moreover, this approach allows us to separate paternalistic and incentive considerations in the optimal tax problem introduced by the second source of heterogeneity.

When some individuals display time inconsistency, the paternalistic approach, according to which the planner is interested in giving to hyperbolic individuals the "correct" amount of consumption, appears to be more appropriate (Blomquist and Micheletto, 2006, Gruber and Köszegi, 2004, O'Donoghue and Rabin, 2006, Cremer et al., 2008). This approach is justified by the fact that hyperbolic consumers regret about the lack of self-control that generated undersaving. Assuming that private commitment devices are not available, the tax policy represents an instrument used by the planner to overcome the inconsistency issue and to drive consumers to the "right" consumption-savings path.

In the spirit of the optimal taxation literature, we assume that neither individual productivity level at time t, θ_t^i, nor labor supply, l_t^i, are observable by the tax administration. Instead, before tax income at time t, $I_t^i = w_t \theta_t^i l_t^i$, is observable. This rules out first best lump sum taxation of labor income. We also consider agents' behavioral type as private information. For simplicity, we assume, as in Cremer et al. (2008), that saving and individuals' consumption levels are publicly observable, so that planner can tax[7] individuals non-linearly over two variables: *effective labor* and return from savings: $T(I_t, rs_t)$, where r is the interest rate. The government has not exogenous revenue requirements; taxes are purely redistributive, *i.e.* $G = 0$. Finally, we assume that the government is able to commit itself, at the beginning of period 1, to the tax policy $(\tau_K,\ T(I))$. Finally, we avoid unnecessary complications by assuming that the government can fully commit to the initial tax system. This can be justified if all the uncertainty about individuals' private information is not revealed after the initial period: if initial skill levels are subject every period to idiosyncratic shocks, the screening function of the system remains valid and the commitment problem is softened.

[7] An alternative specification is to assume that neither savings nor individuals' consumption levels are publicly observable. The tax authority has, however, information on the payment of capital income and anonymous transactions. In this case, the tax policy consists of a nonlinar tax on before tax income, $T(I_t)$ and a linear tax on interests income, τ_K. As shown in Appendix A, the two models provide the same qualitative results. Moreover, as shown by Cremer et al. (2008), in a model similar to ours, any allocation that can be achieved with observable savings can always be implemented with a linear tax, simply by setting a very high capital tax that crows out private savings and redistributing income through period $t + 1$ consumption.

Incentive Compatibility In a second-best setting[8], being the values of the two parameters β and θ_t individuals' private information, the optimal dynamic tax system must induce truthful revelation of both θ and β and allocations should satisfy incentive compatibility conditions.

A *reporting strategy* $\sigma : \Theta^T \times B^T \rightarrow \Theta^T \times B^T$, where $\sigma(\theta^T, \beta^T) = (\theta^{T'}, \beta^{T'})$, for some $\theta^{T'}$ and $\beta^{T'}$. Let be Σ the set of all possible reporting strategies and define as:

$$W(.; c, y) : \Sigma \rightarrow \mathbb{R}$$

$$W(\sigma; c, y) = \int_{B^T} \int_{\theta^T} \left[u(c_t(\sigma)) - v(I_t(\sigma)/\phi_t) + \sum_{\tau=t}^{T} \beta \delta^{\tau+1-t} u(c_{\tau+1}(\sigma)) - v(I_{\tau+1}(\sigma)/\phi_{\tau+1}) \right] d\mu_\theta d\mu_B$$

the expected utility of reporting strategy σ, given an allocation (c, y). We define as σ_{TT} the truth-telling reporting strategy, such that $\sigma_{TT}(\theta^T, \beta^T) = (\theta^T, \beta^T)$. The allocation (c, y, K) is *incentive compatible* if:

$$W(\sigma_{TT}; c, I) \geq W(\sigma; c, I), \forall \sigma \in \Sigma$$

We define *incentive-feasible* an allocation that is both feasible and incentive compatible.

The *optimal* allocation is the (c, y, s) that solves the following maximization problem:

$$\max_{\{c_t^i\}, \{I_t^i\}} \quad \int_{B^T} \int_{\theta^T} \left[u(c_t(\sigma)) - v(I_t(\sigma)/\phi_t) + \sum_{\tau=t}^{T} \widehat{\beta} \delta^{\tau+1-t} \left(u(c_t(\sigma)) - v(I_t(\sigma)/\phi_t) \right) \right] d\mu_\theta d\mu_B \tag{4}$$

subject to

$$W(\sigma_{TT}; c, I) \geq W(\sigma; c, I), \forall \sigma \in \Sigma$$

$$C_t + K_{t+1} + G_t \leq F(K_t, Y_t) + (1-d)K_t, \forall t$$

where $\widehat{\beta}$ in (4) represents the planner's short-term discount factor. As stressed before, if myopia characterizes individuals' preferences, the planner's social welfare function does not reflect paternalism, in the sense that agents' future utilities are discounted at the individuals' true discount factor $\widehat{\beta} = \beta$. However, if individuals are hyperbolic discounters, our framework allows for paternalism: in this case, $\widehat{\beta} = 1$. As a complete and general characterization of the optimal dynamic, time inconsistent, allocation is analytically complicated, in the following section we present two simple examples, that considers two common forms of bounded rationality, myopia and time inconsistency.

[8] We do not characterize first-best allocations, in which individuals' characteristics are publicly observable. Our shortcut can be justified in two ways: first, if we assume that the planner is concerned only by redistribution but not by individuals' bounded rationality, then traditional results in optimal taxation literature applies. The first best allocation entails equal consumption levels, with high-type individuals supplying more labor. Lump-sum transfers from high to low ability agents implement this allocation. Second, if paternalistic consideration are introduced, then the analysis of Cremer et al. (2008) holds: lump-sum transfers should now be supplemented by a Pigouvian subsidy aimed to help individuals to overcome their self-control problem, and save the appropriate amount.

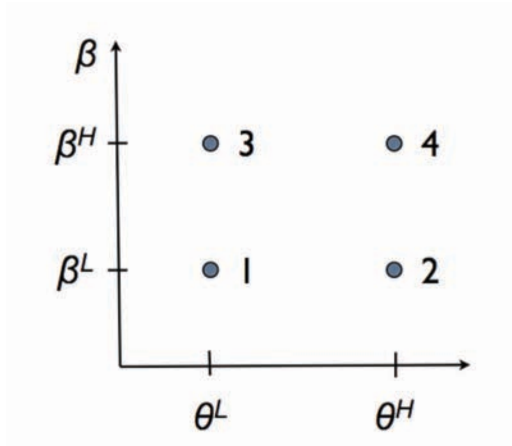

Figure 1: Distribution of Types

4 Application 1 - Myopia

4.1 Setup

In this application, the form of bounded rationality originating asymmetric information between the tax authority and agents is myopia, defined as the situation in which some individuals (*myopic*) attach a lower weight to future utilities relative to far-sighted (or *exponential*) agents. Except for the behavioral assumption, the setup of the model is similar to Cremer et al. (2008).

Our economy is inhabited by a continuum of individuals living for two periods: at $t = 1$, agents supply labor and save for second period consumption; at $t = 2$, they only supply labor and consume savings accumulated at period 1. At the beginning of period 1, each individual receives the shocks on θ and β. At period 2, only the productivity shock affects individual behavior. We assume that productivity can take two values, θ^L and θ^H, with $\theta^H > \theta^L$. Concerning the behavioral shock, there are two possible values for the short-term discount factor, β^H and β^L, with $\beta^H > \beta^L$. We normalize $\beta^H = 1$. It follows that an individual assigned with β^L is myopic. Due to his short-sightedness, at $t = 1$, he saves less (and consume more) than an individual with the same θ but who has been assigned with β^H.

Four types of individuals, indexed by $i = 1, ..., 4$, coexist in our economy, as depicted in Figure 1. The proportion of type i in the population is denoted π_i. Since there is a continuum of individuals, the law of large numbers applies, and proportions π_i coincide with the probability of belonging to group i.

The intertemporal utility function is the same for all individuals:

$$u(c_1^i) - v(l_1^i) + \beta^i \delta \left(u(c_2^i) - v(l_2^i) \right) \tag{5}$$

where c_t^i and l_t^i represent, respectively, individual i's consumption and labor supply at period t.

The production function $F(K, L)$ has constant return to scale with respect to capital, K, and labor,

L, with:

$$L_t = \sum_{i=1}^{4} \pi_i \theta^i l^i \qquad \text{and} \qquad K_{t+1} = \sum_{i=1}^{4} \pi_i (s_t^i + K_t^i) \tag{6}$$

where s_t^i represent saving accumulated at $t = 1$ by individual of type i. We assume that each individual of type i is endowed, at the beginning of his life, with s_0^i units of capital. Total capital available at period 1 is then $K_0 = \sum_{i=1}^{4} \pi_i s_0^i$.

Under perfect competition, firms do not make any profits, and factors are remunerated on the basis of their marginal productivities:

$$1 + r = F_K \qquad \text{and} \qquad w = F_L \tag{7}$$

The resource constraint at $t = 1$ is:

$$K_0 + F(K_0, L_1) = K_1 + \sum_{i=1}^{4} \pi_i c_1^i \tag{8}$$

Whereas at time 2 is:

$$K_1 + F(K_1, L_2) = \sum_{i=1}^{4} \pi_i c_2^i \tag{9}$$

4.2 Individuals' Problem

Given the informational structure highlighted in the introduction, the tax policy at every time t consists of a nonlinear[9] tax on labor income and capital income, rs_t, $T(I_t^i, rs_{t-1}^i)$.

The maximization problem for a representative individual of type i is:

$$\max_{c_1^i, I_1^i, I_2^i} \quad u(c_1^i) - v\left(\frac{I_1^i}{w\theta_1^i}\right) + \beta^i \delta E_{\theta_2}\left(u(c_2^i) - \left(\frac{I_2^i}{w\theta_1^i}\right)\right) \tag{10}$$

subject to:

$$\begin{aligned} c_1^i + s_1^i &= I_1^i - T(I_t^i, rs_0^i) + (1 + r)s_0^i \equiv R_1^i \\ c_2^i &= (1 + r)s_1^i + I_2^i - T(I_2^i, rs_1^i) \equiv R_2^i \end{aligned}$$

where $I_t^i = w_t \theta^i l_t^i$ and R_t^i denotes, respectively, individual's i before tax income and disposable income, obtained from gross income by subtracting the tax on labor income and adding net capital income. Maximization with respect to I_t^i and s_1^i yields to:

$$\frac{u'(c_1^i)}{E_{\theta_2}(u'(c_2^i))} = \beta^i \delta \left(1 + r\left(1 - \frac{\partial T(I_2^i, rs_1^i)}{\partial s_1^i}\right)\right) = \beta^i \delta (1 - \Gamma^i) \tag{11}$$

$$\frac{v'(l_1^i)}{E_{\theta_t}(u'(c_t^i))} = \left(1 - \frac{\partial T(I_t^i, rs_t^i)}{\partial I_t^i}\right) = w\theta_t^i (1 - \Upsilon^i), \quad \forall t \tag{12}$$

[9]The case with linear taxes on capital income is studied in Appendix A.

Define

$$\Gamma^i = r\left(\frac{\partial T(I_2^i, rs_1^i)}{\partial s_1^i} - 1\right) \tag{13}$$

$$\Upsilon^i = \frac{\partial T(I_t^i, rs_t^i)}{\partial I_t^i} \tag{14}$$

which represent implicit marginal tax on, respectively, savings and labor income. When $\Gamma^i > 0$ (< 0), individuals face a marginal tax (subsidy) on savings. If $\Upsilon^i > 0$, individual i's labor income is taxed. In the following, we will concentrate only on the optimal wedge Γ^i, and we let the determination of the optimal wedge Υ^i for further research.

4.3 The Government's Problem

In this subsection, we characterize the shape of the optimal tax function. We assume that the planner is only interested in redistributing resources, and maximizes the following social welfare function:

$$\max_{c_1^i, c_2^i, I_1^i, I_2^i} \quad E_\beta E_\theta \sum_{i=1}^{4} \pi_i \gamma_i \left(u(c_1^i) - v\left(\frac{I_1^i}{w\theta_1^i}\right) + \beta^i \delta E_{\theta_2}\left(u(c_2^i) - v\left(\frac{I_2^i}{w\theta_2^i}\right)\right)\right)$$

Given the informational structure between the tax authority and individuals, allocations must be incentive compatible:

$$V^i \equiv u(c_1^i) - v\left(\frac{I_1^i}{w\theta_1^i}\right) + \beta^i \delta E_{\theta_2}\left(u(c_2^i) - v\left(\frac{I_2^i}{w\theta_2^i}\right)\right) \geq V^{ih} \equiv u(c_1^h) - v\left(\frac{I_1^h}{w\theta_1^i}\right) + \beta^i \delta E_{\theta_2}\left(u(c_2^h) - v\left(\frac{I_2^h}{w\theta_2^i}\right)\right) \tag{15}$$

where (c_1^h, I_1^h) are the consumption-labor supply choices of type i individual mimicking type h. Denoting with η^{ih} and ψ_t the Lagrangean multipliers associated with, respectively, (15) and the resources constraints at periods 1 and 2, the Lagrangean of the problem can be stated as follows:

$$\begin{aligned}
\Lambda^{myopia} =& \sum_{i=1}^{4} \gamma_i \pi_i V^i + \psi_1 \left[\sum_{i=1}^{4} \pi_i K_0^i (1+r) + F\left(\sum_{i=1}^{4} \pi_i K_0^i, \sum_{i=1}^{4} \pi_i I_1^i\right) - \sum_{i=1}^{4} \pi_i R_1^i\right] + \\
&+ \psi_2 \delta \left[\sum_{i=1}^{4} \pi_i K_1^i (1+r) + F\left(\sum_{i=1}^{4} \pi_i K_1^i, \sum_{i=1}^{4} \pi_i I_2^i\right) - \sum_{i=1}^{4} \pi_i R_2^i\right] + \\
&\sum_{i,h} \eta^{ih} (V^i - V^{ih})
\end{aligned} \tag{16}$$

4.4 Optimal Capital Taxes with Myopia

The first order conditions associated to problem (16) are:

$$\frac{\partial \Lambda^{myopia}}{\partial c_1^i} : \left(\pi_i \gamma^i + \sum_{h:h\neq i} \eta^{ih} - \sum_{h:h\neq i} \eta^{hi}\right) u'(c_1^i) - \psi_1 \pi_i = 0 \tag{17}$$

$$\frac{\partial \Lambda^{myopia}}{\partial c_2^i} : \left(\beta^i \pi_i \gamma^i + \delta(\sum_{h:h\neq i} \beta^i \eta^{ih} - \sum_{h:h\neq i} \beta^h \eta^{hi})\right) u'(c_2^i) = 0 \tag{18}$$

Combining the two FOCs, we have:

$$\frac{u'(c_1^i)}{E_{\theta_2}(u'(c_2^i))} = \beta^i \delta \frac{\pi_i \gamma^i + \sum\limits_{h:h\neq i} \eta^{ih} - \sum\limits_{h:h\neq i} \frac{\beta^h}{\beta^i} \eta^{hi}}{\pi_i \gamma^i + \sum\limits_{h:h\neq i} \eta^{ih} - \sum\limits_{h:h\neq i} \eta^{hi}} \tag{19}$$

The expression for the optimal wedge is given by:

$$\Gamma^i = \frac{\sum\limits_{h:h\neq i} (\beta^h - \beta^i) \eta^{hi}}{\pi_i \gamma^i + \sum\limits_{h:h\neq i} \eta^{ih} - \sum\limits_{h:h\neq i} \eta^{hi}} \tag{20}$$

The expression for the optimal distortion on private marginal rate of substitutions is similar to Cremer et al. (2008). The difference is the absence of the paternalistic term, since we assume that the planner is interested only in redistributing resources.

The optimal tax is then composed only by the *incentive* term, which reflects the fact that the optimal allocation is constructed in such a way that mimicking is prevented in all periods.

To understand how the asymmetry of information between individuals and the tax authority modifies the optimal tax, let us assume that there is only a productivity level and that the binding incentive constraints are those from far-sighted to myopic. In this case, it is immediate to see that the more severe is the degree of myopia, the more savings should be subsidized. The intuition is that top individuals (in this very simplifying example, far-sighted) should receive a rent in order to induce them to truthfully report their type. Given that myopic have a lower propensity to save, their allocation (in absence of any paternalistic consideration) is characterized by more consumption in period 1 than in period 2. Far-sighted individuals, who saving more than myopic, may find profitable to mimic myopic, if this strategy gives higher payoffs. However, if the optimal tax system is such that the after tax return from savings is high ($\Gamma^i < 0$), the opportunity cost of mimicking is also high, as mimickers must give up more period 2 consumption when lying about their type. Intuitively, the higher is the degree of myopia, $(\beta^h - \beta^i)$, the lower this subsidy must be. Individuals' myopia makes the incentive problem tighter, and by lowering the tax rate the principal increases the cost of mimicking by making hyperbolic allocation less attractive for exponential. The informational rent ensured to high types assumes in our model the form of a higher after-tax return of saving.

5 Application 2 - Quasi Hyperbolic Preferences

5.1 Setup

This section analyzes a different form of bounded rationality, time inconsistency. To understand how our analysis is modified, we present another model, that extends Mirrlees (1971), Golosov and Tsyvinski (2004), Albanesi and Sleet (2005), Kocherlakota (2006) and Cremer et al. (2008).

Time Inconsistency Time inconsistent individuals, when facing intertemporal trade-offs, change their preferences over time, in such a way that what is preferred at one point in time in the future is incoherent with what is preferred today. Time inconsistency implies that certain agents are more impatient in the short-term than in the long-run. Present-biased individuals would increase their utility level if a commitment device that force them to respect their long run plans would have been made available to them (Laibson, 1997). Laboratory and field studies (Laibson et al. 1998) confirm that discount rates are much greater in the short-run than in the long-run. Following O'Donoghue and Rabin (2007), we consider *sophisticated* hyperbolic discounters, who are aware of their self-control problem, but in absence of commitment devices, as assumed, can not to stick with their optimal plans. The rest of the population includes time consistent individuals (*exponential*), who display exponential discounting and are able to commit themselves, at period 1, to the their entire consumption/saving plan up to period T.

Setting The economy lasts for three periods, $T = 3$, and it is inhabited by a continuum of individuals having preferences additively separable between labor and consumption, as in Application 1.

$$u(c_1) - v(l_1) + \beta\delta\left[(u(c_2) - v(l_2)) + \delta\left(u(c_3) - v(l_3)\right)\right] \tag{21}$$

where β denotes he short-term subjective discount factor, and δ is the long-run one. The increasing functions $u(.)$ and $v(.)$ take specific functional forms[10]:

$$u(c_t) = \frac{c_t^{1-\rho} - 1}{1-\rho}$$
$$v(l_t) = \frac{l_t^2}{2}$$

Shocks Individuals are are heterogeneous with respect to the productivity level θ and the short-term discount factor β. Each variable can take two values: formally, $\Theta = \{\theta_L, \theta_H\}$ and $B = \{\beta_{TI}, \beta_{TC}\}$. Let $\beta_{TC} = 1$. As in Application 1, four types of individuals exist in our economy (see Figure 1), and each of them represents a fraction π_i of the entire population, with $i = 1, 2, 3, 4$. To further simplify, we assume that, while the productivity shock operates every period, the shock on β affects individuals only once, at the beginning of period 1. In the two following periods, the behavioral type remains the same. The (unconditional) probabilities of belonging to group i at time t are given, respectively, by $Pr(\theta_t = \theta_m) \equiv p_m$, for $m = H, L$ and $Pr(\beta_t = \beta_j) \equiv p_j$ for $j = TI, TC$. Since there is a continuum of individuals, the law of large numbers applies, and these probabilities coincide with the fractions π_i. Moreover, the zero correlation assumption between productivity and hyperbolic shocks ensures that the probability of being of type i is the product of the two unconditional probabilities: the fraction of individuals belong to group i at period t is then $\pi_t = p_j p_m$, for $j = TI, TC$ and $m = H, L$.

[10] The specific form for $u(c_t)$ satisfies the condition of Proposition 1 in Treich and Salanié (2007), $P_t(c) - 2A_t(c) \geq 0$, where $P_t(c) \equiv -\frac{u'''(c)}{u''(c)}$ represents the index of absolute prundence, and measures the propensity to increase savings when future become riskier (Kimball, 1990), whereas $A_t(c) \equiv -\frac{u''(c)}{u'(c)}$ is the Arrow-Pratt coefficient for absolute risk aversion.

Government's Preferences and Information Structure With hyperbolic agents, who regret the lack of self-control that generates undersaving, the paternalistic approach (Blomquist and Micheletto, 2006, Gruber and Köszegi, 2004, O'Donoghue and Rabin, 2006, Cremer et al., 2008) is justified, since the social planner is interested in giving to hyperbolic individuals the "correct" amount of consumption. Assuming that private commitment devices are not available, the tax policy is the way the planner tries to overcome the time inconsistency of certain individuals and drive them to the "right" consumption-savings path. To distinguish between the impact of the multidimensional incentive problem and the paternalistic motive of the planner, we present first the case of a laissez-faire planner, and then the case in which the government maximizes a time consistent social welfare function with $\widehat{\beta} = 1$.

The tax policy consists of a nonlinear[11] tax $T(I_t^i, rs_{t-1}^i)$ on labor income at time t and on capital income accumulated at period $t-1$. The exogenous revenue requirement is zero, and taxes are purely redistributive. The government has the power to fully commit itself, at the beginning of period 1, to the tax policy $T(I_t^i, rs_{t-1}^i)$

Production The production function at time t, $F(K, L)$, has constant return to scale with respect to total accumulated capital, K, and total labor supply, L, with:

$$L_t = \sum_{i=1}^{4} \pi_i \theta_i l_t^i \qquad \text{and} \qquad K_{t+1} = \sum_{i=1}^{4} \pi_i s_t^i + K_t \tag{22}$$

where s_t^i represent saving accumulated at time t by individual of type i and K_t capital accumulated in the previous period (in other words, capital depreciation is zero). Each type i is endowed at the beginning of his life with s_0^i unit of capital. Total capital endowment is $K_0 = \sum_{i=1}^{4} \pi_i s_0^i$. Under perfect competition, firms do not make any profits and factors are remunerated according to:

$$1 + r = F_K \qquad \text{and} \qquad w = F_L \tag{23}$$

The resource constraint of the economy, $\forall t$, is:

$$K_t + F(K_t, L_t) = K_{t+1} + \sum_{i=1}^{4} \pi_i c_t^i \tag{24}$$

Preferences We adopt the formulation of Salanié and Treich (2007) to state the individuals' problem. Utility at period 1, before the two shocks are realized, can be written as:

$$E_{\beta_1}\left[E_{\theta_1} u\left(c_1^i, \frac{I_1^i}{w\theta_1^i}\right) + \beta\delta E_{\theta_2}\left(u\left(c_2^i, \frac{I_2^i}{w\theta_2^i}\right)\middle|\theta^1, \beta_1\right) + \beta\delta^2 E_{\theta_3}\left(u\left(c_3^i, \frac{I_3^i}{w\theta_3^i}\right)\middle|\theta^2, \beta_1\right)\right] \tag{25}$$

where θ^{t-1} represents the history of shocks up to period $t-1$.

[11]The case with linear taxes on capital income is studied in Appendix A.

The agent is dynamically inconsistent if the shock received on his short-term discount factor is such that $\beta < 1$. In this case, period 2 preferences are given by:

$$E_{\theta_2}\left(u\left(c_2^i, \frac{I_2^i}{w\theta_2^i}\right)\Big|\theta_1, \beta_1\right) + \beta\delta E_{\theta_3}\left(u\left(c_3^i, \frac{I_3^i}{w\theta_3^i}\right)\Big|\theta_2, \beta_1\right) \tag{26}$$

It follows that, from the point of view of the individual at time 1, the discount factor between period 2 and 3 is δ, but becomes $\beta\delta$ from the point of view of self 2. Our sophistication assumption implies that, when choosing saving at period 1, hyperbolic discounters maximize (25), knowing that saving at period 2 will be chosen by his future self according to (26).

To isolate the effect of lack of self-control from the effect of a higher discount factor, we rewrite model (25)-(26) as follows:

$$E_{\beta_1}\left[E_{\theta_1}u_1\left(c_1^i, \frac{I_1^i}{w\theta_1^i}\right) + E_{\theta_2}\left(u_2\left(c_2^i, \frac{I_2^i}{w\theta_2^i}\right)\Big|\theta_1, \beta_1\right) + \mu E_{\theta_3}\left(u_3\left(c_3^i, \frac{I_3^i}{w\theta_3^i}\right)\Big|\theta_2, \beta_1\right)\right] \tag{27}$$

At period 2, preferences change and become:

$$E_{\theta_2}\left(u_2\left(c_2^i, \frac{I_2^i}{w\theta_2^i}\right)\Big|\theta_1, \beta_1\right) + \lambda E_{\theta_3}\left(u_3\left(c_3^i, \frac{I_3^i}{w\theta_3^i}\right)\Big|\theta_2, \beta_1\right) \tag{28}$$

Notice that model (27)-(28) matches model (25)-(26) if $u_2 = u_3 = \beta\delta u$, $u_1 = u$, $\lambda = \beta\delta$, $\mu = \delta$. In case of a time inconsistent shock on β at period 1, we have that $\lambda < \mu$. If the individual has not received the shock, then $\lambda = \mu$, and period 2 preferences match period 1's ones. This formulation allow us to separate the discount problem from the self control problem: varying λ and keeping constant μ is equivalent to vary β in (26).

5.2 The Individual's Problem

In this section we solve the problem for an individual of type $i = 1, ..., 4$. Taking into account the explicit functional forms for $u(.)$ and $v(.)$ and equations (27)-(28), the problem at time 1 is[12]:

$$\max_{c_1^i, c_2^i, c_3^i, I_1^i, I_2^i, I_3^i} \underbrace{\frac{(c_1^i)^{1-\rho}-1}{1-\rho} - \frac{\left(\frac{I_1^i}{w\theta_1^i}\right)^2}{2}}_{u_1} + \underbrace{\beta\delta E_{\theta_2}\left(\frac{(c_2^i)^{1-\rho}-1}{1-\rho} - \frac{\left(\frac{I_2^i}{w\theta_2^i}\right)^2}{2}\right)}_{u_2} + \underbrace{\beta\delta^2 E_{\theta_3}\left(\frac{(c_3^i)^{1-\rho}-1}{1-\rho} - \frac{\left(\frac{I_3^i}{w\theta_3^i}\right)^2}{2}\right)}_{\mu u_3} \tag{29}$$

At time $t = 1, 2$, individual i's budget constraint is given by:

$$E_{\theta_t}(c_t^i + s_t^i - I_t^i + T(I_t^i, rs_{t-1}^i) - (1+r)s_{t-1}^i) = 0 \tag{30}$$

At time 3, each individual consume alls income:

$$E_{\theta_3}(c_3^i - I_t^i + T(I_3^i, rs_2^i) - (1+r)s_2^i) = 0 \tag{31}$$

[12]Notice that maximization with respect to l_t^i and I_t^i are formally equivalent.

5.2.1 Labor Supply

Labor supply is not affected by time inconsistency, as it is chosen period-by-period. Therefore, $\forall i$, $(l_t^i)^*$ satisfies the following first order conditions, at $t = 1, 2$ and $t = 3$, respectively:

$$\left((1+r)s_{t-1}^i + I_t^i - T(I_t^i, rs_{t-1}^i) - s_t^i\right)^{-\rho} \left(1 - \frac{\partial T(I_t^i, rs_{t-1}^i)}{\partial I_t^i}\right) = \frac{1}{w\theta_t^i} \tag{32}$$

$$\left((1+r)s_2^i + I_3^i - T(I_3^i, rs_2^i)\right)^{-\rho} \left(1 - \frac{\partial T(I_3^i, rs_2^i)}{\partial I_3^i}\right) = \frac{1}{w\theta_3^i}$$

where we have replaced c_2 and c_3 with the budget constraints. First order conditions state that, at the optimum, individuals supply labor in such a way the marginarl benefit of working more, the LHS, is equal to the marginal cost, the RHS.

5.2.2 Savings

To see how time inconsistency modifies consumption/saving choices, let us consider a sophisticated individual ($\beta = \beta^{TI} < 1$) with productivity $m = H, L$. Adopting the alternative formulation (28)-(25), the agent anticipates that time 2 saving will be the solution to the following maximization problem:

$$\begin{aligned}\max_{s_2^{m,TI}} \quad & \frac{\left((1+r)s_1^{m,TI} - s_2^{m,TI} + I_2^m - T\left(I_2^m, rs_1^{m,TI}\right)\right)^{1-\rho} - 1}{1-\rho} - \frac{(I_2^m/\theta_2^m w)^2}{2} \\ & + \lambda E_{\theta_3}\left[\frac{\left((1+r)s_2^{m,TI} + I_3^m - T\left(I_3^m, rs_2^{m,TI}\right)\right)^{1-\rho} - 1}{1-\rho} - \frac{(I_3^m/\theta_3^m w)^2}{2}\right]\end{aligned}$$

which yields the following first order condition:

$$\begin{aligned}&\left((1+r)s_1^{m,TI} - s_2^{m,TI} + I_2^m - T\left(I_2^m, rs_1^{m,TI}\right)\right)^{-\rho} = \\ &\left(1 + r\left(1 - \frac{\partial T(I_3^i, rs_2^{m,TI})}{\partial s_t^i}\right)\right) \lambda E_{\theta_3}\left((1+r)s_2^{m,TI} + I_3^m - T\left(I_3^m, rs_2^{m,TI}\right)\right)^{-\rho}\end{aligned} \tag{33}$$

Rearranging, we obtain an expression for saving at time 2, which are a function of both saving accumulated at $t = 1$ and the parameter of self control, λ:

$$\left(s_2^{m,TI}\right)^* = \frac{(1+r)s_1^{m,TI} + I_2^m - T\left(I_2^m, rs_1^{m,TI}\right) + \lambda^{-\frac{1}{\rho}} E_{\theta_3}\left(I_3^m - T\left(I_3^m, rs_2^{m,TI}\right)\right)}{\lambda^{-\frac{1}{\rho}}\left(1 + r\left(1 - \frac{\partial T(I_3^i, rs_2^{m,TI})}{\partial s_t^{M,TI}}\right)\right)^{-\frac{1+\rho}{\rho}}} \tag{34}$$

In the following, we will refer to this expression as $\widetilde{s}_2^{m,TI} \equiv s(s_1^{m,TI}(r, I_1^m), \lambda, r, I_2^m)$.

At period 1, a sophisticated individual chooses saving by playing a Stackelberg game with his period

2 self[13]:

$$\max_{s_2^{m,TI}} \quad \frac{\left((1+r)s_0^{m,TI} - s_2^{m,TI} + I_1^m - T\left(I_1^i, rs_0^{m,TI}\right)\right)^{1-\rho} - 1}{1-\rho} - \frac{(I_1^m/\theta_1^m w)^1}{2} +$$
$$+ E_{\theta_2}\left[\frac{\left((1+r)s_1^{m,TI} - \widetilde{s}_2^{m,TI} + I_2^m - T\left(I_2^m, rs_1^{m,TI}\right)\right)^{1-\rho} - 1}{1-\rho} - \frac{(I_2^m/\theta_2^m w)^2}{2}\right] +$$
$$+ \mu E_{\theta_3}\left[\frac{\left((1+r)\widetilde{s}_2^{m,TI} + I_3^i - T\left(I_3^m, rs_2^{m,TI}\right)\right)^{1-\rho}}{1-\rho} - \frac{(I_3^m/\theta_3^m w)^2}{2}\right] \tag{35}$$

The first order condition with respect to $s_1^{m,TI}$ is given by:

$$- \left((1+r)s_0^{m,TI} - s_1^{m,TI} + I_1^m - T\left(I_1^m, rs_2^{m,TI}\right)\right)^{-\rho} +$$
$$+ E_{\theta_2}\left((1+r)s_2^{m,TI} - \widetilde{s}_2^{m,TI} + I_2^m - T\left(I_2^m, rs_1^{m,TI}\right)\right)^{-\rho}\left((1+r) - r\frac{\partial \widetilde{s}_2^{m,TI}}{\partial s_1^i}\frac{\partial T(I_2^m, rs_1^{m,TI})}{\partial s_1^{m,TI}}\right) =$$
$$\mu E_\theta\left((1+r)\widetilde{s}_2^{m,TI} + I_3^m - T\left(I_3^i, r\widetilde{s}^{m,TI}\right)\right)^{-\rho}\left(1 + r\left(1 - \frac{\partial T(I_2^m, rs_1^{m,TI})}{\partial s_1^{m,TI}}\right)\right)\frac{\partial \widetilde{s}_2^{m,TI}}{\partial s_1^{m,TI}}$$

Replacing the second term on the LHS with (33), we get:

$$\begin{aligned}
&\left((1+r)s_0^{m,TI} - s_1^{m,TI} + I_1^m - T\left(I_1^m, rs_0^{m,TI}\right)\right)^{-\rho} = \\
&\left(1 + r\left(1 - \frac{\partial T(I_2^m, rs_1^{m,TI})}{\partial s_1^{m,TI}}\right)\right) E_{\theta_3}\left((1+r)\widetilde{s}_2^{m,TI} + I_3^m - T\left(I_3^m, r\widetilde{s}_2^{m,TI}\right)\right)^{-\rho} * \\
&* \left(\mu\left(1 + r\left(1 - \frac{\partial T(I_2^m, rs_1^{m,TI})}{\partial s_1^{m,TI}}\right)\right) - (\mu - \lambda)\frac{\partial \widetilde{s}_2^i}{\partial s_1^{m,TI}}\right)
\end{aligned} \tag{36}$$

notice that, for $\lambda \neq \mu$, this expression is equivalent to the Hyperbolic Euler Equation in Harris and Laibson (2003), whereas, for $\lambda = \mu$, we have the traditional Euler Equation *i.e.* the marginal rate of substitution between consumption at time 1 and expected consumption at time 3 is equal to the marginal cost of consuming more today, $\mu\left(1 + r\left(1 - \frac{\partial T(I_2^m, rs_1^{m,TI})}{\partial s_1^{m,TI}}\right)\right)^2$:

$$\underbrace{\frac{\left((1+r)s_0^{m,TI} - s_1^{m,TI} + I_1^m - T\left(I_1^m, rs_0^{m,TI}\right)\right)^{-\rho}}{E_\theta\left[\left((1+r)\widetilde{s}_2^i + I_3^i - T\left(I_3^i, r\widetilde{s}_2^{m,TI}\right)\right)^{-\rho}\right]}}_{MRS^i_{c_1,c_3}} = \mu\left(1 + r\left(1 - \frac{\partial T(I_2^m, rs_1^{m,TI})}{\partial s_1^{m,TI}}\right)\right)^2$$

Following Salanié and Treich (2007) and Harris and Laibson (2003), we study now how changes in λ affect capital accumulation for hyperbolic discounters. Next proposition shows that time inconsistency leads to underaccumulation of capital and to a drop in third-period consumption levels.

[13] Notice that the exponential individual solves exactly the same problem, but with $\lambda = \mu$.

Proposition 1 *If the utility function is additively separable between consumption and leisure, and utility from consumption has the CRRA specification, then sophisticated individuals save less than exponential.*

Proof. To show that time inconsistency leads to overconsumption, we have to check how the solution to program (35) varies in response to marginal changes in λ. Since decreasing λ represents a loss of self-control, we want to show that $\frac{\partial \widetilde{s}_2^i}{\partial \lambda} > 0$. To see that, let us take the total derivative of the objective function (35) for group i with respect to λ:

$$E_\theta \left[-\left((1+r)s_1^i - \widetilde{s}_2^i + I_2^i - T\left(I_2^i, rs_1^i\right)\right)^{-\rho} + \mu(1+r)\left((1+r)\widetilde{s}_2^i + I_3^i - T\left(I_3^i, r\widetilde{s}_2^i\right)\right)^{-\rho}\right] \frac{\partial \widetilde{s}_2^i}{\partial \lambda}$$

Replacing (33), we obtain:

$$(1+r)E_\theta \left[\left((1+r)\widetilde{s}_2^i + I_3^i - T\left(I_3^i, r\widetilde{s}_2^i\right)\right)^{-\rho}\right] \left((\mu - \lambda)\frac{\partial \widetilde{s}_2^i}{\partial \lambda}\right)$$

that can be written, in a more compact form, as:

$$(1+r)E_\theta u_3'(.) \left((\mu - \lambda)\frac{\partial \widetilde{s}_2^i}{\partial \lambda}\right) \tag{37}$$

By applying the Implicit Function Theorem to (33), we get:

$$\frac{\partial \widetilde{s}_2^i}{\partial s_1^i} = -\frac{(1+r)E_\theta u_3'(.)}{u_2''(.) + \lambda(1+r)E_\theta u_3''(.)(1+r)^2}$$

Replacing into (37) yields to:

$$(\lambda - \mu)\underbrace{\frac{(1+r)^2 E_\theta (u_3'(.))^2}{u_2''(.) + \lambda(1+r)E_\theta u_3''(.)(1+r)^2}}_{X<0}$$

To see whether time inconsistent preferences lead to overconsumption, we have to check the single crossing property, *i.e.* the sign of $(\lambda - \mu)X$. In particular, as shown by Salanié and Treich (2007), this reduces to check, $\forall t$, the sign of $P_t(c) - 2A_t(c)$. In particular, overconsumption occurs if and only if $P_t(c) - 2A_t(c) > 0$. Our specification for the utility function (1) implies that, $\forall t$:

$$P_t(c) - 2A_t(c) > 0 \Leftrightarrow (\rho + 1)c_t^{-1} \geq 2\rho c_t^{-1} \Leftrightarrow \rho \leq 1$$

and the condition is satisfied. ■

Once individuals' choices are known, indirect utility functions for far-sighted and hyperbolic agents, for productivity level $m = H, L$, are given by:

$$V^{m,TC} = \underbrace{u(c_1^{m,TC}) - v\left(\frac{I_1^i}{w\theta_i}\right)}_{\widetilde{U}_1^{m,TC}} + \delta E_\theta \underbrace{\left(u(c_2^{m,TC}) - v\left(\frac{I_2^i}{w\theta_i}\right)\right)}_{\widetilde{U}_2^{m,TC}} + \delta^2 E_\theta \underbrace{\left(u(c_3^{m,TC}) - v\left(\frac{I_3^i}{w\theta_i}\right)\right)}_{\widetilde{U}_3^{m,TC}} \tag{38}$$

$$V^{m,TI} = \underbrace{u(c_1^{m,TI}) - v\left(\frac{I_1^i}{w\theta_i}\right)}_{\widetilde{U}_1^{m,TI}} + \beta\delta E_\theta \underbrace{\left(u(c_2^{m,TI}) - v\left(\frac{I_2^i}{w\theta_i}\right)\right)}_{\widetilde{U}_2^{m,TI}} + \beta\delta^2 E_\theta \underbrace{\left(u(c_3^{m,TI}) - v\left(\frac{I_3^i}{w\theta_i}\right)\right)}_{\widetilde{U}_3^{m,TI}} \tag{39}$$

By adopting the same representation of model (27)-(28), group's i indirect utility function at $t = 1$ can be rewritten as:

$$V^i = E_\beta E_\theta \left(U_1^i + U_2^i + \mu U_3^i \right) \tag{40}$$

and period two utility level as:

$$V_2^i = E_\beta E_\theta (U_2^i + \lambda^i U_3^i) \tag{41}$$

where, $\forall t$, U_t^i is reduced form notation for $U_t(I^{t-1}, I_t^i, R^{t-1}, R_t^i)$, the period t utility level and I^{t-1} and R^{t-1} denote, respectively, the history of individual reports on before tax income and disposable income up to period $t-1$[14].

5.2.3 Wedges

Rearranging individuals' first order conditions, we can characterize the wedges on individuals' marginal rate of substitutions created by the tax system. In particular, combining (32) and (36), we get:

$$\frac{v'(l_1^i)}{E_{\theta_t}(u'(c_t^i))} = \left(\frac{\partial T(I_t^i, rs_t^i}{\partial I_t^i} \right) = w\theta_t^i \Upsilon^i, \quad \forall t \tag{42}$$

where $\Upsilon^i = \frac{\partial T(I_t^i, rs_t^i)}{\partial I_t^i}$ denotes the marginal tax (when $\Upsilon^i > 0$) or subsidy (when $\Upsilon^i < 0$) on labor income. Notice that wedge (42) coincides exactly with the wedge found with myopia (see equation (12)). This is not surprising, as labor supply is not affected by time inconsistency, as it is chosen period-by-period.

On the other hand, rearranging FOC (36) yields to:

$$\frac{u'(c_1^i)}{E_{\theta_3}(u'(c_3^i))} = \mu^i \left[(1+r)(1 + r - \frac{\partial \widetilde{s}_2^i}{\partial s_1^i} \left(1 - \frac{\lambda^i}{\mu^i} \right) \right] - \Gamma^i \tag{43}$$

where $\Gamma^i \equiv \mu^i \left[r \frac{\partial T(I_2^m, rs_1^{m,TI})}{\partial s_1^{m,TI}} \left(2(1+r) - r \frac{\partial T(I_2^m, rs_1^{m,TI})}{\partial s_1^{m,TI}} - \frac{\partial \widetilde{s}_2^i}{\partial s_1^i} \left(1 - \frac{\lambda^i}{\mu^i} \right) \right) \right]$ is the marginal tax (subsidy) on savings.

Notice that $\Gamma^i > 0$ implies that savings of type i are taxed, whereas $\Gamma^i < 0$ implies that savings are subsidized.

This wedge has an intuitive interpretation: for hyperbolic individuals the term $\left(1 - \frac{\lambda^i}{\mu^i}\right)$, *i.e.* the coefficient of hyperbolic discounting, is lower than 1. This implies that hyperbolic discounting lowers the the marginal return from savings compared to exponential agents, for whom $\lambda^i = \mu^i$.

In the following section, we determine how this wedge can be implemented in a competitive equilibrium, taking into account the information problem between the government and agents[15].

[14] Notice that the indirect utility functions are equivalent if we impose $U_2 = \beta^i \delta \widetilde{U}_2$, $U_3 = \beta^i \delta \widetilde{U}_3$, $\mu = \delta$ and $\lambda = \beta^i \delta$. The agent display time inconsistency whenever $\mu > \lambda$.

[15] In this paper we focus on the characterization of Γ^i. The implementation of the optimal Υ^i is more complicated, and it requires further research.

5.3 The Planner's Problems

The planner's objective is to maximize the expected discounted sum of individuals' utilities. We first consider (Section 5.2.1) the case in which the planner fully respect consumers' sovereignty and maximizes the Pareto social welfare function (21). We then consider (Section 5.2.3) a paternalistic planner which maximizes a time consistent social welfare function. Furthermore, in both applications, the planner may weight differently individuals' well-being through the exogenous parameter γ_i, which reflects the social weight attached to group i.

5.3.1 Paretian Planner

With a Paretian planner, the maximization problem is:

$$\max_{s_t^i, I_t^i} \quad E_\beta E_\theta \sum_{i=1}^{4} \pi_i \gamma_i V^i(.) \tag{44}$$

The asymmetric information between the tax authority and individuals requires that the optimal tax system has to satisfy a set of incentive constraints.

First, taxes must be *globally* incentive compatible; defining $R_t^i \equiv I_t^i + (1+r)s_{t-1}^i - T(I_t^i, rs_{t-1}^i)$ individual i's disposable income and I_t^i his before-tax income, utility level V^i has to be at least equal to V^{ih}, the utility level individual i would get by choosing, at any period, any other bundle (I_t^h, R_t^h), for $h = 1, ..., 4$. This implies, $\forall h, i$:

$$V^i \geq V^{ih} \equiv u(c_1^h) - v\left(\frac{I_1^h}{w\theta_1^i}\right) + \beta^i \delta E_{\theta_2}\left(u(c_2^h) - v\left(\frac{I_2^h}{w\theta_2^i}\right)\right) + \beta^i \delta^2 E_\theta \left(u(c_3^h) - v\left(\frac{I_3^h}{w\theta_3^i}\right)\right) \tag{IC}$$

To simplify notation, global incentive constraints can be written as follows:

$$V^i \geq V^{ih} \equiv E_\beta E_\theta \left(U_1^{ih} + U_2^{ih} + \mu U_3^{ih}\right) \tag{45}$$

where U_t^{ih} is individual's i utility level at period t when he reports himself as type h: $U_t^{ih}(I^{t-1}, I_t^h, c^{t-1}, c_t^{ih})$.

Second, taxes should be *temporarily* incentive compatible; this additional incentive constraint, not necessary for application 1, implies that the agent, at time t and after each history of shocks received up to period $t-1$, (θ^{t-1}, β), should be better off by truthfully reporting the shocks rather than lying now and being truthful thereafter. Formally, we have:

$$\forall t, I^{t-1}, I_t^i, I_t^h, c^{t-1}, c_t^i, c_t^{ih} : \tag{TIC1}$$

$$\begin{aligned} &U_t^i(I^{t-1}, I_t^i, c^{t-1}, c_t^i) + \beta^i \sum_{k=t+1}^{T} \delta^{k-1} U_k^i(I^{k-1}, I_k^i, c^{k-1}, c_k^i) \\ \geq\ &U_t^{ih}(I^{t-1}, I_t^h, c^{t-1}, c_t^{ih}) + \beta^h \sum_{k=t+1}^{T} \delta^{k-1} U_k^{ih}(I^{k-1}, I_k^h, c^{k-1}, c_k^{ih}) \end{aligned}$$

In a finite time horizon and for time consistent individuals, it is immediate to see that temporary always implies global incentive compatibility and viceversa (Albanesi and Sleet, 2007). Therefore, these constraints can be neglected in the maximization problem. However, for hyperbolic discounters, this is not necessarily the case. To see why, consider for simplicity our three periods model and a time inconsistent individual who has reported truthfully both shocks at $t = 1$. When computing individual choices, we show that hyperbolic discounting makes what is preferred at time 2 inconsistent with the plans established at time 1. In other words, a time inconsistent agent modifies his preferences (modeled as a difference between μ^i and λ^i) between period 2 and period 3. Therefore, it is possible that this change of preferences may induce these consumers to *temporarily* lie about their type even if the tax system is constructed to be globally incentive compatible. This implies that our allocation need to be incentive compatible also between periods 2 and 3. In our three-periods setting, constraint TIC1 reduces to:

$$\forall t, I^{t-1}, I_t^i, I_t^h, c^{t-1}, c_t^i, c_t^{ih}, \lambda^i : \tag{TIC2}$$

$$U_t^i(I^1, I_t^i, c^{t-1}, c_t^i) + \lambda^i \sum_{k=t+1}^{T} U_k^i(I^{t-1}, I_t^i, c^{t-1}, c_t^i) \geq U_t^h(I^{t-1}, I_t^h, c^{t-1}, c_t^h) + \lambda^i \sum_{k=t+1}^{T} U_k^h(I^{t-1}, I_t^i, c^{t-1}, c_t^h)$$

or, more compactly:

$$V_2^i \geq V_2^{ih} \tag{46}$$

Finally, optimal taxes has to satisfy, at every time t, the following resource constraint[16]:

$$\sum_{i=1}^{4} \pi_i s_t^i + F\left(\sum_{i=1}^{4} \pi_i s_{t-1}^i (1+r), \sum_{i=1}^{4} \pi_i I_t^i\right) - \sum_{i=1}^{4} \pi_i (c_t^i + s_t^i) = 0 \tag{RC}$$

Denoting with η^{ih}, ν^{ih} and ψ_t the Lagrangean multipliers associated with, respectively, IC, TIC2 and RC, the Lagrangean of the problem can be stated as follows:

$$\begin{aligned} \Lambda = & \sum_{i=1}^{4} \gamma_i \pi_i V^i + \sum_t \psi_t \delta^{t-1} \left[\sum_{i=1}^{4} \pi_i K_t^i + F\left(\sum_{i=1}^{4} \pi_i K_t^i, \sum_{i=1}^{4} \pi_i I_t^i\right) - \sum_{i=1}^{4} \pi_i R_t^i\right] + \\ & + \sum_{i,h} \eta_t^{ih} (V^i - V^{ih}) + \sum_{i,h} \nu^{ih} (V_2^i - V_2^{ih}) \end{aligned} \tag{47}$$

[16]As for Application 1, the CRS property of the production function ensures that government's budget constraint is implied by the resource constraint.

5.3.2 Optimal Paretian Taxes

The FOCs associated with problem (47) are:

$$
\begin{aligned}
\frac{\partial\Lambda}{\partial s_1^i} &: \left(\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih} - \sum_{h:h\neq i}\eta^{hi}\right)\left(-\frac{\partial U_1^i}{\partial c_1^i} + \frac{\partial U_2^i}{\partial c_2^i}\left(1+r-\frac{\partial \widetilde{s}_2^i}{\partial s_1^i}\right)\right) + \left(\sum_{h:i\neq h}\nu^{ih} - \sum_{h:h\neq i}\nu^{hi}\right)\frac{\partial U_2^i}{\partial c_2^i}\left(1+r-\frac{\partial \widetilde{s}_2^i}{\partial s_1^i}\right) + \\
&+ \left(\mu^i\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih}\mu^i - \sum_{h:h\neq i}\eta^{hi}\mu^h + \sum_{h:i\neq h}\nu^{ih}\lambda^i - \sum_{h:h\neq i}\nu^{hi}\lambda^h\right)\frac{\partial U_3^i}{\partial c_3^i}\frac{\partial \widetilde{s}_2^i}{\partial s_1^i}(1+r) + \\
&+\pi_i\left(\psi_1 + \delta\psi_2\frac{\partial \widetilde{s}_2^i}{\partial s_1^i} + (1+r)\delta r\left(\psi_2 + \psi_3\delta\frac{\partial \widetilde{s}_2^i}{\partial s_1^i}\right)\right) = 0 \qquad (48)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial\Lambda}{\partial s_2^i} &: -\left(\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih} - \sum_{h:h\neq i}\eta^{hi} + \sum_{h:i\neq h}\nu^{ih} - \sum_{h:h\neq i}\nu^{hi}\right)\frac{\partial U_2^i}{\partial c_2^i} + \qquad (49)\\
&+ \left(\mu^i\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih}\mu^i - \sum_{h:h\neq i}\eta^{hi}\mu^h + \sum_{h:i\neq h}\nu^{ih}\lambda^i - \sum_{h:h\neq i}\nu^{hi}\lambda^h\right)\frac{\partial U_3^i}{\partial c_3^i} + \pi_i\left(\psi_2\delta + \delta^2\psi_3 r(1+r)\right) = 0
\end{aligned}
$$

Replacing (49) into (48), and rearranging, we get an expression for the optimal wedge, *the modified quasi-hyperbolic, incentive-compatible, Euler Equation*:

$$
\begin{aligned}
\frac{u'(c_1^i)}{u'(c_3^i)} = \; & \underbrace{\mu^i\left[(1+r)(1+r-\frac{\partial \widetilde{s}_2^i}{\partial s_1^i}\left(1-\frac{\lambda^i}{\mu^i}\right)\right]}_{I} + \qquad (50)\\
& \underbrace{-(1+r)\left(r + \frac{\sum_{h:h\neq i}\eta^{hi}(\mu^h-\mu^i) + (\sum_{h:h\neq i}\nu^{hi}\lambda^h - \sum_{h:i\neq h}\nu^{ih}\lambda^i)\left(\mu^i(1+r) - \frac{2r}{\mu^i}\frac{\partial \widetilde{s}_2^i}{\partial s_1^i}\right)}{\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih} - \sum_{h:h\neq i}\eta^{hi}}\right)}_{II} + \\
& \underbrace{-\frac{1}{\mu^i}\frac{\partial \widetilde{s}_2^i}{\partial s_1^i}\left[\frac{(\mu^i-\lambda^i)(1+r(\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih}))}{\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih} - \sum_{h:h\neq i}\eta^{hi}} + r\frac{\mu^i\pi_i\gamma_i + \sum_{h:i\neq h}(\eta^{ih}\mu^i - \nu^{ih}\lambda^i) - \sum_{h:h\neq i}(\eta^{hi}\mu^h - \nu^{hi}\lambda^h)}{\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih} - \sum_{h:h\neq i}\eta^{hi}}\right]}_{III} + \\
& \underbrace{+\frac{\pi_i\left(\psi_1 + 2\psi_2\delta + (1+r)^2\delta^2\psi_3\right)}{\pi_i\gamma_i + \sum_{h:i\neq h}\eta^{ih} - \sum_{h:h\neq i}\eta^{hi}}}_{IV}
\end{aligned}
$$

The following proposition illustrates the properties of the optimal marginal tax (subsidy):

Proposition 2 *The marginal tax (subsidy) on savings is given by:*

$$\Gamma^i = II + III - IV \qquad (51)$$

Let us briefly discuss the four terms that compose (50).

Term labeled II was already part of the optimal marginal tax on savings with myopia. The other three terms are peculiar to this application. Term I is the modified Quasi Hyperbolic Euler equation in absence of any distortion induced by taxation (Harris and Laibson, 2003). Term II represents the *incentive* term, which reflects the fact that the optimal allocation is constructed in such a way that mimicking is prevented in all periods. More precisely, it takes into account the double deviation strategy available to mimickers, as reflected by the two elements composing this term: the first one takes into account that, at $t = 1$, from the perspective of individual i's self 1, it could appear profitable to mimic a type h individual at every periods, as captured by the term $\sum_{h:h\neq i} \eta^{hi}(\mu^h - \mu^i)$.

However, this does not guarantee that the allocation would be temporally incentive compatible. The second term, $\left(\sum_{h:h\neq i} \nu^{hi}\lambda^h - \sum_{h:i\neq h} \nu^{ih}\lambda^i\right)$, addresses specifically the incentive problem emerging at $t = 2$. Since hyperbolic discounting causes a change of preferences from period 1 to period 2, it is possible that even if from the perspective of self 1 the optimal strategy is to report truthfully both shocks in all periods, that would not be true anymore for his period 2 self, for whom preferences for the future have changed. The relevance of this effect ultimately depends over the magnitude of the self-control problem, measured by the self control index, $1 - \frac{\lambda^i}{\mu^i}$. The more severe is the lack of self-control, the more savings should be subsidized. Individuals' lack of self-control makes the incentive problem tighter, and by lowering the tax rate the principal increases the cost of mimicking by making hyperbolic allocation less attractive for exponential. The informational rent ensured to high types assumes in our model the form of a higher after-tax return of saving. Furthermore, the weight attached to the incentive term, $\frac{\partial \tilde{s}_2^i}{\partial s_1^i}$, adapts the optimal tax to the effects of time inconsistency on saving behavior. More precisely, the derivative measures how period 2 savings changes in response to a marginal change in period 1 savings. Intuitively, the higher is the absolute value of this derivative (given that it enters with a negative sign into Γ), the greater is the variation of period 2 savings when the after-tax return from savings is marginally increased and the higher the marginal tax can be. It follows that the incentive problem is softened, as the rent that should be guaranteed to high types is lowered.

Term III, the *self-control term*, takes into account the differences between time inconsistency and myopia: in fact, hyperbolic individuals not only discount the future at a higher rate than far-sighted, but also show a lack of self-control that makes them regretting of past choices. This term adapts the optimal tax to this peculiarity: more precisely, the difference $\sum_{h:i\neq h} (\eta^{ih}\mu^i - \nu^{ih}\lambda^i) - \sum_{h:h\neq i} (\eta^{hi}\mu^h - \nu^{hi}\lambda^h)$ captures the fact that individual i can mimic type h degree of time inconsistency. Individuals must find optimal not only to report truthfully their shocks, but also not to misreport their degree of their self-control, $\mu - \lambda$. Notice that, like the incentive term, also the self control term is weighted by $\frac{\partial \tilde{s}_2^i}{\partial s_1^i}$, reflecting that the incentive problem is soften if savings at $t = 2$ are very responsive to marginal changes in period 1 savings.

Term $IV > 0$, the *general equilibrium term*, considers the impact of a variation of the tax on the before tax return from savings.

5.3.3 Paternalistic Planner

If the planner is concerned by the fact hyperbolic individuals regret about their (suboptimal) consumption/savings choices, the social welfare function should be modified as follows:

$$\max_{s_t^i, I_t^i} \quad E_\beta E_\theta \sum_{i=1}^{4} \pi_i \gamma_i \widehat{V}^i(.) \tag{52}$$

where:

$$\widehat{V}^i \equiv u(c_1^i) - v\left(\frac{I_1^i}{w\theta_1^i}\right) + \widehat{\beta}\delta E_{\theta_2}\left(u(c_2^i) - v\left(\frac{I_2^i}{w\theta_2^i}\right)\right) + \widehat{\beta}\delta^2 E_\theta\left(u(c_3^i) - v\left(\frac{I_3^i}{w\theta_3^i}\right)\right)$$

When $\widehat{\beta} = \beta^i$, (52) and (44) coincide. If the planner is interested in correcting hyperbolic choices, then $\widehat{\beta} = 1$. Morever, as in Caplin and Leahy (2000), our formulation allows for the possibility that the planner weighs more future utilities relative to exponential individuals, $\widehat{\beta} > 1 > \beta^i$.

The set of incentive constraints that should be satisfied at the optimal, second-best, allocation is exactly the same as in the laissez-faire case. This is intuitive, as the planner should still prevent global and temporary misreporting of the shocks, irrespectively from the evaluation of individuals' welfare. Finally, notice that individuals' choices continue to be given by (32) and (36).

The planner's Lagrangean becomes:

$$\begin{aligned}\Lambda^{pat} =& \sum_{i=1}^{4} \gamma_i \pi_i \widehat{V}^i + \sum_t \psi_t \delta^{t-1}\left[\sum_{i=1}^{4} \pi_i K_t^i + F\left(\sum_{i=1}^{4} \pi_i K_t^i, \sum_{i=1}^{4} \pi_i I_t^i\right) - \sum_{i=1}^{4} \pi_i R_t^i\right] + \\ &+ \sum_{i,h} \eta_t^{ih}(V^i - V^{ih}) + \sum_{i,h} \nu^{ih}(V_2^i - V_2^{ih})\end{aligned} \tag{53}$$

5.3.4 Optimal Paternalistic Taxes

The first order conditions of the problem with respect to s_1^i and s_2^i are, respectively:

$$\begin{aligned}&\frac{\partial \Lambda^{pat}}{\partial s_1^i} : -\left(\pi_i \gamma_i \frac{\widehat{\beta}}{\beta^i} + \sum_{h:i\neq h} \eta^{ih} - \sum_{h:h\neq i} \eta^{hi}\right) \frac{\partial U_1^i}{\partial c_1^i} + \\ &+ \left(\pi_i \gamma_i \frac{\widehat{\beta}}{\beta^i} + \sum_{h:i\neq h} \eta^{ih} - \sum_{h:h\neq i} \eta^{hi} + \sum_{h:i\neq h} \nu^{ih} - \sum_{h:h\neq i} \nu^{hi}\right) \frac{\partial U_2^i}{\partial c_2^i}\left(1 + r - \frac{\partial \widetilde{s}_2^i}{\partial s_1^i}\right) + \\ &+ \left(\mu^i \pi_i \gamma_i \frac{\widehat{\beta}}{\beta^i} + \sum_{h:i\neq h} \eta^{ih}\mu^i - \sum_{h:h\neq i} \eta^{hi}\mu^h + \sum_{h:i\neq h} \nu^{ih}\lambda^i - \sum_{h:h\neq i} \nu^{hi}\lambda^h\right) \frac{\partial U_3^i}{\partial c_3^i} \frac{\partial \widetilde{s}_2^i}{\partial s_1^i}(1+r) + \\ &+ \pi_i\left(\psi_1 + \delta\psi_2 \frac{\partial \widetilde{s}_2^i}{\partial s_1^i} + (1+r)\delta r\left(\psi_2 + \psi_3 \delta \frac{\partial \widetilde{s}_2^i}{\partial s_1^i}\right)\right) = 0\end{aligned} \tag{54}$$

$$\frac{\partial \Lambda^{pat}}{\partial s_2^i}:-\left(\pi_i\gamma_i\frac{\widehat{\beta}}{\beta^i}+\sum_{h:i\neq h}\eta^{ih}-\sum_{h:h\neq i}\eta^{hi}+\sum_{h:i\neq h}\nu^{ih}-\sum_{h:h\neq i}\nu^{hi}\right)\frac{\partial U_2^i}{\partial c_2^i}+ \tag{55}$$
$$+\left(\mu^i\pi_i\gamma_i\frac{\widehat{\beta}}{\beta^i}+\sum_{h:i\neq h}\eta^{ih}\mu^i-\sum_{h:h\neq i}\eta^{hi}\mu^h+\sum_{h:i\neq h}\nu^{ih}\lambda^i-\sum_{h:h\neq i}\nu^{hi}\lambda^h\right)\frac{\partial U_3^i}{\partial c_3^i}+\pi_i\left(\psi_2\delta+\delta^2\psi_3 r(1+r)\right)=0$$

Combining (55) into (54), we get the expression for the optimal wedge on individuals' marginal rate of substitution.

$$\begin{aligned}\frac{u'(c_1^i)}{u'(c_3^i)}= \quad & \underbrace{\mu^i\left[(1+r)(1+r-\frac{\partial\widetilde{s}_2^i}{\partial s_1^i}\left(1-\frac{\lambda^i}{\mu^i}\right)\right]}_{I}+ \\ & \underbrace{-(1+r)\left(r+\frac{\mu^i\pi_i\gamma_i\frac{\beta^i-\widehat{\beta}}{\beta^i}+\sum_{h:h\neq i}\eta^{hi}(\mu^h-\mu^i)+(\sum_{h:h\neq i}\nu^{hi}\lambda^h-\sum_{h:i\neq h}\nu^{ih}\lambda^i)\left(\mu^i(1+r)-\frac{2r}{\mu^i}\frac{\partial\widetilde{s}_2^i}{\partial s_1^i}\right)}{\pi_i\gamma_i+\sum_{h:i\neq h}\eta^{ih}-\sum_{h:h\neq i}\eta^{hi}}\right)}_{II'}+ \\ & \underbrace{-\frac{1}{\mu^i}\frac{\partial\widetilde{s}_2^i}{\partial s_1^i}\left[\frac{(\mu^i-\lambda^i)(1+r(\pi_i\gamma_i+\sum_{h:i\neq h}\eta^{ih}))}{\pi_i\gamma_i+\sum_{h:i\neq h}\eta^{ih}-\sum_{h:h\neq i}\eta^{hi}}+r\frac{\mu^i\pi_i\gamma_i\frac{\widehat{\beta}}{\beta^i}+\sum_{h:i\neq h}(\eta^{ih}\mu^i-\nu^{ih}\lambda^i)-\sum_{h:h\neq i}(\eta^{hi}\mu^h-\nu^{hi}\lambda^h)}{\pi_i\gamma_i+\sum_{h:i\neq h}\eta^{ih}-\sum_{h:h\neq i}\eta^{hi}}\right]}_{III'}+ \\ & \underbrace{+\frac{\pi_i\left(\psi_1+2\psi_2\delta+(1+r)^2\delta^2\psi_3\right)}{\pi_i\gamma_i+\sum_{h:i\neq h}\eta^{ih}-\sum_{h:h\neq i}\eta^{hi}}}_{IV'}\end{aligned} \tag{56}$$

The following proposition defines the optimal marginal tax on capital income.

Proposition 3 *The marginal tax (subsidy) on savings is given by:*

$$\Gamma^{i\prime}=II'+III'-IV' \tag{57}$$

From (56), it is immediate to see that planner's paternalism only slightly modifies the expression for the optimal wedge. More precisely, two new terms, which take into account the paternalistic behavior of the planner, appear in (56) compared to (50).

A first additional element appear in II', the incentive term: $\frac{\beta^i-\widehat{\beta}}{\beta^i}$. In particular, this term is negative for time inconsistent individuals (those with $\beta^i<\widehat{\beta}$) and zero for exponential: $\beta^i=\widehat{\beta}$. Being negative, this paternalistic term, which coincides with the Pigouvian subsidy in Cremer et al. (2008), further reduces the optimal wedge, and represents a commitment device that helps individuals to stick to their optimal saving plans. In fact, not only incentive considerations obliges the planner to subsidize savings in order to prevent mimicking and to make the allocation made for lower types less attractive to high types, but also paternalism forces now the planner to introduce in the optimal formulae a commitment device, in the form of a higher after tax return from savings, that reduces hyperbolic's overconsumption.

The second "paternalistic" term appears at the numerator of the second part of III', the self-control term: $\frac{\widehat{\beta}}{\beta^i}$. Notice that this term is lower than one for time inconsistent individuals and equal to one for exponential (in this case expression (56) fully coincides with (50)) . It follows that the optimal wedge for hyperbolic is further reduced compared to the Paretian case. Remember that term III represents the self-control term, which take into account the possibility that hyperbolic period 2 self will change preferences in future. If paternalistic considerations are added, the planner's objective results slightly modified: in the Paretian case, the aim of the subsidy was only to avoid mimicking in the second period and to make the allocation temporarily incentive compatible, now there is an additional objective: push individuals towards the "right" consumption/saving path.

To see the "commitment" role played by the two paternalistic terms, one can eliminate the two sources of asymmetric information, and compute the first best allocation, as in Cremer et al. (2008). The resulting optimal allocation can be implemented in equilibrium by a Pigouvian subsidy exactly equal to $\frac{\beta^i - \widehat{\beta}}{\beta^i}$.

6 Concluding Remarks

This paper studies an optimal taxation problem in a dynamic economy inhabited by individuals with self-control issues. In every period, each agent is subject to two idyosincratic shocks, one on his productivity level (as in Kocherlakota, 2006) and one on his short-term, subjective, discount factor. We have characterized Pareto optimal allocations in a multidimensional screening model where individuals have to report truthfully both their shocks, and we construct the tax system that implements, in a competitive equilibrium, the efficient allocation. The government's policy is characterized by a tax function that is non linear in capital and labor income.

We consider two applications, each one considering a particular form of bounded rationality: in the first case, the shock on the short term discount factor makes some individuals myopic, *i.e.* they discount the future at a higher rate than exponential. In the second application, the shock makes some individuals time inconstent *i.e.* their current plans for the future are systematically changed in the future.

In both applications we find that, if the planner is only concerned by redistribution and not by forcing individuals to save correctly (paternalism), it is optimal, in order to provide the right incentives to far-sighted individuals to truthfully report their shocks, to reduce the optimal capital tax rate compared to the tax that would be optimal in an economy with only asymmetric information in productivity levels. As in the standard screening literature, to prevent mimicking, high types must be guaranteed with an informational rent. This rent assumes the form of a subsidy on savings: by lowering the marginal tax on capital income, the principal increases the cost of mimicking by making allocations designed for consumers with self control issues less attractive for far-sighted. As the after tax return of capital increases, the opportunity cost of mimicking someone with self control issues (who save less) increases, and the mimicker finds less profitable to misreport his type.

Adding paternalism to our framework reinforces our results. A paternalistic behavior is justified only with hyperbolic discounting, since time inconsistent individuals not only discount future payoffs at higher rates than far sighted, but also regret ex post for the inability to follow their plans. In this case, the optimal wedge between private and social marginal rate of substitution is further reduced compared to the Paretian planner, as it includes a commitment device term: by subsidizing savings, the planner provides an instrument that allows hyperbolic discounters to increase savings up the optimal, "time consistent" level.

References

[1] Ainslie, G. (1992), *Picoeconomics*, Cambridge University Press, Cambridge MA.

[2] Albanesi, S. and C. Sleet (2005), "Dynamic Optimal Taxation with Private Information", *Review of Economic Studies*, Vol. 73(1), 1-30.

[3] Armstrong, M. (1996), "Multiproduct Nonlinear Pricing", *Econometrica*, Vol. 64(1), 51-75.

[4] Armstrong, M. and J.C. Rochet (1999), "Multidimensional Screening: A User's Guide", *European Economic Review*, Vol. 43(4-6), 959-979.

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

[6] Atkinson, A.B. and A. Sandmo (1980), "Welfare Implications of the Taxation of Savings", *Economic Journal*, Vol. 90, 529-549.

[7] Balle, J. and A. Spadaro (2006), "Optimal Nonlinear Income Taxation in Dynamic Economies", *mimeo.*

[8] Battaglini, M. and S. Coate (2005), "Pareto Efficient Income Taxation with Stochastic Abilities", *NBER Working Papers* N. 10119, National Bureau of Economic Research, Inc.

[9] Beaudry, P., C. Blackorby and D. Szalay (2007), "Taxes and Employment Subsidies in Optimal Redistribution Programs", *Warwick Economic Research Papers Series*, N.779.

[10] Blomquist, S. and L. Micheletto (2006) "Optimal Redistributive Taxation when Government's and Agents' Preferences Differ", *Journal of Public Economics*, Vol. 90(6-7), 1215-1233.

[11] Brito, D., J. Hamilton, S. Slutsky and J.E. Stiglitz (1991), "Dynamic Optimal Income Taxation with Government Commitment", *Journal of Public Economics*, Vol. 44(1), 15-35.

[12] Camerer, C., S. Issacharoff, G. Loewenstein, T. O'Donoghue and M. Rabin (2003), "Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism", *University of Pennsylvania Law Review*, Vol. 151, 2111-2254.

[13] Caplin, A. and J. Leahy (2000), "The Social Discount Rate", *NBER Working Papers* N. 7983, National Bureau of Economic Research, Inc.

[14] Chamley, C. (1986), "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives", *Econometrica*, Vol. 54(3), 607-622.

[15] Chari, V.V., L.J. Christiano and P.J. Kehoe (1994), "Optimal Fiscal Policy in a Business Cycle Model", *Journal of Political Economy*, Vol. 102(4), 617-652.

[16] Corlett, W.J. and D.C. Hague (1953), "Complementarity and the Excess Burden of Taxation", *Review of Economic Studies*, Vol. 21(1), 21-30.

[17] Cremer, H., P. Pestieau and J.C. Rochet (2001), "Direct Versus Indirect Taxation: The Design of The Tax Structure Revisited", *International Economic Review*, Vol. 42(3), 781-799.

[18] Cremer, H., P. Pestieau and J.C. Rochet (2003), "Capital Income Taxation When Inherited Wealth is Not Observable", *Journal of Public Economics*, Vol. 87(11), 2475-2490.

[19] Cremer, H., P. De Donder, D. Maldonado and P. Pestieau (2008), "Forced Saving, Redistribution and Nonlinear Social Security Schemes", *mimeo*.

[20] Della Vigna, S. and U. Malmendier (2004), "Contract Design and Self-Control: Theory and Evidence", *Quarterly Journal of Economics*, Vol. 119(2), 353-402.

[21] Della Vigna, S. and D. Paserman (2005), "Job Search and Impatience", *Journal of Labor Economics*, Vol. 23(3), 527-588.

[22] Diamond, P.A. (1970), "Incidence of an Interest Income Tax", *Journal of Economic Theory*, Vol. 2(3), 211-224.

[23] Kanbur, R., J. Pirttilä and M. Tuomala (2006), "Non-Welfarist Optimal Taxation And Behavioural Public Economics", *Journal of Economic Surveys*, Vol. 20(5), 849-868.

[24] Kocherlakota, N. (2005a), "Annotated Bibliography of The New Dynamic Public Finance", *mimeo*.

[25] Kocherlakota, N. (2005b), "Zero Expected Wealth Taxes", *Econometrica*, Vol. 73(5), 1587-1621.

[26] Krusell P., B. Kuruscu and A. Smith Jr. (2005), "Temptation and Taxation", *mimeo.*

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

[28] Golosov, M., A. Tsyvinski and I. Werning (2006), "New Dynamic Public Finance: A User's Guide", *mimeo.*

[29] Gruber, J. and B. Köszegi (2004), "Tax Incidence When Individuals are Time Inconsistent: the Case of Cigarette Excise Taxes", *Journal of Public Economics*, Vol. 88(9-10), 1959-1987.

[30] Gul, F. and G. Pesendorfer (2001), "Temptation and Self-Control", *Econometrica*, Vol. 69(6), 1403-1435.

[31] Judd, K. L. (1985), "Short-run Analysis of Fiscal Policy in a Simple Perfect Foresight Model", *Journal of Political Economy*, Vol. 93(2), 298-319.

[32] Judd, K. L. (1999), "Optimal Taxation and Spending in General Competitive Growth Models", *Journal of Public Economics*, Vol. 71(1), 1-26.

[33] Kahneman, D. (1994), "New Challenges to the Rationality Assumption", *Journal of Institutional and Theoretical Economics*, Vol. 150(1), 18-36.

[34] Laibson, D. (1996), "Hyperbolic Discount Functions, Undersaving, and Saving Policy", *NBER Working Papers* N. 5635, National Bureau of Economic Research, Inc.

[35] Laibson, D. (1997), "Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics*, Vol. 30(2), 443-477.

[36] Laibson, D. (1998), "Lyfe-Cycle Consumption and Hyperbolic Consumption Functions", *European Economic Review*, Vol. 42(3-5), 861-871.

[37] Laibson, D. and C. Harris (2001), "Dynamic Choices of Hyperbolic Consumers", *Econometrica*, Vol. 69(4), 935-957.

[38] Mirrlees, J. (1971), "An Exploration of the Theory of Optimum Income Taxation", *Review of Economic Studies*, Vol. 38(114), 175-208.

[39] O'Donoghue, T. and M. Rabin (2006), "Optimal Sin Taxes", *Journal of Public Economics*, Vol. 90(6-7), 1825-1849.

[40] O'Donoghue, T. and M. Rabin (2003), "Studying Optimal Paternalism, Illustrated with a Model of Sin Taxes", *American Economic Review Papers and Proceedings*, Vol. 93(2), 186-191.

[41] Petersen, T.W. (2001), "General Equilibrium Tax Policy with Hyperbolic Consumers", *mimeo.*

[42] Roberts, K. (1984), "The Theoretical Limits to Redistribution", *Review of Economic Studies*, Vol. 51(2), 175-195.

[43] Rochet, J.C. and P. Choné (1998), "Ironing, Sweeping and Multidimensional Screening", *Econometrica*, Vol. 66(4), 783-826.

[44] Salanie, B. (2003), *The Economics of Taxation*, MIT Press, Cambridge, MA.

[45] Shapiro, J. (2001), "Income Maintenance Programs and Multidimensional Screening", *mimeo.*

[46] Tarkiainen, R. and M. Tuomala (1999), "Optimal Nonlinear Income Taxation with a Two-Dimensional Population: A Computational Approach", *Computational Economics*, Vol. 13, 1-16.

[47] Tarkiainen, R. and M. Tuomala (2004), "On Optimal Income Taxation with Heterogeneous Work Preferences", *Tampere Economic Working Paper Series*, N. 32.

[48] Thaler, R.H. and C. R. Sunstein (2003), "Libertarian Paternalism", *American Economic Review Papers and Proceedings*, Vol. 93(2), 175-179.

[49] Tuomala M. (1990), *Optimal Income Tax and Redistribution*, Oxford University Press, Oxford UK.

Appendix

A Linear Taxes on Capital Income

This appendix considers the case where savings and individuals' consumption levels are not observable. The tax authority has, however, information on the payment of capital income and anonymous transactions. In this case, the tax policy consists of a nonlinear tax on before tax income, $T(I_t)$ and a linear tax on interests income, τ_K. In the following, we show that the main insights of the nonlinear model presented in the paper continue to hold.

A.1 Myopia

The setting presented in Section 4 remains the same, except for the tax system, which now includes a nonlinear tax on before tax income, $T(I_t)$ and a linear tax on interests income, τ_K. The maximization problem for a representative individual of type i is:

$$\max_{c_1^i, c_2^i, l_1^i, l_2^i} \quad u(c_1^i) - v(l_1^i) + \beta^i \delta \left(u(c_2^i) - v(l_2^i) \right) \tag{A-58}$$

subject to:

$$c_1^i + s_1^i = I_1^i - T(I_1^i) + (1 + r(1 - \tau_K))s_0^i \equiv R_1^i \tag{A-59}$$

$$c_2^i = (1 + r(1 - \tau_K))s_1^i + I_2^i - T(I_2^i) \equiv R_2^i \tag{A-60}$$

where $I_t^i = w_t \theta^i l_t^i$ and R_t^i denotes, respectively, individual i's before tax income and disposable income, obtained from gross income by subtracting the tax on labor income and adding net capital income. Separability of preferences allows us to rewrite the objective function as follows:

$$\max_{R_1^i, R_2^i, I_1^i, I_2^i} \quad V_1(1, p; R_1^i) - v\left(\frac{I_1^i}{w\theta_1^i}\right) + \beta^i \delta \left(V_2(1, p; R_2^i) - v\left(\frac{I_2^i}{w\theta_2^i}\right)\right) \tag{A-61}$$

subject to:

$$R_1^i = I_1^i - T(I_1^i) + (1 + p)s_0^i$$
$$R_2^i = (1 + p)s_1^i + I_2^i - T(I_2^i)$$

where $p = r(1 - \tau_K)$ denotes net capital income, $V_t(.)$, $t = 1, 2$ is the indirect utility function associated with $\tilde{u}(.)$ at time t, obtained by substituting into (A-61) optimal demand functions: $(c_t^i)^* = c(1, p, R_t^i)$ and $(s_1^i)^* = s(1, p, R_t^i)$.

A.1.1 Optimal Capital Taxes with Myopia

The objective of the planner is to maximize the the sum of utilities U^i. We assume that each type i receives, in the social welfare function, a weight $\gamma_i \pi_i$, with $\sum_{i=1}^{4} \gamma_i = 1$. The following budget constraint has to be satisfied:

$$\sum_{i=1}^{4} \pi_i (T(I_1^i) + T(I_2^i) + r(s_0^i + s_1^i)\tau_K) = 0 \tag{A-62}$$

Being the government only concerned by redistribution, we assume that public expenditures G are zero. Moreover, because of the asymmetry of information between the planner and individuals, a set of incentive compatibility constraints need to be satisfied. More precisely, the utility level of an individual of type $i = 1, ..., 4$, when consumes, at time $t = 1, 2$, the bundle (I_t^i, R_t^i):

$$U^i = V_1(1, p; R_1^i) - v\left(\frac{I_1^i}{w_t \theta_i}\right) + \beta^i \delta \left(V_2(1, p; R_2^i) - v\left(\frac{I_2^i}{w_t \theta_2^i}\right)\right)$$

has to be at least equal to the utility level he would get by choosing any other bundle (I_t^h, R_t^h), for $h = 1, ..., 4$. The incentive compatibilty constraints can be written as follows, $\forall h, i$:

$$U^i \geq U^{ih} \equiv V_1(1, p; R_1^{ih}) - v\left(\frac{I_1^h}{w_t \theta_i}\right) + \beta^h \delta \left(V_2(1, p; R_2^{ih}) - v\left(\frac{I_2^h}{w_t \theta_2^i}\right)\right)$$

where

$$R_1^{ih} = R_1^h + s^i - s^h$$
$$R_2^{ih} = R_2^h + (s^i - s^h)(1 + r)$$

are, respectively, the disposable income of the mimicker in period 1 and 2. The difference $s_t^i - s_t^h$ represent the extra amount of saving that the mimicker can obtain. Denoting with λ_t the Lagrangean multiplier associated with the resources constraints and μ_{ih} the one associated with the incentive compatibility constraint, the Lagrangean of the problem is:

$$\Lambda_3 = \sum_{i=1}^{4} \gamma_i \pi_i U^i + \lambda_1 \left[\sum_{i=1}^{4} \pi_i s_0^i + F\left(\sum_{i=1}^{4} \pi_i s_0^i, \sum_{i=1}^{4} \pi_i l_1^i \right) - \sum_{i=1}^{4} \pi_i R_1^i \right] + \tag{A-63}$$
$$\delta\lambda_2 \left[\sum_{i=1}^{4} \pi_i (R_1^i - c_1^{i*}) + F\left(\sum_{i=1}^{4} \pi_i (R_1^i - c_1^{i*})(1+p), \sum_{i=1}^{4} \pi_i l_1^i \right) - \sum_{i=1}^{4} \pi_i c_2^i \right] + \sum_{i,h} \pi_i \mu_{ih} (U^i - U^{ih})$$

Maximization of (A-63) with respect to R_t^i gives us the following first order condition:

$$\frac{\partial \Lambda_3}{\partial R} = \left(\gamma_i \pi_i + \sum_h \mu_{ih} \right) \frac{\partial U^i}{\partial R} - \lambda \pi_i \left[1 - \delta - \delta \frac{\partial F}{\partial K} \pi_i (1+p) \left(1 - \frac{\partial c^{i*}}{\partial R} \right) \right] - \sum_h^4 \mu_{ih} \frac{\partial U^{ih}}{\partial R} = 0 \tag{A-64}$$

where:

$$\frac{\partial U^i}{\partial R} = \frac{\partial V_1^i}{\partial R} + \beta^i \delta \frac{\partial V_2^i}{\partial R} (1+p)$$
$$\frac{\partial U^{ih}}{\partial R} = \frac{\partial V_1^{ih}}{\partial R} + \beta^i \delta \frac{\partial V_2^{ih}}{\partial R} (1+p)$$

represent, respectively, intertemporal marginal utility of income for individuals of type i and individuals of type i mimicking type h and V_t^i and V_t^{ih} are the reduced forms of $V_t(1, p, R^i)$ and $V_t(1, p, R^{ih})$, for $t = 1, 2$.

On the other hand, maximization of (A-63) with respect to $p = r(1 - \tau_K)$ gives:

$$\frac{\partial \Lambda_3}{\partial p} = \sum_{i=1}^{4} \left(\gamma_i \pi_i + \sum_h \mu_{ih} \right) \frac{\partial U^i}{\partial p} - \lambda \delta \left[\frac{\partial F}{\partial K} \sum_{i=1}^{4} \pi_i \left(\frac{\partial c_1^{i*}}{\partial p} (1+p) + s_1^i \right) + \sum_{i=1}^{4} \pi_i \frac{\partial c_2^{i*}}{\partial p} \right] - \sum_{i,h} \mu_{ih} \frac{\partial U^{ih}}{\partial p} = 0 \tag{A-65}$$

where:

$$\frac{\partial U^i}{\partial p} = \frac{\partial V_1^i}{\partial p} + \beta^i \delta \frac{\partial V_2^i}{\partial p}$$
$$\frac{\partial U^{ih}}{\partial p} = \frac{\partial V_1^{ih}}{\partial p} + \beta^i \delta \left(\frac{\partial V_2^{ih}}{\partial p} + \frac{\partial V_2^{ih}}{\partial R} (s_1^i - s_1^h) \right)$$

Following Cremer et al. (2001 and 2003), (A-64) and (A-65) can be combined as follows:

$$\sum_{i=1}^{4} \frac{\partial L}{\partial R} (c_1^i + c_2^i) + \frac{\partial L}{\partial p} \tag{A-66}$$

Denoting with $\widetilde{c_1}^i$ and $\widetilde{c_2}^i$ compensated demand functions for consumption at $t = 1$ and $t = 2$, Slutsky Equation implies that:

$$\frac{\partial c_t^i}{\partial p} = \frac{\partial \widetilde{c_t}^i}{\partial p} - \frac{\partial c_t^i}{\partial R} c_t^i \qquad t = 1, 2 \tag{A-67}$$

Moreover, by Roy's Identity:

$$c_t^i = -\alpha_t^i \frac{\partial V_1^i}{\partial p} \qquad t = 1, 2 \tag{A-68}$$

where $\alpha_t^i = \frac{\partial V_t^i}{\partial R}$ denotes time t marginal utility of income for an individual of type i. Replacing (A-67) and (A-68) into (A-66), we obtain an expression for the optimal τ_K^*:

$$\underbrace{(1+p)\left[(1+r)\lambda\sum_{i=1}\pi_i\left(-\frac{\partial \widetilde{c_1}^i}{\partial p} + (1-\delta-\delta(1+r))(c_1^i + c_2^i) - \frac{\partial c_1^i}{\partial R}c_2^i\right) + \left(\sum_{i=1}\pi_i\gamma_i\beta^i\delta\alpha_2^i c_2^i\right) + \sum_{i,h}\mu_{ih}\beta^i\delta c_1^i\left(\alpha_2^i - \alpha_2^{ih}\right)\right]}_{I} +$$

$$+ \underbrace{\sum_i \gamma_i\pi_i\alpha_1^i + \sum_h \mu_{ih}c_2^i\left(\alpha_1^i - \alpha_1^{ih}\right)}_{II} - \underbrace{\lambda\sum_i\pi_i\left(\delta\frac{\partial \widetilde{c_2}^i}{\partial p} + (1+r)s^i - \delta\frac{\partial c_2^i}{\partial R}\right)}_{III} -$$

$$- \underbrace{\sum_{i,h}\mu_{ih}\left(\alpha_1^{ih}(c_1^{ih} - c_1^i) + \beta^i\delta\alpha_2^{ih}(c_2^{ih} - c_2^i) + \beta^i\delta\alpha_2^{ih}(s^i - s^h)\right)}_{IV} = 0$$

Notice that wheres term I is always positive, $II - III - IV$ can be either positive or negative: the optimal instrument can be either a tax or a subsidy.

Proposition 4 *The optimal tax (subsidy) rate on capital income, τ_K^* is given by:*

$$\tau_K^* = \frac{1+r}{r} + \frac{II - III - IV}{rI} \tag{A-69}$$

Expression (A-69) reflects, at the same time, the effect of myopia, efficiency and incentive compatibility on the optimal tax rate[17].

First, derivatives $\frac{\partial \widetilde{c_1}^i}{\partial p} < 0$ and $\frac{\partial \widetilde{c_2}^i}{\partial p} > 0$ measure, respectively, the (compensated) substitution effects associated with the capital tax on first and second period consumption.

Substitution terms reflect efficiency considerations, as in Ramsey's (1927) inverse elasticity rule, and are not related to the behavioral assumption. The optimal tax should be a decreasing function of both elasticities: rising[18] p, the net price of period one consumption, decreases the numerator (term III) and increases the denominator (term I) of (A-69). It follows that, the more elastic are demand functions at $t = 1$ and $t = 2$, the higher are the inefficiencies/distortions created on saving accumulation, and the lower should be the tax.

Furthermore, expression (A-69) is an increasing function (terms I and II) of $\alpha_t^i - \alpha_t^{ih}$, the difference in marginal utilities of income when disposable income are, respectively, R_t^i and R_t^{ih}. Since the indirect utility function is increasing and concave in R, $\alpha_t^i - \alpha_t^{ih}$ has the same sign as $R_t^i - R_t^{ih}$. However, without knowing the pattern of the incentive constraints, we can establish the effect of these terms on the optimal

[17]We comment expression (A-69) assuming that $\tau_K^* > 0$. If the value of the parameters is such that $\tau_K^* < 0$, all the intuitions remain the same.

[18]Rising p has the same effects of a reduction of τ_K.

tax rate. A priori, it seems more likely that exponential would like to mimic myopic, but this will ultimately depends on the joint distribution of β and θ.

Term IV addresses specifically the impact of the incentive problem generated by myopia upon the optimal tax. To understand why, one has to fix a productivity level, in order to transform the problem into a one-dimensional screening and assume that the binding incentive constraint is from exponential to myopic: the former is mimicking the latter, in order to receive more consumption today, but being able to transfer more resources at $t = 2$ because of his higher propensity to save. In this case, mimicker's consumption is higher than the truth teller at $t = 1$ ($c_1^{ih} > c_1^i$) but lower at $t = 2$ ($c_2^{ih} < c_2^i$).

Term IV shows that τ_K^* is an increasing in the difference in consumption levels in the second period, $|c_2^{ih} - c_2^i|$, and a decreasing function of the difference in consumption levels in the first period, $|c_1^{ih} - c_1^i|$. The intuition for this result is simple: if mimicking a myopic assures much more consumption today to an exponential (if, for instance, β is very low), the incentive problem of the tax authority becomes stronger; to overcome that, the principal has to increase the opportunity cost of mimicking; by lowering τ_K^*, the mimicker loses the (increased) net income on his accumulated saving. The informational rent provided to high types assumes in our model the form of a higher after-tax return from savings. However, at $t = 2$, the mimicker has to receive less consumption than the truth teller. As this difference increases, for a given τ_k, the incentive problem become softer; in fact, an exponential has less incentive to mimic a myopic if the drop in consumption in the second period is high enough. In this case, *ceteris paribus*, the planner can rise the tax rate, since the incentive problem is now less strong. Notice, moreover, that the differences in consumption levels are weighted by α_t^{ih}, the marginal utility of income of the mimicker.

Terms I and II in (A-69) show that τ_K is increasing in $\frac{\partial c_t^i}{\partial R}$, the income effect associated with true consumption levels at $t = 1$ and $t = 2$[19]. The intuition behind this result is the following: income effects measures how period t consumption responds to changes in disposable income R. Therefore, a very responsive c_t implies that small changes in R creates large variation in consumption levels. The implication for our incentive problem, always assuming that the binding incentive constraint is the one from exponential to myopic, is that the informational rent (a higher after tax return from saving) can be reduced (and the tax rate, increased) since small variations of R due to a higher increases a lot c_t^i: the equity/efficiency trade-off is softened, as it is less costly to give to high types the informational rent.

Finally, τ_K^* is decreasing in $s^i - s^h$, the difference between the mimicker's true saving and the savings of the mimicked. The higher is this difference (and thus the incentive to mimic), the more necessary is to provide a rent to high types, in the form of higher after-tax return on saving, in order to increases their period two consumption. [20].

Summing up, we show that, when individuals differ along two dimensions, myopia and productivity, and the social planner is not concerned by paternalism, the optimal capital tax rate is lower than the

[19] Income effects are positive provided that consumption at time 1 and 2 are normal goods.

[20] Moreover, given that s^h is a decreasing function of β, it follows that increasing the degree of myopia further reduces τ_K.

optimal tax with only asymmetric information in productivity levels. This result is not driven the fact that the planner would like to induce the correct saving accumulation, but to the incentive problem. Supposing that the binding constraint is the one from far sighted to myopic, the optimal capital tax must include a rent, in the form of a higher after tax return from saving, to avoid the former mimicking the latter. Our results are in line with Cremer et al. (2008), which show how incentive considerations modify the structure of the optimal tax rate with a paternalistic planner interested in providing myopic with the right amount of savings. Our result is, however, more general: we show that, in absence of any paternalistic considerations, the optimal tax rates may be a decreasing function of the fraction of myopic in the population and their degree of myopia. In order to provide the right incentives to high types, the planner has to provide them an informational rent, in the form of an higher after tax return from savings.

A.2 Quasi-Hyperbolic Discounting

With hyperbolic preferences and linear taxes on capital income, the analysis presented in section 5 does not change substantially. In particular, Proposition 2 (overconsumption) and the setting of the planner's problem remains qualitatively the same:

$$\begin{aligned}\Lambda_4 =&\sum_{i=1}^{4}\gamma_i\pi_i V^i + \sum_t \psi_t\delta^{t-1}\left[\sum_{i=1}^{4}\pi_i K_t^i + F\left(\sum_{i=1}^{4}\pi_i K_t^i, \sum_{i=1}^{4}\pi_i I_t^i\right) - \sum_{i=1}^{4}\pi_i R_t^i\right] + \\ &+\sum_{i,h}\eta_t^{ih}(V^i - V^{ih}) + \sum_{i,h}\nu^{ih}(V_2^i - V_2^{ih})\end{aligned} \tag{A-70}$$

A.2.1 Optimal Capital Taxes with Time Inconsistency

Optimal capital tax τ_K are obtained by maximizing (A-70) with respect to R_t^i and $p = r(1-\tau_K)$. After having replaced $V^i, V^{ih}, V_2^i, V_2^{ih}$ with their expression, the following first order conditions result:

$$\begin{aligned}\frac{\partial \Lambda_4}{\partial R^i} :& \left(\gamma^i\pi_i + \sum_h \eta^{ih}\right)\left(\frac{\partial U_1^i}{\partial R^i} + \frac{\partial U_2^i}{\partial R^i} + \mu\frac{\partial U_3^i}{\partial R^i}\right) - \sum_h \nu^{ih}\left(\frac{\partial U_2^i}{\partial R^i} + \lambda^i\frac{\partial U_3^i}{\partial R^i}\right) - \sum_h(\eta^{ih} - \nu^{ih})\frac{\partial U_2^{ih}}{\partial R^i} - \\ &+\sum_t \psi_t\delta^{t-1}\pi_i\left[1 - \left(1 - \frac{\partial c_t^i}{\partial R}\right)(1+p)(2+r))\right] - \sum_{i,h}\eta^{ih}\frac{\partial U_1^i}{\partial R^i} - -\sum_h(\eta^{ih}\mu - \nu^{ih}\lambda^i)\frac{\partial U_3^{ih}}{\partial R^i} = 0\end{aligned} \tag{A-71}$$

$$\begin{aligned}\frac{\partial \Lambda_4}{\partial p} :& \sum_{i=1}^{4}\left(\gamma^i\pi_i + \sum_{i,h}\eta^{ih}\right)\left(\frac{\partial U_1^i}{\partial p} + \frac{\partial U_2^i}{\partial p} + \mu\frac{\partial U_3^i}{\partial R^i}\right) - \sum_h \nu^{ih}\left(\frac{\partial U_2^i}{\partial p} + \lambda^i\frac{\partial U_3^i}{\partial p}\right) - \sum_{i,h}(\eta^{ih} - \nu^{ih})\frac{\partial U_2^{ih}}{\partial p} - \\ &+\sum_t \psi_t\delta^{t-1}\sum_{i=1}^{4}\pi_i\left[((2+r)(1+p)+1)\left(s_t^i + \frac{\partial s_t^i}{\partial p} - \frac{\partial c_t^i}{\partial p}\right)\right] - \sum_{i,h}\eta^{ih}\frac{\partial U_1^i}{\partial p} - \sum_{i,h}(\eta^{ih}\mu - \nu^{ih}\lambda^i)\frac{\partial U_3^{ih}}{\partial p} = 0\end{aligned} \tag{A-72}$$

Roy's Identity and Slutsky equations allow us to combine first order conditions (A-71) and (A-72) as follows:

$$\sum_{i=1}^{4}\frac{\partial L}{\partial R^i}\sum_t c_t^i+\frac{\partial L}{\partial p}$$

to obtain:

$$\underbrace{\sum_i \gamma^i\pi_i\left[\alpha_1^i(c_2^i+c_3^i)+\alpha_2^i(c_1^i+c_3^i)+\mu\alpha_3^i(c_1^i+c_2^i)\right]}_{I}+\underbrace{\sum_{i,h}\eta^{ih}\left[(\alpha_1^i-\alpha_1^{ih})(c_2^i+c_3^i)-\alpha_1^{ih}((c_1^{ih}-c_1^i)\right]}_{II}+$$

$$+\underbrace{\sum_{i,h}\left(\eta^{ih}-\nu^{ih}\right)\left[(\alpha_2^i-\alpha_2^{ih})(c_1^i+c_3^i)-\alpha_2^{ih}((c_2^{ih}-c_2^i)\right]}_{III}+\underbrace{\sum_{i,h}\left(\eta^{ih}\mu-\nu^{ih}\lambda^i\right)\left[(\alpha_3^i-\alpha_3^{ih})(c_1^i+c_2^i)-\alpha_3^{ih}((c_3^i-c_3^{ih})\right]}_{IV}+$$

$$\underbrace{+(1+p)\left[\sum_i\sum_t\psi_t\delta^{t-1}\pi_i\left((2+r)\left(s_t^i+\frac{\partial s_t^i}{\partial p}+(1+r)c_t^i\right)\frac{\partial \widetilde{c_t}^i}{\partial p}(1+r)-\frac{\partial c_t^i}{\partial p}\right)\right]}_{V}=\underbrace{\sum_t\sum_i\left(\frac{\partial c_t^i}{\partial p}-\psi_t\delta^{t-1}\pi_i\left(s_t^i+\frac{\partial s_t^i}{\partial p}\right)\right)}_{VI}$$

where $\forall t, \widetilde{c_t}^i$ denotes compensated demands functions of the consumption good, $\frac{\partial \widetilde{c_t^i}}{\partial p}$ the associated substitution effect and α_t^i and α_t^{ih} marginal utilities of income for individuals with, respectively, disposable income R_t^i (truth-tellers) and R_t^{ih} (mimickers).

The following proposition defines the properties of the optimal tax rate on savings with time inconsistent preferences and linear capital taxes.

Proposition 5 *The optimal tax (subsidy) rate on capital income, τ_K^* is given by:*

$$\tau_K^*=\frac{1+r}{r}+\frac{I+II+III+IV-VI}{rV} \tag{A-73}$$

Expression (A-73) reflects both the equity/efficiency and the incentive/efficiency trade-offs.

First, terms $\frac{\partial \widetilde{c}_t^i}{\partial p}$ in (A-73) measure the substitution effects associated with the capital tax on (compensated) consumption levels at time t. As usual, to minimize the deadweight loss associated with the tax, the higher is the elasticity of demand (and the substitution effect) for period t consumption, the lower the tax should be.

As for the optimal tax with myopia (see (A-69)), also (A-73) is increasing in $\alpha_t^i-\alpha_t^{ih}$, the difference between marginal utilities of income when disposable income are R_t^i and R_t^{ih}. Since indirect utility functions are increasing and concave in R: therefore, the sign of $\alpha_t^i-\alpha_t^{ih}$ depends on the sign of $R_t^i-R_t^{ih}$. However, without knowing the pattern of the incentive constraints, we can establish the effect of these terms on the optimal tax rate.

The terms labeled II, III and V address specifically the impact of time inconsistency upon the optimal capital tax. To fix ideas, let us assume that there exists only one productivity level, in order to transform the problem into a one-dimensional screening. Moreover, assume that the binding incentive constraint is the one from time consistent to sophisticated individuals: the former group is mimicking the latter, in order to receive more consumption today, but being able to transfer more resources in period due to the higher propensity to save. In this case, consumption levels of the mimicker are higher at $t = 1$ and $t = 2 (c_t^{ih} > c_t^i)$ but lower at $t = 3$, $(c_3^{ih} < c_3^i)$. According to II, III and V, the optimal tax is an increasing function of the difference in consumption levels in the last period, $|c_3^{ih} - c_3^i|$ and a decreasing function of the difference in consumption levels in the first and second periods, $|c_t^{ih} - c_t^i|$. Intuitively, if mimicking a time inconsistent gives much higher consumption in the first period, the incentive problem becomes tighter, and by lowering the tax rate the principal increases the cost of mimicking. The informational rent assumes again the form of a higher after-tax return of saving. On the other hand, at period 3, the mimicker consumes less than the truth teller. The greater is this difference, for a given tax rate, the softer the incentive problem is. An exponential has less incentive to mimic if the drop in consumption in the second period is high enough. If it is the case, than the planner can increase the tax rate. Furthermore, notice that the differences in consumption levels are weighted by α_t^{ih}, the marginal utility of income when mimicking. Moreover, the degree of time inconsistency, measured by the difference $\mu - \lambda^i$, also decreases the optimal tax (A-73). The intuition is simple: the higher is the gain by mimicking, the higher the rent given to time consistent should be.

To sum up, our conclusions are robust to a change in the kind of bounded rationality that characterize individuals' preferences. Two opposite effects determine the optimal tax rate on savings: from one side, if mimicking an hyperbolic agent assures more consumption now (because β is very low), the incentive problem becomes stronger and the principal softens it by increasing the cost of mimicking through a lower tax tate. The informational rent that has to be provided to the high types assumes the form of a higher after-tax return of saving. From the other side, in the following periods, the mimicker consumes less than the truth teller. As the difference in consumption levels increases, for a given τ_K, the incentive problem become softer; this is intuitive, as exponential have less incentive to mimic myopic if this behavior leads to a consistent drop in consumption levels in period two. In this case, *ceteris paribus*, the planner has the possibility to rise the tax rate, since the incentive problem is less strong.