CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 206

I Will Survive: Capital Taxation, Voter Turnout and Time Inconsistency

Matteo Bassi

September 2008

University of Naples Federico II

University of Salerno

Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance – UNIVERSITY OF SALERNO
84084 FISCIANO (SA) - ITALY
Tel. +39 089 96 3167/3168 - Fax +39 089 96 3167 – e-mail: csef@unisa.it

WORKING PAPER NO. 206

I Will Survive: Capital Taxation, Voter Turnout and Time Inconsistency

Matteo Bassi[♥]

Abstract
This paper reconsiders the debate around the political determination of capital income taxes and explains why such taxes survive in most OECD countries. The political economy literature on redistributive politics (Persson and Tabellini 2003) emphasizes the role played by the lower class in the political arena: being labor more concentrated than capital, the majority of the population benefits by overtaxing capital and undertaxing labour. However, in reality, political participation (voting, lobbying, protesting etc.) is positively correlated with income. Therefore, a paradoxical result emerges: why do the upper class, who is politically more active and own most of the capital, still favour a positive capital tax? Hence, voters' income is not the sole relevant variable in the political determination of the capital tax. To reconcile this apparent puzzle, we propose a model that incorporates time inconsistency *à la* Laibson in individual preferences We show that time inconsistent individuals are politically more homogeneous (or "single-minded") than far-sighted, and prefer to tax more capital income, instead of labor income, since accumulated saving are below the planned (and optimal) level and the distortionary effects of a higher capital tax are not only reduced but also delayed in time. We demonstrate that, since politicians find easier to please hyperbolic voters by proposing a tax policy that includes lower labor and higher capital taxes compared to an economy with only far sighted. Moreover, we show that, as the proportion of time inconsistent individuals in the population increases, the tax policy becomes more and more biased towards capital taxation.

JEL classification: A12, D72, H21, H24, H31

Keywords: Political Economy, Multidimensional Voting, Capital Taxation, Redistribution, Hyperbolic Discounting.

Acknowledgements: I wish to thank Helmuth Cremer, Georges Casamatta for helpful comments.

[♥] Università di Salerno, CSEF and Toulouse School of Economics (GREMAQ).
Address: CSEF, Dipartimento di Economia, Università di Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy. E-mail: matteo.bassi@gmail.com

Table of contents

1 Introduction

Capital income taxes continue to represent a major source of fiscal revenues in most OECD countries: more than 20% of total tax proceeds (OECD, 2007) have reference to various form of capital taxation (corporate income tax, taxes on capital gains etc.)[1].

A common view in the literature (see Auerbach, 2006, for instance) is that the importance of capital income taxes has decreased over time in most OECD countries: (except for France and Italy[2]) marginal tax rate on capital have declined over the period 1973-2004 and have converged towards the same level. This trend has recently stopped: corporate taxes have actually risen as a share of total revenue over the last years (especially in the U.S. and Canada), and still account for 10-25% of total tax revenues. As stressed by Sorensen (2007) and Devereux et al. (2002), the decrease of the corporate income tax rate has been more than compensated by the enlargement of the tax base[3], making the trend in marginal effective tax rates less evident. If follows that, overall, corporate tax revenues have actually increased in most OECD countries[4]. Moreover, if other forms of capital taxation are considered, it is evident that the fiscal burden on capital remain significantly high in the world's leading economies.

Are positive levels of capital taxes justified from an economic standpoint? The normative literature has not achieved a unanimous consensus upon the optimal level of capital taxation as illustrated by the following example presented by Martin Feldstein in a post published on *marginalrevolution.com*.

"Mr. X earns an additional $1,000. If X's marginal tax rate is 35%, he gets to keep $650. X saves $100 of this and spends the rest. If Mr. X invests these saving, he receives a return of 6% before tax and 3.9% after tax. With inflation of 2%, the 3.9% after-tax return is reduced to a real after-tax return of only 1.9%. If Mr. X is now 40 years old, this 1.9% real rate of return implies that the $100 of saving will be worth $193 in today's prices when he is 75. So his reward for the extra work is $550 of extra consumption now and $193 of extra consumption at age 75. But if the tax rate on the income from saving is reduced to 15%, the 6% interest rate would yield 5.1% after tax and 3.1% after both tax and inflation. And with a 3.1% real return, X's $100 of extra saving would grow to $291 in today's prices instead of just $193" (Martin Feldstein, *www.marginalrevolution.com*).

This example illustrates two characteristics of capital taxes.

First, taxes influence welfare and GDP: they may waste potential output, reduce welfare by decreasing the reward for saving and distort the allocation between saving and future consumption. Moreover, by

[1]Capital taxation may take several forms: taxes on interests, dividends, capital gains, business profits, and on the value of the housing services enjoyed by owners. In this work, we will refer indistinctly as "taxes on capital income".

[2]The center-left coalition proposed in his electoral program an increase of capital income tax rate from 12.5% to 20%. So far, however, such reform remains unapproved.

[3]For instance, governments have eliminated special deductions and generous asset depreciation rules. This strategy (the *tax-cut-cum-base-broadening* philosophy, very popular in the 80s and 90s) was encouraged by the practice of profit shifting and improvement in the ability of avoid taxation by corporations.

[4]In the U.S., for example, corporate taxes accounted for a higher share of federal revenues in 2005 than in any year since 1979 (Auerbach, 2006, based on OECD data).

increasing the cost of capital, taxes affect the quantity of investments made by firms, through effects on the relative returns to risk-taking.

Secondly, if lowering capital taxes would be beneficial for both taxpayers *and* the government, why does Mr. X, who is supposed to be rational, vote for parties that propose fiscal platforms distorted towards capital taxation? In a political economy voting model with office-seeking candidates, the equilibrium tax policy platforms that please the majority of voters entails low (possibly zero) taxes on capital income. Many papers try to justify, from a political point of view, why capital taxes account for a large share of total tax proceeds. A "redistributive" explanation is generally invoked: being capital more concentrated than labor, the majority should gain from shifting a larger share of the tax burden to capital. If the income distribution is skewed to the left, this idea presuppose that poor majority is more powerful and better organized than rich in the political process, and are able to impose their preferences to the losing minority. However, we know from the political science literature that rich individuals are more active in the voting process than poor[5], and that are less interested in redistribution. Therefore, the "redistributive " explanation is not robust to the reality and the question: "Why does capital income taxation still survive?" remains unanswered.

This paper justifies this apparent puzzle by considering a model that mixes economic, political and behavioral considerations. We propose a multidimensional voting model with opportunistic parties and voters that differ along two dimensions: *productivity level* and *time inconsistency*. Introducing a second source of heterogeneity allows us to depart from the idea that agents/voters display perfect rationality.

This assumption places our paper into the Economics and Psychology literature (see Laibson, 1997 for a review) that emphasizes how individuals' behavior can be better described by a model with bounded rationality. In particular, we assume that individuals, especially whenever it exists a temporal gap between the costs and the benefits associated with a given action, may be more impatient in the short run than in the long run, thus displaying time inconsistency.

Formally, to capture this idea, each individual is modeled as a collection of selves: hyperbolic discounting leads present selves to overweight current payoffs compared to future ones, giving rise to a conflict between preferences of different intertemporal selves. Moreover, not only a time inconsistent individual makes plans that, in absence of any suitable commitment devices, will he will systematically change, but also regrets, ex-post, of his lack of commitment.

The following intertemporal utility function (Strotz 1956, Phelps and Pollacks 1968, Laibson 1997) describes this possibility:

$$u_0(.) + \beta \sum_{t=1}^{T} \delta^t u_t(.) \tag{1}$$

[5]Rich contribute more in political campaigns, have a higher turnout and have more resources to devote to lobbying activities.

where β represents the short-term psychological discount factor, and δ is the long term one. This formulation implies that the discount function is 1 at $t = 0$ and to $\beta\delta^t$ for $t = 1, 2, ..., T$. It follows that implied discount factor between today and the next period is $\beta\delta$, whereas that between any two subsequent periods in the future is δ: the discount factor is first declining, and constant thereafter[6].

Together with our behavioral assumption, the model takes into account several aspects of the real life politics: in particular, we consider that political participation is increasing with income and some individuals are excluded from the political game. By taking into account real turnouts in political elections, we show that it is hard to justify the idea of poor being able to impose their preferred capital taxes to the rich minority.

Anticipating the results, we show that, when voting over the optimal tax mix that finances a redistributive transfer, poor and time inconsistent agents, for any income level, are "single minded", and both agree to lower labor income tax and to increase capital taxation. The intuition for the result is the following: the lower class, owning less capital, favors naturally high capital taxes. However, since this group participates less in the political process, needs to form a coalition with time inconsistent voters, who share, *for any income level*, the same preferences on the optimal allocation of the tax burden between capital and income taxation. Hyperbolic individuals prefer higher capital taxes for two reasons: first, increasing the after-tax return from savings has only a negligible effect on hyperbolic propensity to save: because of their preferences, they still prefer to consume "too much " when young instead of saving, despite the higher return. Second, labor supply is chosen period-by-period, and thus is unaffected by time inconsistency; increasing labor taxes today (together with a lower capital tax tomorrow) implies a first-order reduction in hyperbolic current utility and only a second-order increase in their future utility. Given individual preferences, opportunistic parties maximize the probability of being elected by proposing a fiscal burden distorted towards capital taxation, as to exploit the single mindedness of hyperbolic and poor voters.

The paper proceeds as follows: in section 2, we present stylized facts about capital taxation, as to show that they account for a substantial part of tax revenues in most OECD countries. In section 3 we review the economic literature on capital taxation, both from a normative and a positive point of view. Section 4 presents stylized facts about political participation. Section 5 presents our basic model, which is solved for individuals (section 6) and for the two parties (section 7). Section 8 concludes.

[6]The empirical relevance of this behavioral assumption has been tested (Ainsle 1992) through experiments, simulations and real data. In particular, Laibson, Repetto and Tobacman (1998 and 2004), using data on credit card borrowing and consumption-income comovement, test whether individuals actually behave patiently in the long term and impatiently in the short term. They find hat the hypothesis that the short term discount factor β coincides with the long run one, δ, should be reject. Moreover, the estimated values for the β and δ are, respectively, around 40% and 4%, thus confirming that the hyperbolic model better explains individuals' decision making.

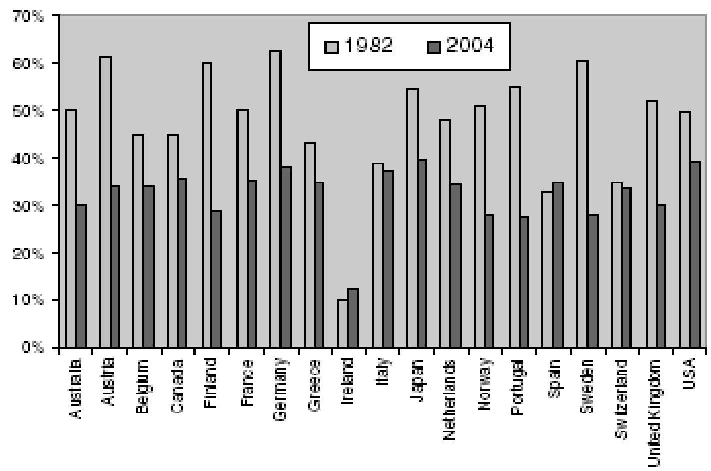

Figure 1: Corporate Tax Rates (*Source*: Sorensen, 2007)

2 Stylized Facts about Capital Taxation

According to Carey and Rabesona (2004), the average level of capital income taxation was around 50% of income in 2002. Data in Persson and Tabellini (2003) show that, in a sample of 14 OECD countries, the average effective tax rates on capital and labor were about the same (around 38%) over the period 1991-1995. In the same period, in U.S. and U.K. capital taxes were higher than labor taxes.

Capital taxes concern both corporations and individuals. For the former, taxes on corporate income (the most important form of capital tax) have fallen in the period 1980-2004 (Figure 1), but the proceeds of this tax have substantially increased, except in Japan, Germany and UK (Figure 2). Since profit shares in the GDP have remained almost the same (Sorensen 2007), this increase in revenues was mainly due to the enlargement of the tax base. Therefore, effective corporate taxation has increased.

It is hard to present evidence for personal capital taxes: the difficulty comes from the fact that OECD statistics do not decompose total revenue from personal income taxes into tax falling on capital income and tax levied on labour income.

Sorensen (2007) estimates the tax structure and the allocation among different sources (capital, labor and property): from figure 3, we see that personal taxes on capital income contribute between 5 and 10 percent of total tax revenue in OECD most countries. On the other hand, the table shows that corporate taxation is a more significant revenue raiser than the personal capital income tax. The importance of property taxes, a mix that includes taxes on the ownership and transfer of real and financial assets, varies quite a lot across countries.

The rest of the section focuses on the structure of capital taxation in the United States, where more data are available, and the puzzle between capital taxes and voting behavior is more evident. In the U.S.

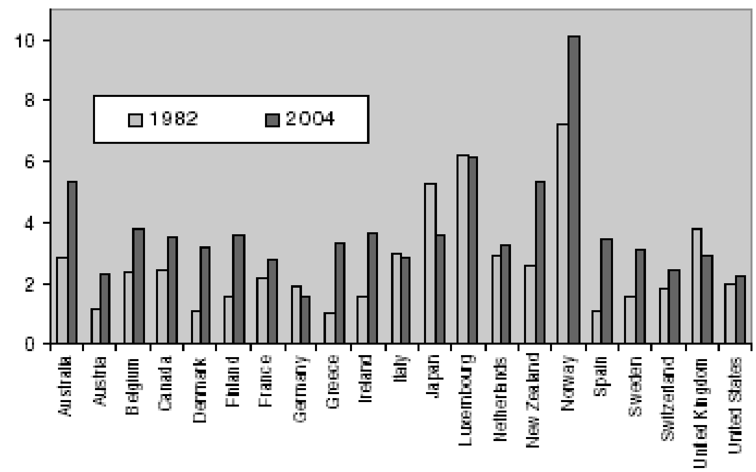

Figure 2: Corporate Tax Revenues (*Source*: Sorensen, 2007)

individuals and corporations pay capital income tax on the net total of all their capital proceeds just as they do on other sorts of income.

Back to 1963, the highest marginal rate of personal (capital and labor) income tax was 93 percent. This rate was reduced, but even as recently as 1980, the top income tax rate was 70 percent and interest and dividend, and the corporate tax rate was around 46 per cent. Moreover, capital gains tax rates were significantly increased in the 1969 and 1976 *Tax Reform Acts*: the minimum tax rate for such gains was increased up to 15 percent, whereas the maximum rate reached 40 percent (Auten 1999). In 1978, Congress reduced capital gains tax rates by eliminating the minimum tax on excluded gains and increasing the exclusion to 60 percent, thereby reducing the maximum rate to 28 percent. The 1981 tax rate reductions further reduced capital gains rates to a maximum of 20 percent. The *Tax Reform Act* signed by president Reagan in 1986 changed substantially the tax code: corporate tax rate was reduced to 35 percent (although, as we have seen the tax base was broaden), but the exclusion of long-term gains was repealed, and the maximum tax rate for short term capital gains was raised to 28 percent (33 percent for taxpayers subject to phaseouts). As an example, Figure 4 illustrates the evolution of nominal and effective tax rates for capital gains for the period 1984-1995: effective tax rates increased during this period.

Until 2003, no substiantial reforms in the tax treatment of capital gains were adopted. In 2003, the tax rate for individuals was lowered for *long-term* capital gains, *i.e.* gains on assets held for over one year before being sold, and increased for *short-term* capital gains. For the former, the tax rate was reduced to 15% (or to 5% for individuals in the lowest income tax brackets). On the other hand, the latter are taxed at the (higher) ordinary income tax rate. This reduced tax rate was scheduled to expire in 2008 but the

Country	Taxes on capital			Other taxes				
	Personal capital income taxes[a]	Taxes on corporations	Taxes on property	Personal labour income taxes[b]	Social security taxes	Payroll taxes	Taxes on goods and services	Other taxes
United States	9.4	8.7	12.0	25.3	26.3	–	18.3	–
Japan	2.5	14.2	10.0	15.3	37.7	–	20.0	0.3
Germany	8.0	4.5	2.5	14.8	40.7	–	29.2	0.0
France	5.8	6.3	7.6	11.2	37.1	2.6	25.6	3.6
Italy	10.2	6.9	6.1	15.2	30.3	–	26.4	5.5
United Kingdom	6.6	8.1	12.0	22.1	18.8	–	32.0	–
Canada	4.2	11.4	10.2	30.9	15.2	2.0	25.9	0.2
Australia	7.2	18.2	8.7	33.0	–	4.4	28.5	–
Austria	5.8	5.4	1.3	18.2	33.9	6.1	28.2	0.9
Belgium	9.5	8.0	3.9	21.1	31.3	–	25	0.0
Czech Republic	3.8	12.4	1.1	8.9	42.3	–	31.2	0.0
Denmark	7.1	9.7	3.8	43.6	2.4	0.4	32.7	0.0
Finland	6.1	8.1	2.6	24.4	26.8	–	31.7	0.1
Greece	6.8	9.4	4.4	7.0	34.7	–	37.1	–
Ireland	4.4	11.9	6.9	23.0	15.0	0.6	37.8	–
Korea	1.0	14.3	11.3	12.6	20.7	0.2	36.3	3.5
Netherlands	5.4	8.2	5.3	11.0	36.9	–	32.0	0.4
New Zealand	9.8	17.2	5.0	34.2	-	–	33.8	–
Norway	4.7	22.6	2.6	18.8	21.6	–	29.7	–
Portugal[c]	1.9	8.3	4.6	14.0	31.8	–	38.6	0.5
Spain	3.2	9.8	8.1	14.5	34.8	–	28.0	0.4
Sweden	5.3	6.3	3.1	26.1	28.4	4.7	25.8	0.1
Switzerland	3.1	8.6	8.5	31.7	24.4	–	23.7	–

Figure 3: Tax Structures in OECD countries in 2004 (percent of total tax revenues; *Source*: Sorensen, 2007)

Economic Growth and Tax Relief Reconciliation Act, signed by President Bush in 2006, has extended this reduced tax rate through 2010. After that date, taxes will revert to the rates in effect before 2003, which were generally 25%[7]. Concerning corporate taxes, the effective average tax rate in the U.S. is around 40%.

This section has presented stylized facts about capital taxation: data for most OECD countries show that *effective* capital taxes, and in particular corporate taxes, remain high. The same appears to be true for the taxation of long term capital gains, since recently implemented tax cuts are temporary and have a clear electoral motive.

The following sections will verify whether the literature on optimal capital income taxation is in line with these empirical observations.

3 Literature Review on Capital Taxation

3.1 Normative Theories

What is the optimal level of the capital income tax? Is replacing capital taxes with other forms of taxation welfare-increasing? Normative public economic literature has tried to answer these questions, but so far unanimity among economists has not been reached. In this section we try to summarize the main findings about this topic[8].

[7]Given that the empirical evidence show that voters of the Republican party are in average richer than Democrats (Krugman 2007, Bartels 2007), this reform could appear, at first sight, harmful for Republican voters.

[8]The background for this section is given by Auerbach and Hines (1998), Barnheim (1999) and Sorensen (2007).

Year	Total Positive Realized Capital Gains ($ billion)	Taxes Paid on Capital Gains ($ billion)	Effective Tax Rate on Capital Gains (%)	Gains as Percent of Gross Domestic Product (%)	Maximum Tax Rate on Long-Term Gains (%)
1984	140,5	21,5	15,3	3,7	20
1985	172	26,5	15,4	4,3	20
1986	327,7	52,9	16,1	7,7	20
1987	148,4	33,7	22,7	3,3	28
1988	162,6	38,9	23,9	3,3	28
1989	154	35,3	22,9	2,9	28
1990	123,8	27,8	22,5	2,2	28
1991	111,6	24,9	22,3	2	28,9
1992	126,7	29	22,9	2,1	28,9
1993	152,3	36,1	23,7	2,4	29,2
1994	152,7	36,2	23,7	2,2	29,2
1995	180,1	44,3	24,6	2,5	29,2

Figure 4: Capital Taxation in the U.S. (1984-1995)

There is a presumption among economists that capital taxes raise revenues in a less efficient way than wage or consumption taxes. Many authors show[9] that capital taxes are desirable only in the short-run: after some initial transition in which savings are discouraged, the long-run capital tax has to converge to zero. The intuition behind this result is related to the classical Ramsey (1927) model; by interpreting consumption at different dates as different commodities, and the capital tax as a selective commodity tax on future consumption, the uniform taxation result applies: capital income should be taxed in the initial period, where the relative price distortion caused by capital income taxation is finite, but never in the following periods, since the size of the distortion increases. This result continues to hold if we assume that individuals have to make a labor supply decision, provided that their utility function is separable between labor and consumption (Atkinson and Stiglitz 1976)[10].

The Chamley-Judd result, however, relies on simplifying assumptions: preferences should be intertemporally separable and isoelastic; capital markets have to be perfectly competitive and complete (individuals may freely reallocate consumption over time by borrowing and lending); there is no uncertainty over the labor income; the time horizon of the representative individual coincides with the one of the planner[11]. Removing these assumptions, positive[12] capital taxes may become optimal. If borrowing

[9]See, for instance, Diamond (1973), Auerbach (1978), Atkinson and Sandmo (1980), Judd (1985 and 1999), Chamley (1986), Chari, Christiano and Kehoe (1994).

[10]The Atkinson-Stiglitz theorem is a particular case of the Corlett-Hague (1953) rule: a commodity tax system that minimize the deadweight loss should impose higher taxes on commondities that are more complementary to leisure, since this will minimize the tax induced subsitution towards leisure. Therefore, if future consumption is more complementary to leisure than present consumption, the former should be reduced through a tax on savings. Since ther is no evidence whether future consumption is more or less substitutable for leisure than present consumption, most economist assume the same degree of substitutability, and therefore a zero optimal capital income tax.

[11]In a recent paper, Abel (2007) challenges the Chamley-Judd result without a substantial depart from the basic model: in an economy with identical infinitely-lived households, if the purchasers of capital are allowed to deduct capital expenditures from the capital income tax base, then a constant and positive tax rate on capital income is non-distortionary. The tax system that implements the optimal allocation consists of a positive tax rate on capital income and a zero tax rate on labor income, the opposite result found by Chamley and Judd.

[12]Or negative taxes (subsidies). See Judd (1997).

constraints, and/or imperfections in the labor and credit markets exist, than it may be optimal to levy a capital tax even if the horizon is infinite (Aiyagari, 1993 and Chamley, 2001). If labor income is subject to stochastic shocks, in absence of market-provided insurances, a capital tax plays the role of a publicly provided insurance device against productivity shocks, and its proceeds may be used to make transfers from high consumption to low consumption states, in order to insure individuals against low-consumption states,Along this line of research, the New Dynamic Public Finance literature (see Kocherlakota 2006, for a review) has recently reconsidered the determination of the optimal tax burden in a dynamic framework, with credit markets imperfections and random shocks on individuals' productivity. The main result that emerges is that the optimal wedge between marginal rate of substitution and marginal rate of transformation is different from zero, *i.e.* saving should be discouraged. The optimal intertemporal allocation can be implemented using a tax system that is linear in current wealth, but equal to zero in expected and aggregate terms.

In settings where consumers' time horizon is shorter than the planner's one (as in OLG models *à la* Diamond), or when future consumption is more complementary to leisure than present consumption (Erosa and Gervais, 2002), a positive capital tax may be optimal.

Redistributive concerns also provide a rational for positive capital taxes: Krusell et al. (2000) and Salanié (2003), by extending the Atkinson-Stiglitz model to a dynamic framework, show that a positive capital income tax is indeed optimal. To be more precise, they assume that saving for future consumption induce capital accumulation and influence pre-tax factor incomes. If skilled labor is more complementary to capital than unskilled labor, it follows that the proceeds of a capital tax that discourages saving can be used to redistribute income in favour of low-income earners, given that the distortion induced by this tax is more than compensated by the welfare gain of a more equitable distribution of income [13]. Finally, a linear tax on capital income represent an optimal instrument to finance a redistributive transfer when the tax authority is not able to observe and to tax directly inherited individual wealth (Cremer et al., 2003, and Boadway et al. 2000).

Even if taxing capital would be optimal, it is also possible that such form of taxation originates substantial welfare losses that can removed by replacing them with labor or consumption taxes. In this sense, Feldstein (1978) shows that replacing capital with labor taxes yielding the same revenues increases welfare by approximately 18%. This conclusion continue to hold in a general equilibrium framework: Chamley (1981) and Judd (1987), in models with infinite-lived individuals, show that the deadweight loss of taxing capital is high (around 11% of total revenue, when the capital tax rate is 30%. Welfare losses are substantial also if in a OLG framework: simulations in Diamond (1970) and Summers (1981) show that steady state welfare would increase by 12% if capital taxation were replaced with consumption

[13] If this complementarity is not taken into account, capital accumulation does not affect the pre-tax distribution of wages, and thus a zero capital income tax is still optimal, provided that utility is separable in consumption and leisure (Ordover and Phelps, 1979)

taxes, and by 5% if were replaced by a labor income tax[14]. Auerbach, Kotlikoff and Skinner (1983) improve upon Summers' analysis, comparing not only steady states welfare levels, but also changes in welfare along the transition path, and confirm that replacing capital with consumption taxes would increase steady state welfare by 6%. However, if the capital income tax is replaced by a wage tax, steady state welfare would *decline* by 4%.

This section shows that, from an efficiency standpoint, capital taxation is in general not desirable, provided that some simplifying assumption are satisfied. However, once redistributive concerns are taken into account, the optimal capital tax may be positive. Simulations show that a reform replacing the capital income tax with other forms of taxation (on wages or consumption) would be welfare-improving.

3.2 Positive Theories

This section reviews the political economy literature on capital taxation; the objective is to understand how the level of capital taxation is determined in the political arena. Several papers have tried to justify the existence of positive capital taxes: we classify these explanations into four groups.

First, capital taxes may exist because for a government it represents an efficient way to collect revenues. Politicians may refrain from eliminating capital taxation if increasing the after tax return of saving does not boost capital accumulation, but only decreases total revenues. The relevance of this explanation depends on the sign and the magnitude of the *interest elasticity of saving*, that measure the responsiveness of saving accumulation to a change in their after-tax return. From a theoretical standpoint, this elasticity can be either positive or negative (Bernheim, 1999), and saving can rise or fall in response to a decrease of the tax rate. If individual preferences are represented by a CES utility function, the sign of this elasticity depend on the sign of the intertemporal elasticity of substitution in consumption: saving rises (*resp.* falls) in response to cut in the tax rate if the elasticity of substitution is high (*resp.* small). Unfortunately, the empirical literature[15] is not able to provide a direct estimate for the value of this elasticity. To overcome these difficulties, a different (indirect) approach has been adopted: in particular, scholars have tried to compute how the introduction of tax-deferred savings account[16] (IRA and 401(k), for instance) has modified the choice of optimal saving. The question is to understand how much less would contributors have saved in absence of these accounts. Unfortunately, the answer is still undetermined: IRAs were effective in attracting new contribution, but it is not clear whether these savings are "new" or simple displacements from other forms of savings (Bernheim, 1999). It is clear that the mixed evidence about

[14] Summers' analisys suffers from several drawbacks: first, he considers only the steady state and not the transition path following the tax reform, and thus he negliges the negative distributional effects that reduce transitional generations' welfare. Second, labor supply is inelastic, and thus the optimal tax rate is zero by assumption.

[15] See Bernheim (1999), Hubbard and Skinner (1996), and Poterba, Venti and Wise (1996).

[16] IRAs and 401(k) were introduced by the U.S. goverment in the 70s, to boost individual saving: these accounts feature tax deductible contributions up to a certain limit, tax-free accumulation, taxation of principal and interest on withdrawal, and penalties for early withdrawal. After an initial popularity (20 billions \$ in the 1986), contributions fell to less than 10 billions \$).

the sign of elasticity of substitution does not allow us to conclude whether a lower tax rate on capital increase/decrease/keep constant savings and therefore we can not infer that individuals and politicians prefer to tax capital as to minimize distortions[17].

A second political justification for capital taxation is related to the lack of credibility of politicians, or the *capital levy problem* (Fischer, 1980): announcing a reduction in capital taxes would not be credible for the politician, since the elasticity of saving already accumulated is zero. In equilibrium, capital will be highly taxed, more than would be efficient for the representative agent. Such a strategy, however, does not work in a repeated model, where politicians care not only about winning the current election, but also maintaining their reputation: announcing low capital taxes before elections and taxing capital later will destroy politicians' credibility for the future.

A third explanation refers to the *strategic political delegation*: rational voters, anticipating that, after the elections, the policy-maker will face a different set of incentive constraints, prefer to elect someone with different preferences from their own. Agents overcome the capital levy problem by using another government at their advantage. This explanation is not entirely satisfactory, since it is know that most of voters have an ideological bias towards a political party, and quite rarely are willing to modify their vote to tie the government's hands.

The forth political explanation for positive levels of capital taxation relies on redistributive concerns, which may also justify capital taxes from a normative standpoint. This view is proposed by Persson and Tabellini (2003): when voting over the composition of the tax burden, the lower class has more political power than the upper class, given that labour income is less concentrated than capital income, and poor represent generally the majority of the population. Therefore, the winning majority is composed by poor individuals that benefit from more redistribution, and the policy vector entails overtaxation of capital income and undertaxation of labour income. However, this model has little empirical support: in real life elections rich are indeed the more involved in the political process and, *ex-ante*, are not interested in redistribution. Moreover, since they own more capital, the resulting positive level of capital taxation is puzzling.

None of the political theories reviewed is fully able to explain, in our opinion, the level of capital taxes observed in reality; our paper, by considering different assumptions about individuals' rationality and some facts about how elections work, will help to understand this puzzle.

[17]Feldstein (2006 and 2007) shows that, even if the interest elasticity of substitution were effectively zero, the negative effects of capital taxation on saving would remain: a tax not only affects *current* consumption, but also *future* consumption that could be actually bought by saving. Feldstein provides the following example: assume that in absence of capital taxes, the return of savings is 10%. If the capital tax is 50%, the net return is only 5%. For an individual who saves at 45 years old and dissaves at 75, each dollar saved increased future consumption to 17$ whereas, with the tax, one dollar today will buy only 4,3$, with a decline of 75%, for a *given* level of saving.

4 The Political Science Literature

The political economy literature has not yet considered three important stylized facts known by political science scholars. Incorporating real world facts into economics would help us to better understand how politicians take decisions and why certain policies are implemented.

First, *not all individuals are politically active*[18]: turnouts (defined as percent of the voting population, *i.e.* everyone above the minimal age for voting, usually 18 years), are much lower than 100%: the average is around 77% in European countries, around 50% in the United States and 54% on average in Latin America countries. While turnout across the globe rose steadily between 1945 and 1980 (increasing from 61% in the 1940s to 68% in the 1980s), since then it has dipped back to 64%, despite the increase in educational levels and economic well-being[19] (Comparative Study of Electoral Systems, 2007). Several reasons justify this tendency[20]: first, burdensome registration procedures may represent a major institutional deterrent to voting. This happens in the U.S. (Rosenstone 1993), but less for Europe, where voting procedures are less complicated. However, also Europe has experienced dramatic declines in voter turnouts (Topf, 1995). Second, also the salience of the issues plays a role in determining voters' participation: political elections have higher turnouts than administrative and local elections, perceived to be less important. Third, turnout is influenced by the attractiveness of parties and candidates: many countries have recently experienced a growing disbelief towards politics and a lower interest for political activity. Fourth, institutional design affects turnout: the choice of the electoral system affects on voters' participation according (Lijphart, 1994): Proportional Representation increases voting participation, by giving citizens more choices and by eliminating *wasted votes* (votes cast for losing candidates or for candidates that win with big majorities), which is typical of systems that use Single-Member districts. The frequency of elections also negatively influenced turnout (Boyd, 1989) by increasing the cost of voting.

Whichever the reasons for low turnouts are, this fact would not represent a issue if non-participation was randomly and evenly distributed among social classes: however, participations is highly unequal, and it is systematically biased in favor of those with higher incomes, greater wealth and better education, against less advantaged citizens (Lijphart 1997). This leads us to the second stylized fact: *political participation increases with income*[21]. A common idea is that self-interest is the main motivation for

[18] In our terminology, the "political process" includes not only voting, but also broader form of political participation, both conventional (working in election campaigns, contribution ot parties or candidates, working informally in the community, lobbying) and unconventional (participation in demonstrastions, boycotts, rents and tax strikes, occupations).

[19] Countries with low literacy rates do not necessarily have a lower turnout: there is no significant statistical correlation between education level and voter turnout, although highly literate countries, on average, have a higher level of political participation. Nevertheless, high illiteracy countries such as Angola and Ethiopia have achieved high turnout rates.

[20] Following the "voting paradox " theory,the striking result that has to be explained is not why 50% of citizens do not vote, but why there is still a 50% of them who continues to do it, since their vote is far from being decisive and not voting is seen as a completely rational activity. However, we take as given the fact the voter's turnout is low, and we do not anayze the determinants of voting.

[21] At the beninning of 20th century, with the adoption of universal suffrage in many countries, political analysts were convinced that the intellectual élite would have preferred not to vote, since its vote would drown among the votes of the

voting: those who have a higher stake in the political process should be the more active. It follows that poor individuals, who in principle benefit more from public policies and redistributive transfers, should be more involved in the political process. However, there is an old and vast empirical evidence that does not confirm this myth[22]: Gosnell (1927) finds that turnout increases with economic status and that "the more schooling the individual has the more likely he is to register and vote in elections". The same pattern is reported also in Arneson (1925) and Tingsten (1937), who reviewed elections' results in Switzerland, Germany, Denmark, Austria, U.S. and Sweden and formulated the rule that "voting frequency rises with rising social standard". This bias is particularly strong in the U.S., where "no matter which form citizen participation takes, the pattern of class equality is unbroken (Lijpart 1997)", and where, over time, the level of voting participation and class inequality are strongly and negatively linked. A study by Comparative Study of Electoral Systems (2007), shows that, for OECD countries, those who voted in the current election have a higher average income than those who did not. An exception to this trend is the participation of senior citizens specifically with regard to Social Security (Campbell, 2002): in this case, participation *decreases* as income rises, in part because lower-income citizens are more dependent on the program.

The positive correlation between income and participation leads us to the third fact: *politicians tend to favor the opinion of rich.* Given that the upper class participates more actively in the political debate, it is not surprising that "*inequalities in political participation are likely to be associated with inequalities in governmental responsiveness* " (Verba, Schlozman, and Brady, 1995),

Bartels (2005) provides some evidence that support this intuition. His paper investigates how responsive U.S. senators are to the preferences of rich, middle-class, and poor constituents; senators appear to be considerably more responsive to the opinions of affluent constituents than to the opinions of the middle-class, while the opinions of poor have no apparent statistical effect on their senators' roll call votes. The sign of the bias is the same both for Democrats and Republican senators; however, the latter appear to be more than twice as responsive as the former to the ideological views of rich constituents.

Forth, *there is a difference in voter turnout between young and old voters*: old's participation rates are higher than young individuals with the same characteristics (income, wealth level, education etc.). For instance, data from the U.S. National Election Study show that citizens aged more than 6 were 7% more likely to vote than their young counterpart.

Another fact, although less clear and more controversial in the political literature, is the relationship between income level and the ideological view of the voter: a persistent myth is that rich people vote Democratic, while workers vote Republican[23]. However, according to data in Krugman (2007) and Bartels

mass. Quite soon, empirical studies showed that status and voting were positively, and not negatively, correlated.

[22] " [...] *Low voter turnout means unequal and socioeconomically biased turnout. This pattern is so clear, strong and well known in the U.S. that it does not need to be elabored further"*. (Lijphart, 1997)

[23] According to MSNBC's political journalist Tucker Carlson: "*Here's the fact that nobody ever, ever mentions —*

(2006), the truth is just the opposite[24]. According to 2006 exit polls, among individuals with less than \$100,000 (78% of the voting population), 55% voted for Democratic Party, and 43% for Republicans. For, individuals with more than \$100,000, 47% voted Democrats and 52% Republicans. A 4-point difference between top and bottom became a 14-point difference.

This analysis shows that poor are more or less excluded from the political arena. They have lower turnouts and are less involved in other political activities (lobbying, campaign financing etc.). It is not surprising that office-seeking parties try to please the more involved in the political life, as Bartels (2006) has stressed. But this is in contrast with the evidence presented in previous sections: if the active electorate is composed mostly by wealthy individuals who own most of the capital in the economy, and political parties are sensitive to rich's preferences, then why is tax burden distorted towards capital taxation? Interestingly, capital taxes appear to be higher than labor taxes in the U.S. than in Europe, although the positive relationship between income and participation is stronger in the U.S. Does it mean that U.S. citizens vote against their interests? Is there really a *Myth of the Rational Voter* (Caplan, 2007) and individuals approve bad policies just because they are misinformed by politicians and unable to fully understand the economic implications of political actions?

We believe that voters are rational, but their behavior can be better described with a model with bounded rationality and, in particular, by *quasi hyperbolic discounting*. The puzzling result about capital taxation can be perfectly understood through a political economy model that embeds more realistic assumptions about individuals preferences: some individuals display a higher preference for present utility, whereas others do not, and these preferences not only matter for economic choices but also for political decisions.

5 The Economic Environment

We consider a three-periods OLG model; in every period, three generations are alive: *old, middle aged* and*young*. Population grows at a constant rate G. The size of each generation is denoted, respectively, with and n^o, $n^{ma} = (1+G)\, n^o$ and $n^y = (1+G)^2\, n^o$.

When young and middle aged, individuals supply labor l and save for post retirement consumption, s: the endowment of units of time is normalized to one. When old, an individual is retired and consumes saving accumulated in previous periods and receive a transfer P, that represent an instrument of intergenerational and intragenerational redistribution, which is financed through the proceeds of two

Democrats win rich people. Over 100,000 in income, you are likely more than not to vote for Democrats. People never point that out. Rich people vote liberal. I don't know what that's all about".

[24] In a post published in his own weblog, economist Paul Krugman states: *"There's a weird myth among the commentary that rich people vote Democratic. There's another strange thing about that myth: the notion that income class doesn't matter for voting, or that it's perverse, has spread even as the actual relationship between income and voting has become much stronger. And the fact that people with higher incomes are more likely to vote Republican has been consistently true since 1972. The interesting question is why so many pundits know for a fact something that simply ain't so".*

proportional taxes[25]: on labor[26], τ_ω, and capital income, τ_K.

Utility of consumption is expressed by the increasing and concave utility function $u(.)$, while the disutility of effort is expressed by $v(.)$, with $v'(.) > 0$ and $v''(.) > 0$. Let r be the constant and exogenous gross return on wealth.

Within each generation, individuals differ with respect to two dimensions[27]: *productivity level* and the *degree of time inconsistency*.

For the former, we assume that each individual, at the beginning of his life, is assigned with a productivity ω, which remains the same in the next period: ω can take two values, ω^P (poor) and ω^R (rich), with the obvious ranking $\omega^R > \omega^P$. Each income group represents, respectively, a fraction ρ^P and ρ^R (or, alternatively, $1 - \rho^P$) of each group, with $\rho^P > \rho^R$. The mean wage of the economy is $\bar{\omega} = \rho^R\omega^R + \rho^P\omega^P$.

For the latter, we assume that certain individuals display a bias toward the present in intertemporal trade-offs and ex-post regret about their lack of commitment. More precisely, the psychological short-term discount factor β between two subsequent periods is lower for time inconsistent than for time consistent individuals: $\beta^{TI} < \beta^{TC}$. Furthermore, we assume that time inconsistent individuals are *sophisticated*, in the sense that they are aware of their self-control issues but, in absence of any commitment device[28], they are not able to stick to their optimal plans. On the other hand, time consistent (or *exponential*) individuals can implement optimal consumption paths. Time consistent and time inconsistent individuals represents, respectively, a fraction λ^{TC} and λ^{TI} of each income group[29]. Therefore, in each generation, we have four group of individuals: poor time consistent (a fraction $\rho^P\lambda^{TC}$ of the population), rich time consistent ($\rho^R\lambda^{TC}$), poor time inconsistent $\left(\rho^P\lambda^{TI}\right)$ and rich time inconsistent $\left(\rho^P\lambda^{TI}\right)$.

The behavioral assumption affects the consumption/saving choice; anticipating the results, we show that time inconsistent old experience a drop in post-retirement consumption, caused by overconsumption when young and middle aged. On the other hand, labor supply, being decided period by period, is not influenced by hyperbolic discounting.

Taking into account the two sources of heterogeneity and the three generations, twelve groups coexist

[25] P can represent either a pension transfer awarded only to retirees, or a public good that increase only old's consumption: health care, for instance, whose consumption increase with age.

[26] If P is interpreted as a pension benefit, then τ_ω is the payroll tax that finances the PAYG system.

[27] We assume that the two sources of heterogeneity are uncorrelated. The existence of a positive (or negative) correlation between income level and degree of time inconsistency is an open empirical question.

[28] Assuming, as we implicitly do, that markets are incomplete, *i.e.* commitment devices for hyperbolic are not available, may appear too strong. However, assuming the completeness of financial markets implies also that we should consider that, together with commitment devices, the market would propose "counter-commitment devices" that exploit the consumers' present bias. For instance in the U.S., the growth of IRA accounts, 401*(k)* plans has been followed by the boom of revolving credit cards. Moreover, as we show in the introduction, there is no sure evidence that the introduction of IRA accounts and 401*(k)* plans has effectively boosted individual savings.

[29] Clearly, $\rho^R + \rho^P = 1$ and $\lambda^{TC} + \lambda^{TI} = 1$. Moreover, we assume that the fractions of rich, poor, time consistent and time inconsistent individuals remain the same across periods. Finally, we do not impose any ranking between λ^{TC} and λ^{TI}.

in our economy (see Figure 5 for a graphical representation):

$$y^P_{TI}, y^P_{TC}, y^R_{TI}, y^R_{TC}, o^P_{TI}, o^P_{TC}, o^R_{TI}, o^R_{TC}, ma^P_{TI}, ma^P_{TC}, ma^R_{TI}, ma^R_{TC}$$

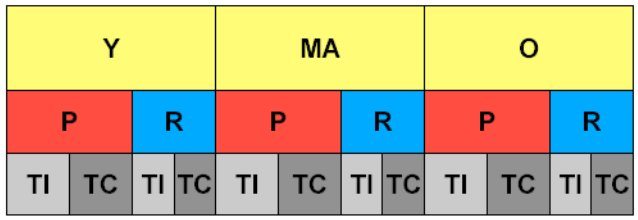

Figure 5: Behavioral and Economic Types

The utility function of an old individual of type $i = R, P$ and $j = TI, TC$ depends only on total consumption:

$$U(c_o) = u(c_o^{i,j}) \tag{2}$$

where:

$$c_o^{i,j} = (1 + r(1 - \tau_K)s_{ma}^{i,j}) + P$$

and $s_{ma}^{i,j}$ is the amount of saving accumulated in the previous period. The pro-capita transfer P is given by:

$$P = \frac{1}{n^o}\left\{\tau_\omega[\omega^R L^P + \omega^P L^R] + \tau_K r\left[\lambda^{TI} S_{TI} + \lambda^{TC} S_{TC}\right]\right\} \tag{3}$$

where $L^i = \rho^i\left(n^y l_y^i + n^{ma} l_{ma}^i\right)$ is the total labor supplied by young and middle aged belonging to the same income group. $S_{TI} = \sum_{i=R,P} \rho^i\left(n^y s_y^{TI,i} + n^{ma} s_{ma}^{TI,i}\right)$ and $S_{TC} = \sum_{i=R,P}\left(n^y s_y^{TC,i} + n^{ma} s_{ma}^{TC,i}\right)$ represent the total amount of saving for time inconsistent and exponential individuals.

The preferences of a *middle aged* depend on consumption, $c_{ma}^{i,j}$, and labor supply, l_{ma}^i:

$$U(c_{ma}, l_{ma}) = u\left(c_{ma}^{i,j}\right) - v(l_{ma}^i) + \beta^j \delta u\left(c_o^{i,j}\right) \tag{4}$$

with:

$$c_{ma}^{i,j} = \omega_i l_{ma}^i(1 - \tau_\omega) + (1 + r(1 - \tau_K))s_y^{i,j} - s_{ma}^{i,j}$$
$$c_o^{i,j} = (1 + r(1 - \tau_K)s_{ma}^{i,j}) + P$$

Finally, the intertemporal utility function for the representative *young* is:

$$U(c_y, l_y) = u\left(c_y^{i,j}\right) - v(l_y^i) + \beta^j \delta\left[u\left(c_{ma}^{i,j}\right) - v(l_{ma}^i) + \delta u\left(c_o^{i,j}\right)\right] \tag{5}$$

where the budget constraints are:

$$
\begin{aligned}
c_y^{i,j} &= \omega_i l_y^i (1-\tau_\omega) - s_y^{i,j} \\
c_{ma}^{i,j} &= \omega_i l_{ma}^i (1-\tau_\omega) + (1 + r(1-\tau_K))s_y^{i,j} - s_{ma}^{i,j} \\
c_o^{i,j} &= (1 + r(1-\tau_K))s_{ma}^{i,j} + P
\end{aligned}
$$

Utility functions (4) and (5) reflect the general intertemporal hyperbolic utility function given by (1): the discount structure implies that individuals, when young, discount the utility level of subsequent periods at the rate $\beta\delta$ (middle aged) and $\beta\delta^2$ (old) meaning that they are impatient when they make short run trade-offs. On the other hand, from the point of of view of a young individual, the discount factor between two periods far in the future (between middle age and old age) is simply δ, implying that the agent is patient in the long run. To simplify our computations and to obtain closed-form solutions, we assume that $u(.)$ and $v(.)$ take the following functional forms:

$$
u(c) = \frac{c^{1-\sigma}}{1-\sigma} \text{ and } v(l) = \frac{l^\gamma}{\gamma} \tag{6}
$$

The utility function belongs to the family of constant utility of substitution, where the elasticity of substitution is given by $\varepsilon = \frac{1}{\sigma}$, with $0 < \sigma \leq 1$[30]. The parameter γ measures the intensity of the disutility of effort.

Let us now move to the political side of our economy. The public policy vector is defined as $q = (\tau_K, \tau_\omega)$: the two parties propose a platform that includes a capital income and a labor income tax. The policy vector is multidimensional, and generally in such a framework an equilibrium may not exists: we adopt a model with probabilistic voting, in the spirit of Lindbeck and Weibull (1987) and Coughlin and Nitzan (1981), which is particularly appropriate in our case since allows us to consider the ideological bias of the different social classes.

We assume that there are two parties, A and B; before the election takes place, parties choose, simultaneously and not cooperatively, the platform q that maximizes his expected number of voters. Politicians can commit to the policies promised during the campaign. We assume that voters are not only interested in the proposed policies, but also in the ideological elements that each party has. To be precise, voters are heterogeneous in terms of ideological preference: voter k in group $x = y_{TI}^P, y_{TC}^P, y_{TI}^R, y_{TC}^R, , ma_{TI}^P, ma_{TC}^P, ma_{TI}^R, ma_{TC}^R, o_{TI}^P, o_{TC}^P, o_{TI}^R, o_{TC}^R$ vote for party A if:

$$
V^x(q^A) + \psi + \sigma^{k,x} > V^x(q^B)
$$

where $V^x(q^A)$ is the indirect utility function of voters in group x if policy q^A is implemented and the term $(\psi + \sigma^{k,x})$ reflects voter's k ideological bias towards A. The component ψ is common to all voters

[30] In particular, for $\sigma = 1$, we have the logarithmic utility function ($\varepsilon = 1$), while $0 < \sigma < 1$ yields $\varepsilon > 1$ (substitutes) and $\sigma < 1$ yields $\varepsilon < 0$ (complements).

and is uniformly distributed on $\left[-\frac{1}{2d}, \frac{1}{2d}\right]$ with mean zero and density d. The ideology of voter k in group x is identified by the idiosyncratic parameter $\sigma^{k,x}$, which has group-specific uniform distributions over the interval $\left[-\frac{1}{2\phi^x}, \frac{1}{2\phi^x}\right]$, with zero mean and density ϕ^x.

The timing of the elections is as follows: (1) The two parties announce their policy platforms; at this stage, economic decisions are already made: therefore, parties knows voters' policy preferences and the distribution of the random variables ψ and $\sigma^{k,x}$, but not their realizations. (2) The value of d is realized and know. (3) Election takes place and the winning party implements his preferred policy.

Each group of individuals has neutral voters (also called swing voters) who are indifferent between A and B. The identity of the swing voters is crucial when a politician consider deviations form the common policy announcement $q^A = q^B$. To better understand this concept, consider only two groups, capitalist (who hold only capital) and workers (who have only labor). Suppose that party A decides to decrease τ_K with a corresponding increase in τ_ω such that the transfer P remains the same. Doing that, the party gains votes from the capitalist equal to the number of swing voters and lose votes from the group of workers equal to the number of swing voters. If the number of swing voters in the first group is greater than the number of swing voters in the second group, the party will have a net gain of votes. Therefore, each party is interested in attracting the more mobile voters in each group. A swing voter in group x is defined by σ^{sv} where:

$$\sigma^{sv} = V^x(q^B) - V^x(q^A) - \psi$$

All voters with $\sigma^{k,x} > \sigma^{sv}$ vote for A and voters with $\sigma^{k,x} < \sigma^{sv}$ vote for B. Therefore, the share of voters in group x that votes for party A is:

$$\pi^{A,x} = \phi^x \left(V^x(q^A) - V^x(q^B) + \psi\right) + \frac{1}{2} \tag{7}$$

Given the definition of the vote share (7), each party maximizes the following objective function:

$$\max_{\{q^A\}} \quad \sum_x v^x \phi^x \left(V^x(q^A) - V^x(q^B)\right) \tag{8}$$

where v^x represents the number of *voters* in each group x listed above. The central point of our paper is that the number of people who actually show up the election day is lower from the number of person alive in each generation. If the number of swing voters in every group x is the same, the problem (8) reduces to a simple maximization of average utilities. However, in our framework, groups differ in how votes can be swayed from one party to the other one. Therefore, parties try to please the more mobile voters by giving them more weight in the objective function.

6 Individuals' Problem

6.1 First Step: Labor Supply and Saving

Young Let us consider the problem for a young of income ω_i, for $i = R, P$. He chooses labor supply and saving for post-retirement consumption as to maximize the following intertemporal utility function, where the superscript j refers to the behavioral type:

$$\max_{l_y^i, c_y^{i,j}} \frac{\left(c_y^{i,j}\right)^{1-\rho}}{1-\rho} - \frac{\left(l_y^i\right)^{\gamma}}{\gamma} + \beta^j \delta \left(\frac{\left(c_{ma}^{i,j}\right)^{1-\rho}}{1-\rho} - \frac{\left(l_{ma}^i\right)^{\gamma}}{\gamma} + \delta \frac{\left(c_o^{i,j}\right)^{1-\rho}}{1-\rho} \right)$$

subject to:

$$\begin{aligned}
&0 < l_y^i < 1 \\
&c_y^{i,j} + s_y^{i,j} = \omega_i l_y^i (1-\tau_\omega) \\
&c_{ma}^{i,j} + s_{ma}^{i,j} = \omega_i l_{ma}^i (1-\tau_\omega) + s_y^{i,j}(1 + r(1-\tau_K) \\
&c_o^{i,j} = (1 + r(1-\tau_K) s_{ma}^{i,j}) + P
\end{aligned}$$

Replacing the budget constraints into the objective function, the maximization problem becomes:

$$\max_{l_y^i, c_y^{i,j}} \frac{\left(\omega_i l_y^i (1-\tau_\omega) - s_y^{i,j}\right)^{1-\rho}}{1-\rho} - \frac{\left(l_y^i\right)^{\gamma}}{\gamma} +$$
$$+ \beta^j \delta \left[\frac{\left(\omega_i l_{ma}^i (1-\tau_\omega) - s_y^{i,j}(1 + r(1-\tau_K) - s_{ma}^{i,j}\right)^{1-\rho}}{1-\rho} - \frac{\left(l_{ma}^i\right)^{\gamma}}{\gamma} + \delta \left(\frac{\left((1 + r(1-\tau_K)) s_{ma}^{i,j} + P\right)^{1-\rho}}{1-\rho} \right) \right]$$

subject to:

$$\begin{aligned}
&0 < s_y^{i,j} < \omega_i l_y^i (1-\tau_\omega) \\
&0 < l_y^i < 1
\end{aligned}$$

The FOCs of the problem are:

$$\begin{aligned}
&FOC\left\{s^y\right\} : \left(\omega_i (1-\tau_\omega) l_y^i - s_y^{i,j}\right)^{-\rho} = \beta^j \delta \left(\omega_i (1-\tau_\omega) l_{ma}^i + s_{ma}^{i,j}(1 + r(1-\tau_K))\right)^{-\rho} (1 + r(1-\tau_K) \\
&FOC\left\{l_y^i\right\} : \left[\omega_i (1-\tau_\omega)\right]^{1-\rho} \left(l_y^i\right)^{-\rho} - \left(l_y^i\right)^{\gamma-1} = 0
\end{aligned}$$

Optimal choices for thus given, respectively, by:

$$l_y^*(\omega_i) = (\omega_i (1-\tau_\omega))^{\alpha} \tag{9}$$

$$\left(s_y^{i,j}\right)^* = s(\beta^j, \omega_i, \tau_\omega, \tau_K) \tag{10}$$

where $\alpha = \frac{1-\rho}{\gamma+\rho-1} < 1$ and $\left(s_y^{i,j}\right)^* = s(\beta^j, \omega_i, \tau_\omega, \tau_K)$ is a function (whose closed form expression is given in the appendix) that describes optimal saving accumulation as a function of the parameters. The following proposition summarizes its the properties.

Proposition 1 *The saving function* $\left(s_y^{i,j}\right)^*$ *has the following properties, for* $i = P, R$ *and* $j = TI, TC$*:*
(i) For given j, $\left(s_y^{i,j}\right)^*$ *is increasing with the productivity level* ω_i*;*
(ii) $\left(s_y^{i,j}\right)^*$ *is decreasing with* τ_ω*; moreover,* $\frac{\partial^2 s_y^{i,j}}{\partial \tau_\omega \partial \omega_i} < 0$*;*
(iii) Depending on whether the substitution effect or income effect prevails, we have $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial \tau_K} > 0$ *or* < 0*;*
(iv) If the income effect prevails, then $\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial \tau_K \partial \omega_i} < 0$*; otherwise,* $\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial \tau_K \partial \omega_i} > 0$*;*
(v) For given i, $\left(s_y^{i,j}\right)^*$ *is increasing with the parameter of time inconsistency* β^j*:* $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial \beta^j} > 0$*;*
(vi) $\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial \tau_\omega \partial \beta^j} < 0$ *and* $\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial \tau_K \partial \beta^j} > 0$*.*

Proof. In appendix. ■

The first three results of the Proposition are intuitive: part *(i)* shows that, for a given level of time inconsistency, rich save more than poor: $s_y^{P,j} < s_y^{R,j}$: this is consistent with the evidence that a minority of rich holds the majority of capital of the economy.

In *(ii)*, we state that saving are a decreasing function of the labor income tax, τ_ω. This reduction is negatively correlated with productivity: for a given level of time inconsistency, if the labor income tax rate rises, poor reduce their savings more than rich.

Result *(iii)* is in line with the theoretical literature on taxation and saving (Bernheim, 1999). More precisely, depending on whether the uncompensated interest elasticity of saving is positive or negative, saving can either decrease or increase in response to a reduction of the capital tax rate, *i.e.* an increase in the after-tax rate of return of saving. From one hand, a reduction of τ_K reduces the price of consumption in periods 1 and 2: the associated *substitution effect* shifts consumption towards the future (*i.e.* saving *increase*), if future consumption is a normal good (as we assume). From the other hand, the income effect increases consumption in both periods (*i.e.* saving *decrease*). Unless we specify further the parameters of our model, we are not able to determine which effect prevails in our model. In the rest of the paper, we consider separately the two cases.

Furthermore, we show that rich and poor respond differently after an increase of τ_K (part *iv*): if the income effect prevails (saving *increases*), rich individuals will increase saving more than a poor individual with the same β^j. On the other hand, if the substitution effect prevails (saving *decreases*), then the derivative is positive: rich individuals decreases less their saving than poor.

In part *(v)*, we demonstrate that, for a given ω_i, time inconsistency leads to overconsumption: $s_y^{i,TC} > s_y^{i,TI}$, for $i = P, R$. This is a classical results in the behavioral literature, which has stressed (Laibson, 1997 and Laibson et al. 1998) that individuals regret about their saving rates and that retirees experience a drop in their post retirement consumption levels (Bernheim, 1998). Moreover, combining this result with part *(i)*, it is possible to show that, if there is enough inequality in the economy, *i.e.* $\omega_R >> \omega_P$, we have that $s_y^{P,TC} < s_y^{R,TI}$. Despite their time inconsistency, hyperbolic rich individuals continue to save more than poor and time consistent agents.

Part *(vi)* focuses on the effects of time inconsistency on saving accumulation; we first show that, keeping constant ω_i, the decrease of saving due to a higher τ_ω is more intense for hyperbolic consumer; the result is intuitive but meaningful: increasing τ_ω reduces individuals' disposable income and saving (see part *(iii)*); time inconsistent individuals, who are more likely to sacrifice future consumption in favor of present consumption, reduce more saving that exponential individuals. The second part of *(vi)* shows an interesting result: when τ_K changes, exponential are more responsive than time inconsistent in adapting their saving. More precisely, when the income (*resp.* substitution) effect prevails and saving increase (*resp.* decrease), exponential increase saving more (*resp.* less) than hyperbolic. The intuition for the result is the following: for hyperbolic young, the effects of a change in the tax are not only postponed in the future but also reduced, given that the weight attached to future utility is lower, and therefore they are less responsive in adapting their saving to the changes in the tax code.

Following Laibson (1997), it is possible to prove the following Corollary, which shows that time inconsistent agents would benefit from an increase of saving from $s_y^{i,TI}$ up to $s_y^{i,TC}$: if a commitment device that forces them to save up to this level would be made available, total welfare would increase. However, our assumption about the absence of such devices makes this Pareto improvement impossible.

Corollary 1 *Increasing saving from $s_y^{i,TC}$ to $s_y^{i,TI}$ is welfare-improving for time inconsistent individuals (young and middle aged)*

Proof. In appendix. ■

Middle Aged The problem of a middle aged individual of type $i = R, P$ and $j = TI, TC$ is:

$$\max_{l_{ma}^i, c_{ma}^{i,j}} \frac{\left(c_{ma}^{i,j}\right)^{1-\rho}}{1-\rho} - \frac{\left(l_{ma}^i\right)^\gamma}{\gamma} + \beta^j \delta \frac{\left(c_o^{i,j}\right)^{1-\rho}}{1-\rho}$$

subject to:

$$\begin{aligned}
&0 < l_{ma}^i < 1 \\
&c_{ma}^{i.j} + s_{ma}^{i,j} = \omega_i l_{ma}^i (1-\tau_\omega) + s_y^{i,j}(1 + r(1-\tau_K)) \\
&c_o^{i.j} = (1 + r(1-\tau_K) s_{ma}^{i,j}) + P
\end{aligned}$$

Replacing the budget constraint into the objective function, we get:

$$\max_{l_{ma}, s_{ma}^{i,j}} \frac{(\omega_i l_{ma}^i (1-\tau_\omega) + s_y^{i,j}(1 + r(1-\tau_K) - s_{ma}^{i,j})^{1-\rho}}{1-\rho} - \frac{\left(l_{ma}^i\right)^\gamma}{\gamma} + \beta^j \delta \left(\frac{((1 + r(1-\tau_K)) s_{ma}^{i,j} + P)^{1-\rho}}{1-\rho} \right)$$

subject to:

$$\begin{aligned}
&0 < s_{ma}^{i,j} < \omega_i l_{ma}^i (1-\tau_\omega) \\
&0 < l_{ma}^i < 1
\end{aligned}$$

for given $\omega_i, \tau_\omega, \tau_K$. It is easy to see that the first order conditions are:

$$FOC\left\{s_{ma}^{i,j}\right\}: \left(\omega_i(1-\tau_\omega)l_{ma}^i + s_y^{i,j}(1+r(1-\tau_K) - s_{ma}^{i,j}\right)^{-\rho} = \beta^j\delta\left(s_{ma}^{i,j}(1+r(1-\tau_K)+P\right)^{-\rho}(1+r(1-\tau_K)$$

$$FOC\left\{l_{ma}^i\right\}: \left[\omega_i(1-\tau_\omega)\right]^{1-\rho}\left(l_{ma}^i\right)^{-\rho} - \left(l_{ma}^i\right)^{\gamma-1} = 0$$

and the optimal levels of saving and labor supply are given by:

$$\left(s_{ma}^{i,j}\right)^* = s(\beta^j, \omega_i, \tau_\omega, \tau_K) \tag{11}$$

$$l_{ma}^*(\omega_i) = \left(\omega_i(1-\tau_\omega)\right)^\alpha \tag{12}$$

From equations (11) and (12), it is easy to see that labor supply does not depend on β^j, while the consumption saving trade-off is influenced by hyperbolic discounting. Comparative statics over the function $s(.)$ yields to the following proposition:

Proposition 2 *The saving function $\left(s_{ma}^{i,j}\right)^*$ has the following properties, for $i = P, R$ and $j = TI, TC$:*
(i) For given j, $\left(s_{ma}^{i,j}\right)^*$ is increasing with income;
(ii) $\left(s_{ma}^{i,j}\right)^*$ is decreasing in τ_ω; moreover, $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_\omega\partial\omega_i} < 0$;
(iii) Depending on whether the substitution effect or income effect prevails, we have $\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K} > 0$ or < 0;
(iv) If the income effect prevails, then $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K\partial\omega_i} < 0$; otherwise, $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K\partial\omega_i} > 0$;
(v) For given i, two effects determines the sign of $\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\beta^j}$: if the hyperbolic dominates the catching up effect, we have $\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\beta^j} > 0$; otherwise, $\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\beta^j} < 0$;
(vi) $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_\omega\partial\beta^j} < 0$ and $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K\partial\beta^j} > 0$.

Proof. In appendix. ■

Corollary 2 *Increasing saving from $s_{ma}^{i,TC}$ to $s_{ma}^{i,TI}$ is welfare-improving for time inconsistent individuals (young and middle aged)*

Proof. In appendix. ■

Intuitions behind Proposition 2 and Corollary 2 are similar to those of Proposition 1 and Corollary 1. The only exception is given by *(iv)*: it is possible that hyperbolic middle aged save *more* than a far-sighted. Depending on the value of β_j, two opposite effects determine the sign of this derivative: from one hand, the bias toward the present leads to overconsumption today (the *hyperbolic* effect). On the other hand, since we assume that hyperbolic are aware of their self-control issues, it is possible that, to finance consumption when old, they decide to save more, compared to an exponential individual (*catching up* effect The conditions determining which effect dominates are given in the appendix.

Old The problem for old individual is simple: they do not make any economic choice and only consume their accumulated saving and the transfer $P_{eq}(\tau_\omega, \tau_K)$.

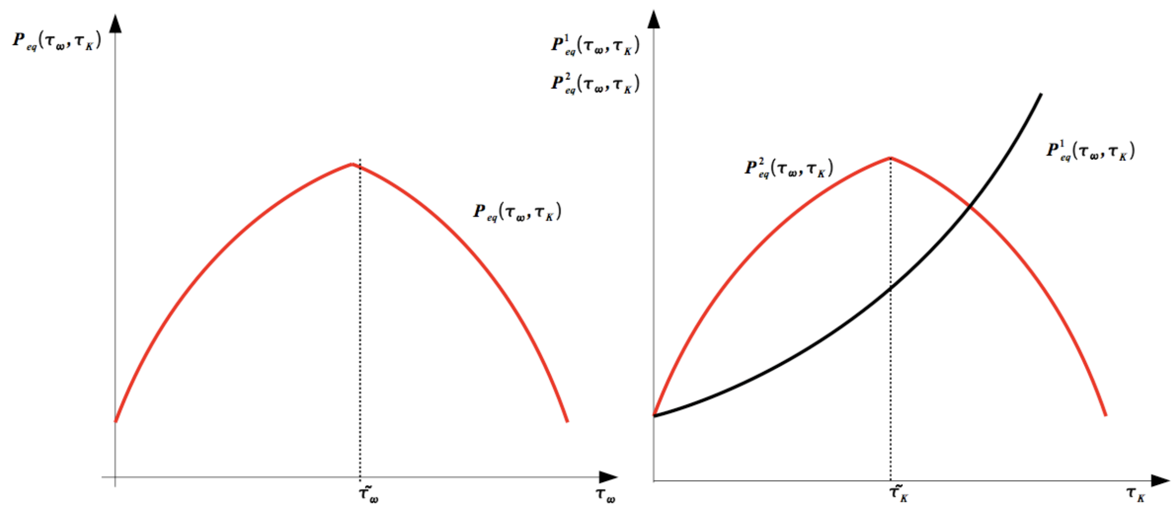

Figure 6: The Pension Function

The Transfer P Once optimal savings and the labor supply for young and middle aged are known, we compute the equilibrium pension transfer $P_{eq}(\tau_\omega, \tau_K)$ received by old (see the Appendix for the closed form expression for $P_{eq}(\tau_\omega, \tau_K)$) [31] and its properties.

Proposition 3 *The equilibrium transfer $P_{eq}(\tau_\omega, \tau_K)$ is:*

(i) Increasing in the level of the labor income tax up to $\tilde{\tau}_\omega$ and then decreasing;

(iia) If the substitution effect prevails, the pension function is increasing with the level of the capital income tax up to $\tilde{\tau}_K$ and then decreasing;

(iib) If the income effect prevails, the pension function is increasing and convex in τ_K.

Proof. In appendix. ■

Proposition 3 shows that $P_{eq}(\tau_\omega, \tau_K)$ is concave both in the labor income and in capital income tax rates (if saving decrease with τ_K), with maxima, respectively, at $\tilde{\tau}_\omega$ and $\tilde{\tau}_K$. However, when the income effect prevails, saving increases with the tax rate: it follows that the pension function is convex, with a maximum at $\tau_K = 1$ (see Figure 6). In the following, to make the analysis non-trivial[32], we will restrict our attention to the interval $\tau_\omega \in [0, \tilde{\tau}_\omega]$.

[31] In the following we denote with $l^*_y(\omega_i)$ and $l^*_{ma}(\omega_i)$ the optimal labor supplies for, respectively, young and middle aged, and with $s^*_y(\beta^j, \omega_i)$ and $s^*_{ma}(\beta^j, \omega_i)$ their optimal saving decisions, for income levels $i = R, P$ and the individuals' degree of time inconsistency $j = TI, TC$.

[32] If the tax rate is above $\tilde{\tau}_\omega$, it is obvious that every individual prefers to tax more capital, since it increases the transfer P_{eq}.

For future references, we determine indirect utility functions for a representative ij-type.

$$
\begin{aligned}
V_y^{i,j} &= \frac{\left[\omega_i(1-\tau_\omega)l_y^*(\omega_i)-s_y^*(\beta^j,\omega_i)\right]^{1-\sigma}}{1-\sigma} - \frac{\left(l_y^*(\omega_i)\right)^\gamma}{\gamma} \\
&+ \beta^j\delta\left[\frac{\left[\omega_i(1-\tau_\omega)l_y^*(\omega_i)-s_{ma}^*(\beta^j,\omega_i)+(1+r(1-\tau_K))s_y^*(\beta^j,\omega_i)\right]^{1-\sigma}}{1-\sigma} - \frac{(l_{ma}^*(\omega_i))^\gamma}{\gamma}\right] \\
&+ \beta^j\delta^2\frac{\left[s_{ma}^*(\beta^j,\omega_i)(1+r(1-\tau_K))+P\right]^{1-\sigma}}{1-\sigma} \qquad (13) \\
V_{ma}^{i,j} &= \frac{\left[\omega_i(1-\tau_\omega)l_y^*(\omega_i)-s_y^*(\beta^j,\omega_i)\right]^{1-\sigma}}{1-\sigma} - \frac{\left(l_y^*(\omega_i)\right)^\gamma}{\gamma} \qquad (14) \\
&+ \beta^j\delta\left[\frac{\left[\omega_i(1-\tau_\omega)l_y^*(\omega_i)-s_{ma}^*(\beta^j,\omega_i)+(1+r(1-\tau_K))s_y^*(\beta^j,\omega_i)\right]^{1-\sigma}}{1-\sigma} - \frac{(l_{ma}^*(\omega_i))^\gamma}{\gamma}\right] \\
V_o^{i,j} &= \frac{\left[(1+r(1-\tau_K))s_{ma}^*(\beta^j,\omega_i)+P_{eq}(\tau_\omega,\tau_K)\right]^{1-\sigma}}{1-\sigma} \qquad (15)
\end{aligned}
$$

6.2 Second step: To Vote or Not to Vote?

To account for the positive correlation between productivity level and political participation (voting turnout, campaign contributions. lobbying etc.)[33], we assume that it exists an exogenous costs C associated with voting activity (watching debates on TV, comparing different political platforms and candidates etc.). If these costs are high enough, an individual chooses not to vote [34].

The cost C is such that only a fraction z of poor votes [35], while all rich vote; the budget constraints are modified as follows:

$$
\begin{aligned}
c_y^{i,j} + s_y^{i,j} &= \omega_i l_y^i(1-\tau_\omega) - C \\
c_{ma}^{i,j} + s_{ma}^{i,j} &= \omega_i l_{ma}^i(1-\tau_\omega) + s_y^{i,j}(1+r(1-\tau_K) - C \\
c_o^{i,j} &= (1+r(1-\tau_K))(s_y^{i,j}+s_{ma}^{i,j}) + P - C
\end{aligned}
$$

Notice that, being C fixed, the comparative statics performed in the previous sections remains valid. Our assumption of lower turnout among poor create a discrepancy between the number of *voters* and the number of individuals *alive*. The number of voters in every group x, denoted by v^x, is given by:

[33]Our model does not want to explain the determinants of this correlation, but only its implications for a probabilistic voting model.

[34]We realize that this is a very simplifying assumption: a more realistic and complicated model should take into account that the voting decision results from as a trade-off between two opposite forces: from one hand, voting is costly and poor may decide not to vote; on the other hand, there are psychological factors, not related to any economic variable, that positively affect the probability of voting: for instance, some individuals perceive voting activity as a "duty", and thus they to do it anyway, whatever the cost is. The psychological motive could be modeled as an *i.i.d.* random variable R, with c.d.f. $F(.)$ and density $f(.)$. In this modified framework, very poor individuals with high psychological motivation may still decide to vote in equilibrium. We believe, however, that all main insights of our simplified model will hold also in such enlarged framework since for poor individuals the first force is still relevant, whereas for rich individuals the cost C remains negligible.

[35]Empirical evidence shows that there is a correlation between age and political participation: senior citizens more involved in the political process: however, for the moment, we neglect this additional stylized fact

$v_y^{P,TI} = z\rho^P \lambda^{TI} n^y$	Poor hyperbolic Y	$v_y^{P,TC} = z\rho^P \lambda^{TC} n^y$	Poor exponential Y
$v_y^{R,TI} = \rho^R \lambda^{TI} n^y$	Rich hyperbolic Y	$v_y^{R,TC} = \rho^R \lambda^{TC} n^y$	Rich exponential Y
$v_{ma}^{P,TI} = z\rho^P \lambda^{TI} n^{ma}$	Poor hyperbolic MA	$v_{ma}^{P,TC} = z\rho^P \lambda^{TC} n^{ma}$	Poor exponential MA
$v_{ma}^{R,TI} = \rho^R \lambda^{TI} n^{ma}$	Rich hyperbolic MA	$v_{ma}^{R,TC} = \rho^R \lambda^{TC} n^{ma}$	Rich exponential MA
$v_o^{P,TI} = z\rho^P \lambda^{TI} n^o$	Poor hyperbolic O	$v_o^{P,TC} = z\rho^P \lambda^{TC} n^o$	Poor exponential O
$v_o^{R,TI} = \rho^R \lambda^{TI} n^o$	Rich hyperbolic O	$v_o^{R,TC} = \rho^R \lambda^{TC} n^o$	Rich exponential O

The parameter z is chosen such that the number of rich individuals do not represent the majority of the electorate (the vote share of rich middle aged plus rich young plus rich old is lower than 1/2 of the total population), so that the policy proposed in equilibrium must also receive the approval of the poor classes, in every generation.

7 The Party's Choices: Solving the Model

Each party maximizes the expected total number of votes from the three generations currently alive, taking into account all the subgroups that exist within each generation, and the different turnouts level among rich and poor. Formally,

$$\max_{\{q^A\}} \sum_x v^x \phi^x (V^x(q^A) - V^x(q^B) \tag{16}$$

where $V_y^{i,j}(q^m), V_{ma}^{i,j}(q^m), V_o^{i,j}(q^m)$ are defined by (13), (14) and (15).

The equilibrium concept adopted is similar to Profeta (2004): the two parties decide the policy vector having in mind the utility of current generations . Young and middle aged expect, in a stationary equilibrium, the policies to be the same in future. Maximization of problem (16) yields to the two following FOCs:

$$FOC\{\tau_\omega\} : \sum_x v^x \phi^x \frac{dV^x}{d\tau_\omega} = 0 \tag{17}$$

$$FOC\{\tau_K\} : \sum_x v^x \phi^x \frac{dV^x}{d\tau_K} = 0$$

8 Equilibrium

8.1 Labor Income Tax Rates

Preferred tax rates for the different groups in our economy have the following properties.

Proposition 4 *Preferred labor tax rates have the following properties:*

(i) For a given degree of time inconsistency, preferred labor tax rates are decreasing with income: $\tau_\omega^y(\omega_P, \beta^j) > \tau_\omega^y(\omega_R, \beta^j)$ *and* $\tau_\omega^{ma}(\omega_P, \beta^j) > \tau_\omega^{ma}(\omega_R, \beta^j)$*;*

(ii) Every old individual set $\tau_\omega^o(\omega_i, \beta^j) = \tilde{\tau}_\omega$, $\forall i, j$*;*

(iii) For a given income level, hyperbolic consumers prefer lower labor income taxes than time consistent ones: $\tau_\omega^y(\omega_i, \beta^{TC} = 1) > \tau_\omega^y(\omega_i, \beta^{TI})$ *and* $\tau_\omega^{ma}(\omega_i, \beta^{TC} = 1) > \tau_\omega^{ma}(\omega_i, \beta^{TI})$*;*

(iv) If there is enough inequality in the economy, we have that, for young individual: $\tau_\omega^y(\omega_R, \beta^{TC} = 1) > \tau_\omega^y(\omega_R, \beta^{TI}) > \tau_\omega^y(\omega_P, \beta^{TC} = 1) > \tau_\omega^y(\omega_P, \beta^{TI})$*;*

(v) If there is enough inequality in the economy, we have that, for middle aged individual: $\tau_\omega^{ma}(\omega_R, \beta^{TC} = 1) > \tau_\omega^{ma}(\omega_R, \beta^{TI}) > \tau_\omega^{ma}(\omega_P, \beta^{TC} = 1) > \tau_\omega^{ma}(\omega_P, \beta^{TI})$*.*

Proof. In appendix. ■

Proposition 4 sheds light on the voting behavior of the different groups: in *(i)*, we show the intuitive result that preferred τ_ω are decreasing with ω_i: poor, looking for more intergenerational redistribution, prefer to increase the tax as to augment the transfer P.

In *(ii)*, we show that all old (rich, poor, time consistent and time inconsistent) set the same $\tau_\omega = \tilde{\tau}_\omega$, namely the tax rate that maximize the value of the transfer. This is intuitive, since all the economic decisions have been already taken, they maximize consumption levels by maximizing P.

In *(iii)*, we analyse the second source of heterogeneity, keeping ω_i constant. We show that hyperbolic individuals set lower tax rates than exponential: the intuition is that the former group faces a different trade-off for labor taxation than the latter: for hyperbolic, increasing τ_ω has a current cost (it reduces labor supply and consumption), and a benefit that is postponed in the future (it increases the transfer P at $t = 3$) as it is discounted by the lower factor $\beta^2\delta$. On the other hand, exponential, who fully understand the intertemporal trade-off at stake, set the "correct" tax rate.

Finally, in *(iv)* and *(v)*, we aggregate for the two sources of heterogeneity and we rank preferred labor tax rates as follows:

$$\begin{aligned} \tau_\omega^o(\omega_i, \beta^j) = \tilde{\tau}_\omega > \tau_\omega^{ma}(\omega_P, \beta^{TC}) > \tau_\omega^{ma}(\omega_P, \beta^{TI}) > \tau_\omega^y(\omega_P, \beta^{TC}) > \\ > \tau_\omega^y(\omega_P, \beta^{TI}) > \tau_\omega^{ma}(\omega_R, \beta^{TC}) > \tau_\omega^{ma}(\omega_R, \beta^{TI}) > \tau_\omega^y(\omega_R, \beta^{TC}) > \tau_\omega^y(\omega_R, \beta^{TI}) \end{aligned} \tag{18}$$

8.2 Capital Income Tax Rates

Depending on whether higher τ_K increases (*resp.* decreases) savings, *i.e* the income (*resp.* substituion effect) prevails, two different cases are possible.

8.2.1 Case (a): Increasing τ_K reduces Saving

If the income effect is lower than the substitution effect, the pension function is increasing and concave in τ_K (see figure 6). Proposition 5 follows immediately.

Proposition 5 (Substituion Effect Dominates) *Preferred tax rates, denoted* $\tau_K^g(\beta^j, \omega_i)$*, for* $g = y, ma, o$*;* $i = R, P$ *and* $j = TI, TC$*, satisfy the following properties:*

(i) $\tau_K^g(\beta^j, \omega_i)$ *are decreasing with income,* $\forall g$*, for a given* j*;*

(ii) $\tau_K^g(\beta^j, \omega_i)$ *are decreasing with the parameter of time inconsistency* β^j*,* $\forall g$*, and for a given* i*;*

(iii) For given i *and* j*, we have:* $\tau_K^o(\beta^j, \omega_i) > \tau_K^y(\beta^j, \omega_i) = \tau_K^{ma}(\beta^j, \omega_i)$*.*

(iv) If there is enough inequality in the economy, we have that, for given g*:*

$$\tau_K^g(\omega_R, \beta^{TC}) < \tau_K^g(\omega_R, \beta^{TI}) < \tau_K^g(\omega_P, \beta^{TC}) < \tau_K^g(\omega_P, \beta^{TI}).$$

Proof. In appendix. ■

Part *(i)* shows that preferred capital taxes are decreasing with income. This result has two reasons: first, poor save less, and a higher tax on capital reduces less consumption levels and utiltity. Second, poor benefit more from redistribution by increasing τ_K and P.

Part *(ii)*, keeping constant τ_ω, analyzes how time inconsistency affects preferred capital tax rates. We show that, within all generations, $\tau_\omega(\omega_i, \beta^{TC}) < \tau_\omega(\omega_i, \beta^{TI})$. Two effects determine this result.

First, there is a direct effect: hyperbolic are less hurt by a reduction of the after tax return of saving, since they save less than far-sighted. Second, there is an indirect effect: Propositions 1*(v)* and 2*(v)* show that the decrease in saving due to a higher τ_K is lower for time inconsistent agents. Therefore, the decrease in current and future utility is lower for this group. Third, there is an hyperbolic effect: capital taxes lead to an intertemporal trade-off not present in labor taxation. Taxing more capital income increases current consumption (which is beneficial, from the perspective of a present biased individual) at a delayed costs (less consumption tomorrow, due to reduced saving and lower after tax capital income).

All effects goes in the same direction: it follows that hyperbolic would like to set higher capital taxes than exponential, in order to keep constant P^{eq}.[36]

Part *(iii)* shows that, for a given β^j and ω_i, old prefer higher taxes than young and middle aged. Like for labor taxes, old do not make any economic decision: they set taxes as to maximize consumption levels. Notice that the preferred tax is lower than $\tilde{\tau}_K$, the tax that maximizes P^{eq}.

Finally, in *(iv)* we aggregate for the two sources of heterogeneity and we rank preferred labor tax rates:

$$\begin{aligned} &\tilde{\tau}_K > \tau_K^o(\omega_P, \beta) > \tau_K^o(\omega_P, 1) > \tau_K^o(\omega_R, \beta) > \tau_K^o(\omega_R, 1) > \\ &> \bar{\tau}_K(\omega_P, \beta) > \bar{\tau}_K(\omega_P, 1) > \bar{\tau}_K(\omega_R, \beta) > \bar{\tau}_K(\omega_R, 1) \end{aligned} \tag{19}$$

where $\bar{\tau}_K(\omega_i, \beta^j)$ is the common preferred tax rate for young and middle aged with the same i and j.

[36] In Proposition 5 we have assumed that middle aged time inconsistent save less than exponential, *i.e.* the exponential effect dominates the catching up effect. If this is not the case, and hyperbolic saves more the two effects described before goes in opposite directions. A priori, we do not know whether the chain of inequalities (20) changes or not. If yes, we have that: $\tau_K^{ma}(\omega_P, 1) > \tau_K^{ma}(\omega_P, \beta) > \tau_K^{ma}(\omega_R, 1) > \tau_K^{ma}(\omega_R, \beta)$.

8.2.2 Case (b): Increasing τ_K increases Saving

If the income effect dominates the substitution effect, the pension function is increasing and convex in τ_K (see figure 6). The following proposition summarize the properties of preferred tax rates.

Proposition 6 (Income Effect prevails) *Preferred tax rates, denoted* $\hat{\tau}_K^g(\beta^j, \omega_i)$*, for* $g = y, ma, o$*;* $i = R, P$ *and* $j = TI, TC$*, satisfy the following properties:*

(i) $\tau_K^g(\beta^j, \omega_i)$ *are increasing with income,* $\forall g$*, for a given* β^j*;*

(ii) $\tau_K^g(\beta^j, \omega_i)$ *are decreasing with the degree of time inconsistency* β^j*,* $\forall g$*, and for a given* ω_i*;*

(iii) For given i *and* j*, we have* $\tau_K^o(\beta^j, \omega_i) > \tau_K^y(\beta^j, \omega_i) = \tau_K^{ma}(\beta^j, \omega_i)$*.*

(iv) If there is enough inequality in the economy, we have that, for given g*:* $\tau_\omega^g(\omega_R, \beta^{TC}) < \tau_\omega^g(\omega_R, \beta^{TI}) < \tau_\omega^g(\omega_P, \beta^{TC}) < \tau_\omega^g(\omega_P, \beta^{TI})$*.*

(v) $\forall i, j, g$*, we have that:* $\hat{\tau}_K^g(\beta^j, \omega_i) > \tau_K^g(\beta^j, \omega_i)$*.*

Proof. In appendix. ■

Results do not change substantially from Case (a): this is not surprising, as the effects (positive or negative) of a change in τ_K are soften for hyperbolic individuals (as accumulated savings are lower) and delayed in time.

However, a different result is given by *(i)*: now, rich individuals prefer *higher* taxes than poor individual: for them, consumption in the future is relatively cheaper, and a higher tax increases it through saving. Finally, in part (v), we claim that preferred capital tax rates are always higher if the income effect prevails than when the substitution effect prevails.

8.3 Political Equilibria

Given the structure of voters' preferred tax rates, it is immediate to see why time inconsistent individuals prefer to have a policy vector in which capital taxes are relatively higher than labor income ones. To simplify, in the following we are going to concentrate on Case (a), *i.e* saving decrease in response to an increase in τ_K. The following lemma shows that time inconsistent individuals are more single minded that time consistent ones.

Lemma 1 *Hyperbolic individuals are more ideologically homogeneous (single minded) than time consistent ones.*

When voting over the composition of the tax burden that finances a redistributive transfer P, individuals take into account not only the factors (labor and capital) they own, but also the timing of the tax. Single mindedness comes form the fact that time inconsistent agents prefer a higher τ_K compared to a far sighted with the same income level. Two effects determines this result: first, hyperbolic own less capital than exponential. Second, the effectsof a higher tax are postponed in the future (if young) and soften by the suboptimality of his choice (if middle aged).

Lemma 1 allows us to fully describe the set of equilibria of the model.

Proposition 7 *Both parties, in equilibrium, converge to the same fiscal platform:* $q^{eq} = (\tau_\omega^{eq}, \tau_K^{eq})$*. The vector* q^{eq} *is characterized as follows:*

(i) if $z = 1$ *and* $\rho^{TI} = 0$*,* q^{eq} *is such that:*

$$\tau_\omega^y(\omega_R) < \tau_\omega^{ma}(\omega_R) < \tau_\omega^y(\omega_P) < \tau_\omega^{ma}(\omega_P) < \boldsymbol{\tau_\omega^{eq}} < \tilde{\tau}_\omega \tag{20}$$

$$\bar{\tau}_K(\omega_R) < \boldsymbol{\tau_K^{eq}} < \bar{\tau}_K(\omega_P) < \tau_K^o(\omega_P) < \tau_K^o(\omega_R) < \tilde{\tau}_K$$

(ii) if $z < 1$ *and* $\rho^{TI} = 0$*,* q^{eq} *is such that:*

$$\tau_\omega^y(\omega_R) < \tau_\omega^{ma}(\omega_R) < \boldsymbol{\tau_\omega^{eq}} < \tau_\omega^y(\omega_P) < \tau_\omega^{ma}(\omega_P) < \tilde{\tau}_\omega \tag{21}$$

$$\boldsymbol{\tau_K^{eq}} < \bar{\tau}_K(\omega_R) < \bar{\tau}_K(\omega_P) < \tau_K^o(\omega_P) < \tau_K^o(\omega_R) < \tilde{\tau}_K$$

(iii) if $z = 1$ *and* $\rho^{TI} > 0$*,* q^{eq} *is such that:*

$$\begin{aligned}&\tau_\omega^y(\omega_R, \beta) < \tau_\omega^y(\omega_R, 1) < \tau_\omega^{ma}(\omega_R, \beta) < \tau_\omega^{ma}(\omega_R, 1) < \\ &< \tau_\omega^y(\omega_P, \beta) < \boldsymbol{\tau_\omega^{eq}} < \tau_\omega^y(\omega_P, 1) < \tau_\omega^{ma}(\omega_P, \beta) < \tau_\omega^{ma}(\omega_P, 1) < \tilde{\tau}_\omega\end{aligned} \tag{22}$$

$$\begin{aligned}&\bar{\tau}_K(\omega_R, 1) < \bar{\tau}_K(\omega_R, \beta) < \bar{\tau}_K(\omega_P, 1) < \bar{\tau}_K(\omega_P, \beta) < \boldsymbol{\tau_K^{eq}} < \\ &< \tau_K^o(\omega_R, 1) < \tau_K^o(\omega_R, \beta) < \tau_K^o(\omega_P, 1) < \tau_K^o(\omega_P, \beta) < \tilde{\tau}_K\end{aligned}$$

(iv) if $z < 1$ *and* $\rho^{TI} > 0$*,* q^{eq} *is such that:*

$$\begin{aligned}&\tau_\omega^y(\omega_R, \beta) < \tau_\omega^y(\omega_R, 1) < \tau_\omega^{ma}(\omega_R, \beta) < \tau_\omega^{ma}(\omega_R, 1) < \boldsymbol{\tau_\omega^{eq}} < \\ &< \tau_\omega^y(\omega_P, \beta) < \tau_\omega^y(\omega_P, 1) < \tau_\omega^{ma}(\omega_P, \beta) < \tau_\omega^{ma}(\omega_P, 1) < \tilde{\tau}_\omega\end{aligned} \tag{23}$$

$$\begin{aligned}&\bar{\tau}_K(\omega_R, 1) < \bar{\tau}_K(\omega_R, \beta) < \boldsymbol{\tau_K^{eq}} < \bar{\tau}_K(\omega_P, 1) < \bar{\tau}_K(\omega_P, \beta) < \\ &< \tau_K^o(\omega_R, 1) < \tau_K^o(\omega_R, \beta) < \tau_K^o(\omega_P, 1) < \tau_K^o(\omega_P, \beta) < \tilde{\tau}_K\end{aligned}$$

In equilibrium, both parties propose the same platform, as problem (17) is the same. Policy vectors coincide, $q^A = q^B$, and individuals reach the same utility levels under the two platforms, $V_g^{i,j}(q^A) = V_g^{i,j}(q^B)$, $\forall i, j, g$.

In part *(i)*, we show that, if all poor vote, $z = 1$, and time inconsistency is not an issue ($\rho^{TI} = 0$), Tabellini-Persson (2003) holds: representing the majority of the electorate, and holding less capital, poor prefer to tax more capital than labor: both parties will then propose a policy vector that includes poor preferred tax rates.

In part (*ii*), we show that, if $\rho^{TI} = 0$, and with turnout positively correlated to income level, the upper class, who saves more, becomes more attractive for the two parties which are willing to reduce both taxes, and the transfer P, as rich individuals are not interested in redistribution, and prefer to keep the transfer as lowest ast possible.

Part (*iii*) considers the case of full turnout and time inconsistency: the policy platform is distorted toward capital taxation, and the equilibrium capital tax is higher than incase *(i)*: in this case, also time inconsistent individuals prefer to tax more capital than labor income, given that their saving are suboptmal, and they are more mobile than exponential rich.

Finally, in part (*iv*), we assume that $z < 1$ and $\rho^{TI} > 0$. To win the elections, parties have to please the swing voters: Lemma 1 shows that hyperbolic care more about labor income taxation and are more "single minded" and more likely to sway their vote if the tax burden is more distorted toward capital taxation. Therefore, proposing hyperbolic's preferred τ_K and τ_ω, both parties receive the support of hyperbolic rich and the fraction of politically active poor.

9 An Illustration

Without loss of generality, let us suppose that it exists only one generation, and that parameters are such that $s^{P,TI} < s^{P,TC} = s^{R,TI} < s^{TC,R}$. There are $n+1$ individuals in our economy; the n agents are equally split into the four groups (*i.e.* each group has size 1/4) and there is also a "lonely" poor[37], who can be either hyperbolic or exponential. Following Propositions 4 and 5, we have that preferred capital tax rates are such that: $\tau_K^{P,TI} > \tau_K^{P,TC} = \tau_K^{R,TI} > \tau_K^{TC,R}$. With exponential preferences and full turnout, there are only two preferred tax rates: τ_K^P and τ_K^R, with $\tau_K^P > \tau_K^R$. The probabilisting voting equilibrium is such that τ_K^P would be proposed in the equilibrium platform q by both parties.

Suppose now that no poor, except for the lonely one, show up at the election day: the implemented τ_K will be τ_K^R, and this is the paradoxical result we do not observe in reality. However, with hyperbolic agents (and limited turnout), an opportunistic party maximizes the number of votes obtainde by proposing the τ_K preferred by $s_{ma}^{R,TI}$, as to gain $\frac{1}{2}+1$ of the *effective* electorate.

[37]The additional individual breaks the eventual ties; moreover, assuming that poor represent the majority of the popolation is a reasonable assumption.

10 Conclusions

This paper sheds light on the political determination of the tax burden (capital taxation and labor taxation) that finances a redistributive transfer: in particular, we justify why we observe positive level of capital taxes, in spite of the popular view and most normative economic theories.

Two ingredients characterize our political model: first, we take into account that political participation is an increasing function of income (Lijpart, 1994) and therefore positive levels of capital taxation can not be justified simply by invoking the political power of the lower class. Second, by introducing time inconsistency in individuals' preferences, we show that also the upper class may be in favor of tax burden distorted toward capital taxation. The intuition for our result is fairly intuitive: time inconsistent individuals are more politically homogeneous ("single-minded") and, for any income level, prefer to tax more capital, instead of labor income, since accumulated savings are below the planned (and optimal) level. An opportunistic political party that maximizes the probability of winning the elections, therefore, finds profitable to propose a policy vector that entails a positive, and possibly high, level of capital taxation: the group of hyperbolic individuals has more "swing voters", and it is more united in his political action.

References

[1] Abel, A. B. (2007), "Optimal Capital Income Taxation", *mimeo.*

[2] Ainslie, G. (1992), *Picoeconomics*, Cambridge, UK: Cambridge University Press.

[3] Arneson, B. (1925), "Non-Voting in a Tipical Ohio Community", *American Political Science Review*, Vol. 19, 816-25.

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

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

[6] Auerbach, A.J. (1979), "The Optimal Taxation of Heterogeneous Capital", *Quarterly Journal of Economics*, Vol. 93, 589-612.

[7] Auerbach, A.J. (2006), "The Future of Capital Income Taxation", *mimeo.*

[8] Auerbach, A.J. and J.R. Hines (1998), "Excess Burden and Optimal Taxation", *Handbook of Public Economics*, Vol. 3, Chapter 18, edited by A.J. Auerbach and M. Feldstein, Elsevier Science Publishers B.V. (North Holland).

[9] Auerbach, A.J., L.J. Kotlikoff and J. Skinner (1983), "The Efficiency Gains from Dynamic Tax Reforms", *International Economic Review*, Vol. 24, 81-100.

[10] Auten, G. (1999), "Capital Gains Taxation", *The Encyclopedia of Taxation and Tax Policy*, edited by Joseph J. Cordes, Robert D. Ebel, and Jane G. Gravelle, Urban Institute Press.

[11] Bartels, L.M. (2005), "Economic Inequality and Political Representation", *mimeo.*

[12] Bartels, L.M. (2007), "What's The Matter with What's Matter with Kansas?", *mimeo.*

[13] Bernheim, B.D. (1999), "Taxation and Savings", *Handbook of Public Economics*, Vol. 3, Chapter 18, edited by A.J. Auerbach and M. Feldstein, Elsevier Science Publishers B.V. (North Holland).

[14] Bernasconi M. and P. Profeta (2007), "Redistribution or Education? The Political Economy of Social Race", *mimeo.*

[15] Boyd, R. (1989), "The Effects of Primaries and Statewide Races on Voter Turnout", *Journal of Politics*, Vol. 51, 730-39.

[16] Caplan, B.D. (2007), *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*, Princeton University Press.

[17] Campante, F.R. (2006), "Redistribution in a Model of Voting and Campaign Contributions", *mimeo.*

[18] Campbell, A. (2002), "Self-Interest, Social Security, and the Distinctive Participation Patterns of Senior Citizens", *American Political Science Review*, Vol. 96(3), 565-574.

[19] Carey, D. and J. Rabesona (2004), "Tax Ratios on Labor and Capital and on Consumption", in P.B.Sorensen (ed.), *Measuring the Tax Burden on Capital and Income*, CESifo Seminar Series, MIT Press, Cambridge, MA.

[20] Chamley, C. (1981), "The Welfare Cost of Capital Income Taxation in a Growing Economy", *Journal of Political Economy*, Vol. 89, 468-96.

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

[22] 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.

[23] Cowen, T. (2007), *www.marginalrevolution.com*

[24] Cremer, H., J.C Rochet and P. Pestieau (2003), “Capital Income Taxation when Inherited Health is not Observable”, *Journal of Public Economics*, Vol. 87, 2475-2490.

[25] Dahl, R. (1961), *Who Governs? Democracy and Power in an American City*, Yale University Press, New Haven, CT.

[26] Devereux, M.P., R. Griffith and A. Klemm (2002), “Corporate Income Tax Reforms and International Tax Competition”, *Economic Policy*, Vol. 61, 8-27.

[27] Diamond, P.A. (1970), “Taxation and Public Production in a Growth Setting”, in *Models of Economic Growth*, J.A. Mirrlees and N.H. Stern (eds,) London: MacMillan, 215-234.

[28] Erosa, A. and M. Gervais (2002), “Optimal Taxation in Life-Cycle Economies ”, *Journal of Economic Theory*, Vol. 105(2), 338-369.

[29] Feldstein, M.S. (1978), “The Welfare Cost of Capital Income Taxation”, *Journal of Political Economy*, Vol. 86(2), 29-51.

[30] Feldstein M.S. (2006), “Taxes on Investment Income Remain Too High and Leas to Multiple Distortions”, *Economists' Voice*.

[31] Fischer, S. (1980), “Dynamic Inconsistency, Cooperation, and the Benevolent Dissembling Government”, *Journal of Economic Dynamics and Control*, Vol. 2, 93-107.

[32] Krugman, P. (2007), *The Conscience of a Liberal*, W. W. Norton.

[33] Judd, K.L. (1985), “Redistributive Taxation in a Simple Perfect Foresight Model”, *Journal of Public Economics*, Vol. 28, 59-83.

[34] Judd, K.L. (1987), “The Welfare Cost of Factor Taxation in a Simple Foresight Model”, *Journal of Political Economy*, Vol.95(4), 675-709.

[35] Judd, K.L. (1997), “The Optimal Tax Rate for Capital Income is Negative”, *NBER Working Papers* N. 6004, National Bureau of Economic Research, Inc.

[36] Judd, K.L. (1999), “Optimal Taxation and Spending in general Competitive Growth Models”, *Journal of Public Economics,*, Vol. 71 (1), 1-25.

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

[38] Laibson, D. (1997), “Golden Eggs and Hyperbolic Discounting”, *Quarterly Journal of Economics*, 112 (2), 443-477.

[39] Laibson, D., A. Repetto and J. Tobacman (1998), “Self-Control and Saving for Retirement” *Brookings Paper on Economic Activity.*

[40] Laibson, D., A. Repetto and J. Tobacman (2007), “Estimating Discount Functions from Lifecycle Consumption Choices ”, *NBER Working Papers* N.13314, National Bureau of Economic Research, Inc.

[41] Lijphart, A. (1994), “Democracies: Forms, Performance, and Constitutional Engineering ”, *European Journal of Political Research*, Vol. 25(1), 1-17

[42] Lijphart, A. (1997), “Unequal Participation: Democracy’s Unresolved Dilemma”, *American Political Science Review*, Vol. 91(1), 1-14.

[43] Lindbeck, A. and J.W. Weibull (1987), “Balanced-Budget Redistributions as the Outcome of Political Competition”, *Public Choice*, Vol. 52, 146 -183.

[44] Mulligan, C. and X. Sala-i-Martin (2003), “Social Security, Retirement, and the Single-Mindedness of the Electorate”, *NBER Working Papers* N. 9691, National Bureau of Economic Research, Inc.

[45] Persson, T. and G. Tabellini (2003), *Political Economics: Explaining Economic Policy*, MIT Press: Cambridge, MA.

[46] Phelps, E. S. and R. A. Pollak (1968), “On Second-Best National Saving and Game-Equilibrium Growth”, *The Review of Economic Studies*, Vol. 35(2), 185-199.

[47] Profeta, P. (2002), “Retirement and Social Security in a Probabilistic Voting Model”, *International Tax and Public Finance*, Vol. 9, 331-348.

[48] Rosenstone, S.J. and J.M. Hansen (1993), *Mobilization, Participation, and Democracy in America*, New York: Macmillan.

[49] Sorensen, P.B. (2007), “Can Capital Income Taxes Survive? And Should They?”, *mimeo*

[50] Strotz, R. H. (1956), “Myopia and Inconsistency in Dynamic Utility Maximization”, *The Review of Economic Studies*, Vol. 23(3), 165-180.

[51] Tingsten, H. (1937), *Political Behavior: Studies in Election Statistics*, London: P. S. King & Son.

[52] Topf, R. (1995), “Electoral Participation”, in *Citizen and the State*, ed. H.D. Klingemann and D. Fuchs, Vol. 1, of *Beliefs in Government*, Oxford: Oxford University Press.

[53] Verba, S., N.Nie, and J. Kim (1978), *Participation and Political Equality: A Seven-Nation Comparison*, New York: Cambridge University Press.

[54] Verba, S., K.L. Schlozman, and H. E. Brady (1995), *Voice and Equality: Civic Voluntarism in American Politics*, Cambridge, MA: Harvard University Press.

A Appendix

A.1 Closed Form expressions for Saving

In Section 6 we provided the generic form for the equilibrium saving functions. If we assume that $u(c)$ and $v(l)$ have the CES specification, $(s_y^{i,j})^*$ and $(s_{ma}^{i,j})^*$, for $i = P, R$ and $j = TI, TC$, are given by:

$$\left(s_y^{i,j}\right)^* = \frac{(\omega_i(1-\tau_\omega))^\alpha\left[1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-2}{\sigma}}\left(1+\left(\beta^j\delta\right)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}\right)^{-1}\right]-\left(\beta^j\delta(1+r(1-\tau_K))\right)^{-\frac{2}{\sigma}}P}{\left(1+(\beta^j\delta)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}+\left((\beta^j\delta)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}\right)^2\right)} \tag{A24}$$

$$\left(s_{ma}^{i,j}\right)^* = \frac{(\omega_i(1-\tau_\omega))^\alpha\left[1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}}M^{\frac{\sigma-2}{\sigma}}\left(1+\left(\beta^j\delta\right)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}\right)^{-1}\right]}{\left(1+(\beta^j\delta)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}+\left((\beta^j\delta)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}\right)^2\right)}+ \tag{A25}$$

$$-\frac{\left(\beta^j\delta(1+r(1-\tau_K))\right)^{-\frac{1}{\sigma}}P}{\left(1+(\beta^j\delta)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}\right)}\left(1+\frac{\left(\beta^j\delta\right)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}}{\left(1+(\beta^j\delta)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}+\left((\beta^j\delta)^{-\frac{1}{\sigma}}(1+r(1-\tau_K))^{\frac{\sigma-1}{\sigma}}\right)^2\right)}\right)$$

where $\beta^j = \left\{\beta^{TI}, \beta^{TC}\right\}$ denotes the short-term discount factor, with $1 = \beta^{TC} > \beta^{TI}$ and $\alpha = \frac{\sigma(\gamma-\sigma)-1}{\sigma(\gamma-1)-1}$.

A.2 Useful Derivatives

The sign of some derivatives will be crucial in the following proofs. First, let us introduce some notation:

$$M = (1 + r(1 - \tau_K)) > 0$$
$$K = \left(\beta^j\delta\right)^{-\frac{1}{\sigma}} M^{\frac{\sigma-1}{\sigma}} > 0$$
$$R = 1 + K + K^2 > 0$$

Expressions (A24) and (A25) can be reduced to:

$$\left(s_y^{i,j}\right)^* = \frac{(\omega_i(1-\tau_\omega))^\alpha\left[1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}}M^{\frac{\sigma-2}{\sigma}}(1+K)^{-1}\right]-\left(\beta^j\delta M\right)^{-\frac{2}{\sigma}}P}{R} > 0 \tag{A26}$$

$$\left(s_{ma}^{i,j}\right)^* = \frac{(\omega_i(1-\tau_\omega))^\alpha\left[R+M\left(1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}}M^{\frac{\sigma-2}{\sigma}}(1+K)^{-1}\right)\right]}{R(1+K)} - \frac{\left(\beta^j\delta M\right)^{-\frac{1}{\sigma}}P}{(1+K)}\left(1+\frac{K}{R}\right) > 0 \tag{A27}$$

We are interested in determining how K, R and M vary with the capital income tax rate, τ_K, and the short-term discount factor, β^j:

$\frac{\partial M}{\partial \tau_K} = -r < 0;$

$\frac{\partial K}{\partial \tau_K} = \left(\frac{\sigma-1}{\sigma} \left(\beta^j \delta M\right)^{-\frac{1}{\sigma}} \frac{\partial M}{\partial \tau_K} \right) = \left(\frac{1-\sigma}{\sigma} \left(\beta^j \delta M\right)^{-\frac{1}{\sigma}} r \right) > 0$, since $\sigma \in (0,1)$

$\frac{\partial R}{\partial \tau_K} = \frac{\partial K}{\partial \tau_K} (1+2K) > 0$

$\frac{\partial M}{\partial \beta^j} = 0$

$\frac{\partial K}{\partial \beta^j} = -\left(\frac{1}{\sigma} \left(\beta^j \delta\right)^{-\frac{1}{\sigma}-1} M^{\frac{\sigma-1}{\sigma}} \right) < 0$

$\frac{\partial R}{\partial \beta^j} = \frac{\partial K}{\partial \beta^j} (1+2K) < 0$

For future references, we compute also how differences in β^j influence the marginal variation of saving in response to changes in τ_K:

$\frac{\partial^2 M}{\partial \tau_K \partial \beta^j} = 0$

$\frac{\partial^2 K}{\partial \tau_K \partial \beta^j} = -\left(\frac{1-\sigma}{\sigma^2} \delta M \left(\beta^j \delta M\right)^{-\frac{1}{\sigma}-1} r \right) < 0$

$\frac{\partial R}{\partial \tau_K \partial \beta^j} = \frac{\partial^2 K}{\partial \tau_K \partial \beta^j} (1+2K) + 2 \frac{\partial K}{\partial \tau_K} \frac{\partial K}{\partial \beta^j} < 0$

A.3 Proof of Proposition 1

Part *(i)* is straightforward: from (A26) it is immediate to see that $\frac{\partial \left(s_y^{i,j}\right)^*}{\partial \omega_i} > 0$.

For *(ii)*, differentiating (A26) with respect to τ_ω, we obtain:

$$\frac{\partial \left(s_y^{i,j}\right)^*}{\partial \tau_\omega} = -\alpha \left(\omega_i (1-\tau_\omega)\right)^{\alpha-1} \frac{\left[1 - \left(\beta^j \delta\right)^{-\frac{2}{\sigma}} M^{\frac{\sigma-2}{\sigma}} (1+K)^{-1}\right]}{R} < 0$$

Moreover,

$$\frac{\partial^2 \left(s_y^{i,j}\right)^*}{\partial \tau_\omega \partial \omega_i} = -\alpha \left(\alpha-1\right) \left(\omega_i \left(1-\tau_\omega\right)\right)^{\alpha-2} \frac{\left[1 - \left(\beta^j \delta\right)^{-\frac{2}{\sigma}} M^{\frac{\sigma-2}{\sigma}} \left(1+K\right)^{-1}\right]}{R} > 0$$

In part *(iii)* we show that two opposite effects determine how savings (A26) vary with τ_K:

$$\frac{\partial \left(s_y^{i,j}\right)^*}{\partial \tau_K} = \underbrace{\frac{\left(\beta^j \delta M\right)^{-\frac{2}{\sigma}}}{R} \left(\left(\omega_i (1-\tau_\omega)\right)^\alpha M (1+K)^{-1} \frac{\partial K}{\partial \tau_K} + \frac{P}{R} \frac{\partial R}{\partial \tau_K} \right)}_{\textit{Income Effect (I)}} +$$

$$- \underbrace{\frac{r\left(\beta^j \delta M\right)^{-\frac{2}{\sigma}}}{\sigma R} \left(\frac{2P}{M} + \left(\omega_i (1-\tau_\omega)\right)^\alpha (1+K)^{-1} (2-\sigma) \right) - \frac{1 - \left(\beta^j \delta M\right)^{-\frac{2}{\sigma}} M (1+K)^{-1}}{R} \frac{\partial R}{\partial \tau_K}}_{\textit{Substitution Effect (II)}}$$

If $II > I$, (*resp.* $I > II$), saving decrease (*resp.* increase) with τ_K. Only a further specification of the parameters of our model (but with a substantial loss of generality) will determine which effect prevails.

For part (*iv*), we have that:

$$\frac{\partial \left(s_y^{i,j}\right)^*}{\partial \tau_K \partial \omega_i} = \frac{\left(\beta^j \delta M\right)^{-\frac{2}{\sigma}}}{R} (1-\tau_\omega)^\alpha (1+K)^{-1} \left[M \frac{\partial K}{\partial \tau_K} - \frac{r(2-\sigma)}{\sigma} \right]$$

The whole derivative has the same sign as the term in square brackets: if the income effect outweighs the substitution effect, then the first term is higher than the second, and $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial\tau_K\partial\omega_i} > 0$; otherwise, $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial\tau_K\partial\omega_i} < 0$.

In part *(v)*, we show that saving are increasing in the short term discount factor, *i.e. sign* $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial\beta^j} > 0$. To prove that, notice that (A26) can be written as:

$$\left(s_y^{i,j}\right)^* = \frac{a+b}{c}$$

where:

$$\begin{aligned} a &\equiv \left(\omega_i(1-\tau_\omega)\right)^\alpha \left[1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}} M^{\frac{\sigma-2}{\sigma}} (1+K)^{-1}\right] \\ b &\equiv -\left(\beta^j\delta M\right)^{-\frac{2}{\sigma}} P \\ c &\equiv R \end{aligned}$$

A sufficient (but not necessary) condition for $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial\beta}$ to be positive is that both a and b are increasing with β^j, given that we already show (see section A.2) that $\frac{\partial R}{\partial\beta^j} < 0$. Differentiation of the two terms yields to:

$$\begin{aligned} \frac{\partial a}{\partial\beta^j} &= \left(\omega_i(1-\tau_\omega)\right)^\alpha \left[\frac{\delta\left(\beta^j\delta\right)^{-\frac{2}{\sigma}-1} M^{\frac{\sigma-2}{\sigma}} (2+K)}{\sigma(1+K)^2}\right] > 0 \\ \frac{\partial b}{\partial\beta^j} &= \frac{2}{\sigma}\left(\beta^j\delta M\right)^{-\frac{2+\sigma}{\sigma}} > 0 \end{aligned}$$

It follows that $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial\beta^j} > 0$.

In *(vi)*, we first show that increasing the labor income tax lead to a greater reduction in saving for hyperbolic individuals:

$$\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial\tau_\omega\partial\beta^j} = -\alpha\left(\omega_i(1-\tau_\omega)\right)^{\alpha-1} sign\left(\frac{\partial}{\partial\beta^j}\left[\frac{1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}} M^{\frac{\sigma-2}{\sigma}} (1+K)^{-1}}{R}\right]\right)$$

The numerator of term in square brackets is increasing with β^j (see *(v)*): it follows that the $\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial\tau_\omega\partial\beta} < 0$.

Secondly, we show that, no matter which effect prevails, increasing τ_K has a greater effect on exponential saving: $\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial\tau_K\partial\beta^j} > 0$. To prove that, we use the fact that $\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial\tau_K\partial\beta^j}$ has the same sign as $\frac{\partial^2\left(s_y^{i,j}\right)^*}{\partial\beta^j\partial\tau_K}$. Moreover, by applying the the same decomposition of *(v)*, this reduces to check the signs of $\frac{\partial^2 a}{\partial\beta^j\partial\tau_K}$ and $\frac{\partial^2 b}{\partial\beta^j\partial\tau_K}$. A sufficient condition for the whole derivative to be positive, is that both derivatives are positive[38]. After some algebra, we get:

$$\begin{aligned} \frac{\partial^2 a}{\partial\beta^j\partial\tau_K} = & \frac{\frac{\delta r}{\sigma^2} M^{\frac{\sigma-2}{\sigma}}\left(2+M^{\frac{\sigma-2}{\sigma}}\left(\beta^j\delta\right)^{-\frac{1}{\sigma}}\right)\left(\beta^j\delta\right)^{-\frac{2}{\sigma}-1}}{(1+K)^3} * \\ & * \left[2-\sigma+M^{\frac{\sigma-1}{\sigma}}\left(\sigma\left(2-\left(\beta^j\delta\right)^{-\frac{1}{\sigma}}\right)-2\left(1-\left(\beta^j\delta\right)^{-\frac{1}{\sigma}}\right)\right)\right] + W \end{aligned}$$

[38] We already show that the denominator of (A26), R, is decreasing with τ_K: $\frac{\partial^2 R}{\partial\beta^j\partial\tau_K} < 0$.

where $W = \delta r \frac{1-\sigma}{\sigma^2} \left(\beta^j \delta\right)^{-\frac{3}{\sigma}-1} M^{\frac{2\sigma-3}{\sigma}} (1+K) > 0$. It follows that:

$$\frac{\partial^2 a}{\partial \beta^j \partial \tau_K} > 0 \Leftrightarrow \sigma \left(2 - \left(\beta^j \delta\right)^{-\frac{1}{\sigma}}\right) > 2\left(1 - \left(\beta^j \delta\right)^{-\frac{1}{\sigma}}\right) \Rightarrow \sigma > \frac{2\left(1 - \left(\beta^j \delta\right)^{-\frac{1}{\sigma}}\right)}{\left(2 - \left(\beta^j \delta\right)^{-\frac{1}{\sigma}}\right)}$$

The right hand side of this expression is negative, given that $\left(\beta^j \delta\right)^{-\frac{1}{\sigma}} < 1$. Since by assumption $\sigma \in (0,1)$, $\frac{\partial^2 a}{\partial \beta^j \partial \tau_K} > 0$.

Finally,

$$\frac{\partial^2 b}{\partial \beta^j \partial \tau_K} = \frac{2(2+\sigma)}{\sigma^2} \left(\beta^j \delta M\right)^{-\frac{\sigma+2}{\sigma}-1} r > 0$$

We conclude that $\frac{\partial^2 \left(s_y^{i,j}\right)^*}{\partial \beta^j \partial \tau_K} > 0$.

A.4 Proof of Corollary 1

In part *(v)* of Proposition 1 we show that $\left(s_y^{i,TC}\right)^* > \left(s_y^{i,TI}\right)^*$. To show the Pareto-improvement, we use the following notation. Let $U^y\left(s_y^{i,j}\right)$ the utility of a young individual of type j, given his saving decision, $s_y^{i,j}$. We want to show that:

$$f(\beta) = U^y\left(s_y^{i,TC}\right) - U^y\left(s_y^{i,TI}(\beta)\right) > 0$$

Note that $U^y\left(s_y^{i,TI}(\beta)\right)$ depends on β in two ways: from one hand, β is the degree at which future utility is discounted, so changes in β affect future consumption. Moreover, β enters directly into the expression that determines $\left(s_y^{i,TI}\right)$ itself. For exponential individuals, note that β only influences $U^y\left(s_y^{i,TC}\right)$ through the first mechanism, as $\left(s_y^{i,TC}\right)$ does not depend on β. In the proof, we characterize the value of $f(.)$ in a neighborhood of $\beta = 1$. First, note that $f(1) = 0$, since $\left(s_y^{i,TC}\right) = \left(s_y^{i,TI}\right)$. In order to evaluate the marginal variation of $f(.)$ around $\beta = 1$, we consider $f"(1)$ and $f'(1)$; we have:

$$f'(\beta) = \frac{\partial U^y\left(s_y^{i,TC}\right)}{\partial \beta} - \frac{\partial U^y\left(s_y^{i,TI}\right)}{\partial \beta} - \frac{\partial U^y\left(s_y^{i,TI}\right)}{\partial s_y^{i,TI}} \frac{d s_y^{i,TI}}{d\beta}$$

Note that $f'(1) = 0$, as $\left.\frac{\partial U^y\left(s_y^{i,TC}\right)}{\partial \beta}\right|_{\beta=1} = \left.\frac{\partial U^y\left(s_y^{i,TI}\right)}{\partial \beta}\right|_{\beta=1}$ and $\left.\frac{\partial U^y\left(s_y^{i,TI}\right)}{\partial s_y^{i,TI}}\right|_{\beta=1} = \left.\frac{\partial U^y\left(s_y^{i,TC}\right)}{\partial s_y^{i,TC}}\right|_{\beta=1} = 0$. It is possible to show that $f"(1) > 0$. Given that $f(1) = 0$, $f'(1) = 0$ and $f"(1) > 0$, there exists an interval $\left(\bar{\beta}, 1\right)$ such that $f(\beta) > 0$ for $\beta \in \left(\bar{\beta}, 1\right)$. This shows that a sophisticated individual is made better off by saving up to $s_y^{i,TC}$. Pareto dominance follows from the fact that an individual is made better off in two ways: first, they prefer to save $s_y^{i,TC}$ that increases consumption when middle age and, second, they would gain a higher transfer P when old, for a giving τ_K, as implied by $s_y^{i,TC}$.

A.5 Proof of Proposition 2

Part *(i)* of the proposition is straightforward: from (A27), we see that $\frac{\partial \left(s_{ma}^{i,j}\right)^*}{\partial \omega_i} > 0$.

For part *(ii)*, differentiating (A27) with respect to τ_ω gives us:

$$\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\tau_\omega} = -\alpha\left(\omega_i\left(1-\tau_\omega\right)\right)^{\alpha-1}\frac{\left[R+M\left(1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}}M^{\frac{\sigma-2}{\sigma}}\left(1+K\right)^{-1}\right)\right]}{R(1+K)} < 0 \tag{A28}$$

$$\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_\omega\partial\omega_i} = -\alpha\left(\alpha-1\right)\left(\omega_i(1-\tau_\omega)\right)^{\alpha-2}\frac{\left[R+M\left(1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}}M^{\frac{\sigma-2}{\sigma}}\left(1+K\right)^{-1}\right)\right]}{R(1+K)} > 0$$

In part (iii), we compute the marginal variation of saving in response to an increase of the capital income tax. To do that, let us rewrite (A27) as follows:

$$\left(s_{ma}^{i,j}\right)^* = \underbrace{\frac{\left(\omega_i(1-\tau_\omega)\right)^\alpha}{(1+K)}}_{I}\underbrace{\left[1+\frac{M-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}}M^{\frac{\sigma-2}{\sigma}}(1+K)^{-1}}{R}\right]}_{II}\underbrace{-\frac{\left(\beta^j\delta M\right)^{-\frac{1}{\sigma}}P}{(1+K)}}_{III}\underbrace{\left(1+\frac{K}{R}\right)}_{IV}$$

It follows that $sign\left(\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K}\right) = sign\left(\frac{\partial I}{\partial\tau_K}II+\frac{\partial II}{\partial\tau_K}I+\frac{\partial III}{\partial\tau_K}IV+\frac{\partial IV}{\partial\tau_K}III\right)$. Giving that I, II, III, $IV>0$, and computing all the derivatives, it emerges that two opposite effects determine the sign of the derivative:

Substitution effect <0

$$-\left(\omega_i(1-\tau_\omega)^\alpha\left(1+K\right)^{-2}\frac{\partial K}{\partial\tau_K}\right)II+$$
$$-\left[r\left(1+\left(\beta^j\delta M\right)^{-\frac{2}{\sigma}}\left(1+K\right)^{-1}\frac{(2-\sigma)}{\sigma}\right)+\frac{\left(M-\left(\beta^j\delta M\right)^{-\frac{2}{\sigma}-1}(1+K)^{-1}\right)\frac{\partial R}{\partial\tau_K}}{R}\right]\frac{I}{R}+$$
$$-\left(\frac{1}{\sigma}r(1+K)^{-1}P\left(\beta^j\delta M\right)^{-\frac{2}{\sigma}}\right)IV+\frac{\partial K}{\partial\tau_k}\frac{1}{R}III$$

Income Effect >0

$$\left[\left(\beta^j\delta M\right)^{-\frac{2}{\sigma}+1}\left(1+K\right)^{-2}\frac{\partial K}{\partial\tau_K}\frac{1}{R}\right]I+$$
$$+\left[P\left(\beta^j\delta M\right)^{-\frac{1}{\sigma}}\left(1+K\right)^{-1}IV\right]+\left[\frac{\partial R}{\partial\tau_K}\frac{K}{R^2}III\right]>0$$

As for Proposition 1, a priori, we do not know which effect prevails.

For part (iv), we have that:

$$\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K\partial\omega_i} = \frac{\partial\left(\mathit{Income\ Effect}\right)}{\partial\omega_i}+\frac{\partial\left(\mathit{Substitution\ Effect}\right)}{\partial\omega_i}$$

The relative magnitude of the two terms determines the sign of the derivative: if the substitution effect prevails over the income effect, the first term is higher than the second, and $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial\tau_K\partial\omega_i}<0$; otherwise, $\frac{\partial\left(s_y^{i,j}\right)^*}{\partial\tau_K\partial\omega_i}>0$.

To determine the sign of $\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\beta^j}$(part v), we rewrite (A27) as follows

$$\left(s_{ma}^{i,j}\right)^* = a-bc \tag{A29}$$

where:

$$a \equiv (\omega_i(1-\tau_\omega))^\alpha \left[\frac{1}{1+K} + \frac{M\left(1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}} M^{\frac{\sigma-2}{\sigma}} (1+K)^{-1}\right)}{R(1+K)}\right]$$
$$b \equiv \left(\beta^j\delta M\right)^{-\frac{2}{\sigma}} P$$
$$c \equiv \frac{K+R}{R(1+K)}$$

Differentiation of (**??**) with respect to β^j yields to:

$$\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\beta^j} = \frac{\partial(a)}{\partial\beta^j} - \left[\frac{\partial(b)}{\partial\beta^j}c + b\frac{\partial(c)}{\partial\beta^j}\right]$$

We already show that $\frac{\partial(a)}{\partial\beta^j} > 0$ and $\frac{\partial(b)}{\partial\beta^j} > 0$ (see part (*iv*) of Proposition 1). Moreover, differentiating *(c)* with respect to β_j gives:

$$\frac{\partial(c)}{\partial\beta_j} = -\frac{K^2\frac{\partial R}{\partial\beta^j} + R^2\frac{\partial K}{\partial\beta^j}}{\left(R(1+K)\right)^2} > 0$$

The sign of the whole derivative is then undetermined, as it depends on the relative magnitude of two opposite effects:

$$\frac{\partial\left(s_{ma}^{i,j}\right)^*}{\partial\beta^j} = \underbrace{-b\frac{\partial(c)}{\partial\beta^j}}_{Catching\ Up\ Effect\ (<0)} + \underbrace{\frac{\partial(a)}{\partial\beta^j} + \frac{\partial(b)}{\partial\beta^j}c}_{Exponential\ Effect\ (>0)}$$

In *(vi)*, we show that $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_\omega\partial\beta^j} < 0$. By differentiating (A28), we obtain:

$$\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_\omega\partial\beta^j} = -\alpha\left(\omega_i(1-\tau_\omega)\right)^{\alpha-1}\frac{\partial}{\partial\beta^j}\left[\frac{R+M\left(1-\left(\beta^j\delta\right)^{-\frac{2}{\sigma}} M^{\frac{\sigma-2}{\sigma}} (1+K)^{-1}\right)}{R(1+K)}\right] < 0$$

The derivative has the same sign as the term in square brackets, which coincides with term a of (v)). We already know that $\frac{\partial a}{\partial\beta^j} > 0$.

In the second part of (vi), we show that $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K\partial\beta^j} > 0$. To show that, we proceed in two steps; first, notice that $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K\partial\beta^j}$ has the same sign as $\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\beta^j\partial\tau_K}$. Second, by applying the same decomposition of (A29) and differentiating, we obtain:

$$\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\beta^j\partial\tau_K} = \frac{\partial^2(a)}{\partial\beta^j\partial\tau_K} - \frac{\partial^2(b)}{\partial\beta^j\partial\tau_K}c - \frac{\partial^2(c)}{\partial\beta^j\partial\tau_K}b \tag{A30}$$

A sufficient but not necessary condition for the derivative to be positive is that the first term on the RHS of (A30) is positive, and the other terms are negative. We know from Proposition 1 that $sign\frac{\partial^2(a)}{\partial\beta^j\partial\tau_K} > 0$ and $sign\frac{\partial^2(b)}{\partial\beta^j\partial\tau_K} < 0$. Moreover,

$$\frac{\partial^2(c)}{\partial\beta^j\partial\tau_K} = -\frac{1}{\sigma\beta^j}\frac{\partial K}{\partial\tau_K}\left[K^7+8K^6+9K^5-2K^4-10K^3-6K^2-2K-1\right] < 0 \Leftrightarrow K > 1$$

Since $K > 1$, we have that $sign\frac{\partial^2\left(s_{ma}^{i,j}\right)^*}{\partial\tau_K\partial\beta^j} > 0$.

A.6 Proof of Corollary 2

In part *(v)* of Proposition 2 we show that, if the *exponential* effect overcome the *catching up* effect, $\left(s_{ma}^{i,TC}\right)^* > \left(s_{ma}^{i,TI}\right)^*$. To show the Pareto-improvement, let us define $U^{ma}\left(s_{ma}^{i,j}\right)$ as the utility of a middle aged individual of behavioral type j, given its choice about of saving, $\left(s_{ma}^{i,j}\right)$. We want to show that:

$$f(\beta) = U^{ma}\left(s_{ma}^{i,TC}\right) - U^{ma}\left(s_{ma}^{i,TI}(\beta)\right) > 0$$

Note that $U^{ma}\left(s_{ma}^{i,TI}(\beta)\right)$ depends on β in two ways: from one hand, β is the degree at which future utility is discounted, so changes in β affect future consumption. Moreover, β enters directly into the expression that determines $\left(s_{ma}^{i,TI}\right)$ itself. For exponential individuals, note that β only influences $U^{y}\left(s_{ma}^{i,TC}\right)$ through the first mechanism, as $\left(s_{ma}^{i,TC}\right)$ does not depend on β. In the proof, we characterize the value of $f(.)$ in a neighborhood of $\beta = 1$. First, note that $f(1) = 0$, since $\left(s_{ma}^{i,TC}\right) = \left(s_{ma}^{i,TI}\right)$. In order to evaluate the marginal variation of $f(.)$ around $\beta = 1$, we consider $f"(1)$ and $f'(1)$; we have:

$$f'(\beta) = \frac{\partial U^{ma}\left(s_{ma}^{i,TC}\right)}{\partial \beta} - \frac{\partial U^{ma}\left(s_{ma}^{i,TI}\right)}{\partial \beta} - \frac{\partial U^{ma}\left(s_{ma}^{i,TI}\right)}{\partial s_{ma}^{i,TI}}\frac{ds_{ma}^{i,TI}}{d\beta}$$

Note that $f'(1) = 0$, as $\left.\frac{\partial U^{ma}\left(s_{ma}^{i,TC}\right)}{\partial \beta}\right|_{\beta=1} = \left.\frac{\partial U^{ma}\left(s_{ma}^{i,TI}\right)}{\partial \beta}\right|_{\beta=1}$ and $\left.\frac{\partial U^{y}\left(s_{ma}^{i,TI}\right)}{\partial s_{ma}^{i,TI}}\right|_{\beta=1} = \left.\frac{\partial U^{ma}\left(s_{ma}^{i,TC}\right)}{\partial s_{ma}^{i,TC}}\right|_{\beta=1} = 0$. Finally, it is possible to show that $f"(1) > 0$. Given that $f(1) = 0$, $f'(1) = 0$ and $f"(1) > 0$, there exists an interval $\left(\bar{\beta}, 1\right)$ such that $f(\beta) > 0$ for $\beta \in \left(\bar{\beta}, 1\right)$. This shows that a hyperbolic individual is made better off by saving up to $s_{ma}^{i,TC}$. Pareto dominance follows from the fact that an individual is made better off in two ways: first, they prefer to save $s_{ma}^{i,TC}$ that increases consumption when middle age and, second, they would gain a higher transfer P when old, for a giving τ_K, as implied by $s_{ma}^{i,TC}$. If, on the other hand, the catching up effect prevails, $\left(s_{ma}^{i,TI}\right)^* > \left(s_{ma}^{i,TC}\right)^*$: in this case, the Pareto improvement depends on whether $\left(s_{ma}^{i,TI}\right)^*$ is the optimal level of saving.

A.7 The Transfer P

The equilibrium transfer P, financed through the proceeds of the two income taxes, depends on individuals' choices (labor supply and saving) and satisfies the following budget constraint:

$$P = \frac{1}{n^o}\left\{\tau_\omega[\omega^P L^P + \omega^R L^R] + \tau_K r\left[\lambda^{TI} S_{TI} + \lambda^{TC} S_{TC}\right]\right\} \tag{A31}$$

where $L^i = \rho^i\left(n^y l_y^i + n^{ma} l_{ma}^i\right)$ denotes the total labor supply by young and middle aged for a given ω and $S_{TI} = \sum_{i=R,P} \rho^i\left(n^y s_y^{TI,i} + n^{ma} s_{ma}^{TI,i}\right)$ and $S_{TC} = \sum_{i=R,P} \rho^i\left(n^y s_y^{TC,i} + n^{ma} s_{ma}^{TC,i}\right)$ represent total saving for time inconsistent and far-sighted individuals. Replacing into (A32) the expression for labor supply and saving, and rearranging, we have:

$$P_{eq}(\tau_\omega, \tau_K) = \tag{A32}$$
$$= \frac{1}{n^o} \frac{(1-\tau_\omega)^\alpha (n^y+n^{ma})(\rho_R\omega_R^\alpha+\rho_P\omega_P^\alpha)\left\{\tau_\omega(\rho_R\omega_R+\rho_P\omega_P)+\frac{\tau_K r(\delta M)^{-\frac{1}{\sigma}}}{R}\left[M\left(1-\delta^{-\frac{2}{\sigma}}\left(\frac{\lambda^{TI}\beta^{-\frac{2}{\sigma}}}{1-K\beta^{-\frac{1}{\sigma}}}+\frac{\lambda^{TC}}{1-K}\right)\left(M^{-\frac{2}{\sigma}}+M^{\frac{2+\sigma}{\sigma}}\right)\right)+R\right]\right\}}{1+\frac{\tau_K r(\delta M)^{-\frac{1}{\sigma}}}{R}\left(\lambda^{TI}\left((\beta^{-\frac{1}{\sigma}}\delta M)^{-2}+\beta^{-\frac{1}{\sigma}}(R+K\beta^{-\frac{1}{\sigma}})\right)+\lambda^{TC}(R+K+(\delta M)^{-2})\right)}$$

A.8 Proof of Proposition 3

To show that $P_{eq}(\tau_\omega, \tau_K)$ is decreasing in τ_ω (part *(i)* of the Proposition), we differentiate (A32) with respect to τ_ω:

$$\frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_\omega} = -\frac{(n^y + n^{ma})}{n^o} \frac{\alpha(1-\tau_\omega)^{\alpha-1}}{J} \left[\left(\rho_R\omega_R^{\alpha+1} + \rho_P\omega_P^{\alpha+1}\right)(\alpha\tau_\omega(1-\tau_\omega)-1) - \alpha\left(\rho_R\omega_R^\alpha + \rho_P\omega_P^\alpha\right)F\right] \tag{A33}$$

where:

$$F = \frac{\tau_K r(\delta M)^{-\frac{1}{\sigma}}}{R}\left[M\left(1-\delta^{-\frac{2}{\sigma}}\left(\frac{\lambda^{TI}\beta^{-\frac{2}{\sigma}}}{1-K\beta^{-\frac{1}{\sigma}}}+\frac{\lambda^{TC}}{1-K}\right)\left(M^{-\frac{2}{\sigma}}+M^{\frac{2+\sigma}{\sigma}}\right)\right)+R\right] > 0$$
$$J = 1 + \frac{\tau_K r(\delta M)^{-\frac{1}{\sigma}}}{R}\left(\lambda^{TI}\left((\beta^{-\frac{1}{\sigma}}\delta M)^{-2}+\beta^{-\frac{1}{\sigma}}(R+K\beta^{-\frac{1}{\sigma}})\right)+\lambda^{TC}\left(R+K+(\delta M)^{-2}\right)\right) > 0$$

are constant terms that do not depend on τ_ω.

Expression (A33), has not a clear sign; however, it is possible to show that it exists a threshold level $\tilde{\tau}_\omega$ such that $\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial\tau_\omega} > 0$ for $0 \le \tau_\omega \le \tilde{\tau}_\omega$ and $\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial\tau_\omega} < 0$ otherwise. In this interval, moreover, it is immediate to see that $\frac{\partial^2 P_{eq}(\tau_\omega,\tau_K)}{\partial\tau_\omega\partial\omega_i} < 0$.

Comparative statics on $P_{eq}(\tau_\omega, \tau_K)$ with respect to τ_K (part *(ii)*), lead to the following results:

$$\frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K} = \tau_K r\left[\lambda^{TI}\frac{\partial S_{TI}}{\partial \tau_K} + \lambda^{TC}\frac{\partial S_{TC}}{\partial \tau_K}\right] + r\left[\lambda^{TI}S_{TI} + \lambda^{TC}S_{TC}\right]$$

The sign of the derivative depends on the sign of the term in square brackets: if the substitution effect is greater than the income effect, both $\frac{\partial S_{TI}}{\partial\tau_K}$ and $\frac{\partial S_{TC}}{\partial\tau_K}$ are positive: the equilibrium transfer is linear and increasing in τ_K (Figure 6a). If the income effect prevails, then $P_{eq}(\tau_\omega, \tau_K)$ is concave in τ_K: it exist a threshold value for the tax rate, $\tilde{\tau}_K$, suche that the transfer increases up to $\tilde{\tau}_K$, and decreses thereafter. Moreover, by Propositions 1*(iii)* and 2*(iii)*, it is immediate to see that, whichever effect prevails, $\left|\frac{\partial S_{TI}}{\partial\tau_K}\right| < \left|\frac{\partial S_{TC}}{\partial\tau_K}\right|$, *i.e.* the variation of individuals' saving in response to a variation of τ_K is bigger for time consistent agents.

A.9 Proof of Proposition 4

Each group of individual set labor income tax rates according to:

$$\frac{\partial V_o^{i,j}}{\partial \tau_\omega} : u'\left(c_o^{i,j}\right)\frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_\omega} = 0 \tag{A34}$$

$$\frac{\partial V_{ma}^{i,TI}}{\partial \tau_\omega} : u'\left(c_{ma}^{i,TI}\right)\omega_i l_{ma}^*(\omega_i) = \beta\delta\left(u'\left(c_o^{i,TI}\right)\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega}\right) \tag{A35}$$
$$\frac{\partial V_{ma}^{i,TC}}{\partial \tau_\omega} : u'\left(c_{ma}^{i,TC}\right)\omega_i l_{ma}^*(\omega_i) = \delta\left(u'\left(c_o^{i,TC}\right)\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega}\right)$$

$$\frac{\partial V_{y}^{i,TI}}{\partial \tau_\omega} : u'\left(c_{y}^{i,TI}\right)\omega_i l_{y}^*(\omega_i) = \beta\delta\left[-u'\left(c_{ma}^{i,TI}\right)\omega_i l_{ma}^*(\omega_i) + \delta\left(u'\left(c_o^{i,TI}\right)\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega}\right)\right] \tag{A36}$$
$$\frac{\partial V_{y}^{i,TC}}{\partial \tau_\omega} : u'\left(c_{y}^{i,TC}\right)\omega_i l_{y}^*(\omega_i) = \delta\left[-u'\left(c_{ma}^{i,TC}\right)\omega_i l_{ma}^*(\omega_i) + \delta\left(u'\left(c_o^{i,TC}\right)\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega}\right)\right]$$

In part *(i)* of the Proposition we show that, for a giving degree of time inconsistency, labor income tax rates are increasing with income. To see that, let us consider, without loss of generality, a middle aged individual of type i, j; by the implicit function theorem:

$$sign\left(\frac{\partial \tau_\omega}{\partial \omega_i}\right) = -\frac{\partial^2 V_{ma}^{i,j}/\partial \tau_\omega \partial \omega_i}{\partial^2 V_{ma}^{i,j}/\partial \tau_\omega^2}$$

Assuming that the SOC are satisfied, $\partial \tau_\omega/\partial \omega_i$ has the same sign of $\partial^2 V_{ma}^{i,j}/\partial \tau_\omega \partial \omega_i$. Taking into account that an interior solution for $s_{ma}^{i,j}$ implies that:

$$-u'\left(c_{ma}^{i,j}\right) + \delta(1 + r(1-\tau_K))u'\left(c_o^{i,j}\right) = 0$$

The FOC can be rewritten as follows:

$$\frac{\partial V_{ma}^{i,j}}{\partial \tau_\omega} : \delta u'\left(c_o^{i,j}\right)\left(-(1+r(1-\tau_K))\omega_i l_{ma}^*(\omega_i) + \beta^j\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega}\right) = 0 \tag{A37}$$

Therefore,

$$\frac{\partial^2 V_{ma}^{i,j}}{\partial \tau_\omega \partial \omega_i} = \delta u'\left(c_o^{i,j}\right)\left[-(1+r)(1-\tau_K)\left(1+\frac{\alpha}{\omega_i(1-\tau_\omega)}\right)l_{ma}^*(\omega_i) + \beta^j\frac{\partial^2 P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega \partial \omega_i}\right] +$$
$$+\,\delta u''\left(c_o^{i,j}\right)\left(\frac{\partial c_o^{i,j}}{\partial \omega_i}\right)\left[-(1+r)(1-\tau_K)\omega_i l_{ma}^* + \frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega}\right] < 0$$

We already show that $\frac{\partial^2 P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega \partial \omega_i} < 0$ and $\left(-(1+r)(1-\tau_K)\omega_i l_{ma}^* + \frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega}\right)$ has to be positive in order to have an interior solution for (A37).

In *(ii)*, we show that all old prefer $\tilde{\tau}_\omega$. This follows from (A34): the FOC is satisfied if $\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega} = 0$, or $\tau_\omega^o(\beta^j,\omega_i) = \tilde{\tau}_\omega$, $\forall i, j$, *i.e.* old maximize their consumption level by setting the tax rate that maximizes the pension function[39]¿

Part *(iii)* can be proved in the following way; the implicit function theorem implies that:

$$sign\left(\frac{\partial \tau_\omega}{\partial \beta^j}\right) = -\frac{\partial^2 V_{ma}^{i,j}/\partial \tau_\omega \partial \beta^j}{\partial^2 V_{ma}^{i,j}/\partial \tau_\omega^2}$$

[39] Also $\tau_\omega = 1$ implies $\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial \tau_\omega} = 0$, but this is a *minimum* and not a maximum.

Therefore,

$$\frac{\partial^2 V_{ma}^{i,j}}{\partial \tau_\omega \partial \beta^j} = \delta u'\left(c_o^{i,j}\right)\frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_\omega} > 0$$

Parts *(iv)* and *(v)* follow immediately.

A.10 Proof of Proposition 5

Each group of individuals of type $i = R, P$ and belonging to generation $g = y, ma, o$ set taxes, denoted by $\tau_K^g(\beta^j, \omega_i)$, such that:

$$\frac{\partial V_o^{i,TI}}{\partial \tau_K} : u'(c_o^{i,TI})\left(-rs_{ma}^{i,TI} + \frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K}\right) = 0 \tag{A38}$$
$$\frac{\partial V_o^{i,TC}}{\partial \tau_K} : u'(c_o^{i,TC})\left(-rs_{ma}^{i,TC} + \frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K}\right) = 0$$

$$\frac{\partial V_y^{i,TI}}{\partial \tau_K} : \beta\delta\left[-u'\left(c_{ma}^{i,TI}\right) rs_y^{i,TI} + \delta u'\left(c_o^{i,TI}\right)\left(-rs_{ma}^{i,TI} + \frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K}\right)\right] = 0 \tag{A39}$$
$$\frac{\partial V_y^{i,TC}}{\partial \tau_K} : \delta\left[-u'\left(c_{ma}^{i,TC}\right) rs_y^{i,TC} + \delta u'\left(c_o^{i,TC}\right)\left(-rs_{ma}^{i,TC} + \frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K}\right)\right] = 0$$

$$\frac{\partial V_{ma}^{i,TI}}{\partial \tau_K} : -u'\left(c_{ma}^{i,TI}\right) rs_y^{i,TI} + \beta\delta u'\left(c_o^{i,TI}\right)\left(-rs_{ma}^{i,TI} + \frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K}\right) = 0 \tag{A40}$$
$$\frac{\partial V_{ma}^{i,TC}}{\partial \tau_K} : -u'\left(c_{ma}^{i,TC}\right) rs_y^{i,TC} + \delta u'\left(c_o^{i,TC}\right)\left(-rs_{ma}^{i,TC} + \frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K}\right) = 0$$

In part *(i)*, we claim that tax rates are decreasing with ω. For old, simply notice that from Proposition 3 we know that $\forall i$, $s_{ma}^{P,j} < s_{ma}^{R,j}$, for a given j. From (A38), we have that: $rs_{ma}^{i,TC} = \frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K}$. Moreover, the concavity of the pension function implies that:

$$\tilde{\tau}_K > \tau_K^o(\beta^j, \omega_P) > \tau_K^o(\beta^j, \omega_R)$$

Comparing (A39) and (A40), it is immediate to see that the problem is very similar for middle aged and young: without loss of generality, we are going to consider only middle aged agents. The implicit function theorem implies that:

$$sign\left(\frac{\partial \tau_K}{\partial \omega_i}\right) = -\frac{\partial^2 V_{ma}^{i,j}/\partial \tau_K \partial \omega_i}{\partial^2 V_{ma}^{i,j}/\partial \tau_K^2}$$

Assuming that the SOC are satisfied, $\partial \tau_\omega/\partial \omega_i$ has the same sign of $\partial^2 V_{ma}^{i,j}/\partial \tau_\omega \partial \omega_i$. If saving are positive for every agent, the FOC can be rewritten as:

$$\frac{\partial V_{ma}^{i,j}}{\partial \tau_K} : \delta u'\left(c_o^{i,j}\right)\left(-(1 + r(1-\tau_K))rs_y^{i,j} + \beta^j\left(-rs_{ma}^{i,j} + \frac{\partial P_{eq}(\tau_\omega, \tau_K)}{\partial \tau_K}\right)\right) = 0 \tag{A41}$$

then we have:

$$\frac{\partial^2 V_{ma}^{i,j}}{\partial\tau_K\partial\omega_i} = \delta u'\left(c_o^{i,j}\right)\left[-(1+r(1-\tau_K))r\frac{\partial s_y^{i,j}}{\partial\omega_i} + \beta^j\left(-r\frac{\partial s_y^{i,j}}{\partial\omega_i} + \frac{\partial^2 P_{eq}(\tau_\omega,\tau_K)}{\partial\tau_K\partial\omega_i}\right)\right] +$$
$$+\delta u''\left(c_o^{i,j}\right)\left(\frac{\partial c_o^{i,j}}{\partial\omega_i}\right)\left[-(1+r(1-\tau_K))rs_y^{i,j} + \beta^j\left(-rs_{ma}^{i,j} + \frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial\tau_K}\right)\right] < 0$$

The sign of the equation follows from the fact that individual saving are increasing with productivity and $\frac{\partial^2 P_{eq}(\tau_\omega,\tau_K)}{\partial\tau_K\partial\omega_i} < 0$.

In *(ii)* we show that preferred τ_K are decreasing with β^j. For old agents, notice that Proposition 3 assures that $\forall i$, $s_{ma}^{i,TI} < s_{ma}^{i,TC}$, for given i. The FOC (A38) implies that: $rs_{ma}^{i,TC} = \frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial\tau_K}$. Concavity of P ensures: $\tau_K^o(\beta^{TI},\omega_i) > \tau_K^o(\beta^{TC},\omega_i)$, for $i = P, R$. If there is enough inequality in the economy,far-sighted agents save less than hyperbolic rich: $s_{ma}^{R,TI} < s_{ma}^{P,TC}$. It follows that preferred tax rates can be ranked as follows:

$$\tilde{\tau}_K > \tau_K^o(\beta^{TI},\omega_P) > \tau_K^o(\beta^{TC},\omega_P) > \tau_K^o(\beta^{TI},\omega_R) > \tau_K^o(\beta^{TC},\omega_R)$$

For middle aged and young, the implicit function theorem implies that:

$$sign\left(\frac{\partial\tau_K}{\partial\beta^j}\right) = -\frac{\partial^2 V_{ma}^{i,j}/\partial\tau_K\partial\beta^j}{\partial^2 V_{ma}^{i,j}/\partial\tau_K^2}$$

Assuming that the SOC are satisfied, $\partial\tau_\omega/\partial\beta^j$ has the same sign of $\partial^2 V_{ma}^{i,j}/\partial\tau_\omega\partial\beta^j$, namely:

$$\frac{\partial^2 V_{ma}^{i,j}}{\partial\tau_K\partial\beta^j} = -u'\left(c_{ma}^{i,TI}\right)r\frac{\partial s_y^{i,TI}}{\partial\beta^j} - \delta u'\left(c_o^{i,TI}\right)rs_{ma}^{i,TI} - \beta^j\delta u'\left(c_o^{i,TI}\right)r\frac{\partial s_{ma}^{i,TI}}{\partial\beta^j}\frac{\partial P_{eq}(\tau_\omega,\tau_K)}{\partial\tau_K} < 0 \quad \text{(A42)}$$

In *(iii)* we show that, for a given productivity level and β^j, preferred τ_K for old are greater than those for young and middle aged. Comparing (A38) and (A40) we notice that FOC for middle aged has an additional negative term compared to the FOC for old. It follows that $\tau^o(\beta^j,\omega_i) > \tau_K^{ma}(\beta^j,\omega_i)$. Finally, since FOCs (A39) and (A40) are the same, we have that: $\tau_K^{ma}(\beta^j,\omega_i) = \tau_K^y(\beta^j,\omega_i)$[40].

Part *(iv)* follows immediately from previous discussions.

A.11 Proof of Proposition 6

Preferred tax rates satisfies equations (A38), (A39) and (A40). Given that P is linear in τ_K, old, for any type, prefer the maximum tax rate, $\tau_K = 1$. The proof of rest of the Proposition follows exactly the proof of Proposition 5.

[40] In the proof of part iii, we have implicitly assumed that the the catching up effect is smaller than the exponential effect, *i.e.* $\frac{\partial s_{ma}^{i,TI}}{\partial\beta^j} > 0$. If this is not the case, the sign of (A42) is undetermined.

A.12 Proof of Proposition 7

First, let us consider the equilibrium labor tax, τ_ω^{eq}. Equation (17) can be rewritten as follows, for $g = y, ma, o$:

$$Nz\lambda^P \left[\rho^{TI}\phi^{TI}\sum_g \frac{dV_g^{TI,P}}{d\tau_\omega} + \rho^{TC}\phi^{TC}\sum_g \frac{dV_g^{TC,P}}{d\tau_\omega}\right] +$$
$$+N\lambda^R \left[\rho^{TI}\phi^{TI}\sum_g \frac{dV_g^{TI,R}}{d\tau_\omega} + \rho^{TC}\phi^{TC}\sum_g \frac{dV_g^{TC,R}}{d\tau_\omega}\right] = 0$$

Depending on the values of the key parameters of our model, z and ρ^{TI}, several cases are possible.

Case 1: $z = 1$ **and** $\rho^{TI} = 0$

$$\tau_\omega^y(\omega_R) < \tau_\omega^{ma}(\omega_R) < \tau_\omega^y(\omega_P) < \tau_\omega^{ma}(\omega_P) < \boldsymbol{\tau_\omega^{eq}} < \tilde{\tau}_\omega$$

Poor and old receive more weight in the party's objective function; since τ_ω increases with income, their most preferred tax rate will be the equilibrium one.

Case 2: $z < 1$ **and** $\rho^{TI} = 0$

$$\tau_\omega^y(\omega_R) < \tau_\omega^{ma}(\omega_R) < \boldsymbol{\tau_\omega^{eq}} < \tau_\omega^y(\omega_P) < \tau_\omega^{ma}(\omega_P) < \tilde{\tau}_\omega$$

If only a fraction z of poor votes, they receive less weight in the objective function: the platform is distorted towards rich's preferred labor tax, which is lower.

Case 3: $z = 1$ **and** $\rho^{TI} > 0$

$$\tau_\omega^y(\omega_R,\beta) < \tau_\omega^y(\omega_R,1) < \tau_\omega^{ma}(\omega_R,\beta) < \tau_\omega^{ma}(\omega_R,1) < \tau_\omega^y(\omega_P,\beta) <$$
$$< \boldsymbol{\tau_\omega^{eq}} < \tau_\omega^y(\omega_P,1) < \tau_\omega^{ma}(\omega_P,\beta) < \tau_\omega^{ma}(\omega_P,1) < \tilde{\tau}_\omega$$

With full turnout and time inconsistency, the equilibrium τ_ω is lower than case 1 but higher than case 2.

Case 4: $z < 1$ **and** $\rho^{TI} > 0$

$$\tau_\omega^y(\omega_R,\beta) < \tau_\omega^y(\omega_R,1) < \tau_\omega^{ma}(\omega_R,\beta) < \tau_\omega^{ma}(\omega_R,1) < \boldsymbol{\tau_\omega^{eq}} < \tau_\omega^y(\omega_P,\beta) <$$
$$< \tau_\omega^y(\omega_P,1) < \tau_\omega^{ma}(\omega_P,\beta) < \tau_\omega^{ma}(\omega_P,1) < \tilde{\tau}_\omega$$

If only a fraction z of poor votes, they receive less weight in the objective function: however, now the platform is distorted towards hyperbolic rich's preferred labor tax, which is lower.

Let us consider now the equilibrium capital tax, denoted τ_K^{eq}. Equation (18) can be rewritten as follows, for $g = y, ma, o$:

$$Nz\lambda^P \left[\rho^{TI}\phi^{TI}\sum_g \frac{dV_g^{TI,P}}{d\tau_K} + \rho^{TC}\phi^{TC}\sum_g \frac{dV_g^{TC,P}}{d\tau_K}\right] +$$
$$+N\lambda^R \left[\rho^{TI}\phi^{TI}\sum_g \frac{dV_g^{TI,R}}{d\tau_K} + \rho^{TC}\phi^{TC}\sum_g \frac{dV_g^{TC,R}}{d\tau_K}\right] = 0$$

Depending on the values of z and ρ^{TI}, several cases are possible:

Case 1: $z = 1$ **and** $\rho^{TI} = 0$

$$\bar{\tau}_K(\omega_R) < \boldsymbol{\tau_K^{eq}} < \bar{\tau}_K(\omega_P) < \tau_K^o(\omega_P) < \tau_K^o(\omega_R) < \tilde{\tau}_K$$

Poor and old are more attractive for both parties: since $\lambda^P > \lambda^R$, the equilibrium tax is the one preferred by them. Given that τ_K increases with income, their most preferred tax rate will be the equilibrium one.

Case 2: $z < 1$ **and** $\rho^{TI} = 0$

$$\boldsymbol{\tau_K^{eq}} < \bar{\tau}_K(\omega_R) < \bar{\tau}_K(\omega_P) < \tau_K^o(\omega_P) < \tau_K^o(\omega_R) < \tilde{\tau}_K$$

Poor are not decisive anymore, since $z\lambda^P < \lambda^R$: the equilibrium tax rate is the lowest possible.

Case 3: $z = 1$ **and** $\rho^{TI} > 0$

$$\bar{\tau}_K(\omega_R, 1) < \bar{\tau}_K(\omega_R, \beta) < \bar{\tau}_K(\omega_P, 1) < \bar{\tau}_K(\omega_P, \beta) < \tau_K^o(\omega_R, 1) <$$
$$< \boldsymbol{\tau_K^{eq}} < \tau_K^o(\omega_R, \beta) < \tau_K^o(\omega_P, 1) < \tau_K^o(\omega_P, \beta) < \tilde{\tau}_K$$

With full turnout, and time inconsistency, equilibrium capital taxes are higher than case 1: poor form a coalition with time inconsistent individuals, who are politically more homogeneous.

Case 4: $z < 1$ **and** $\rho^{TI} > 0$

$$\bar{\tau}_K(\omega_R, 1) < \bar{\tau}_K(\omega_R, \beta) < \boldsymbol{\tau_K^{eq}} < \bar{\tau}_K(\omega_P, 1) < \bar{\tau}_K(\omega_P, \beta)$$
$$\tau_K^o(\omega_R, 1) < \tau_K^o(\omega_R, \beta) < \tau_K^o(\omega_P, 1) < \tau_K^o(\omega_P, \beta) < \tilde{\tau}_K$$

With less than full turnout and time inconsistency, equilibrium capital taxes are lower than case 3 but higher than case 1: the fraction of politically active poor forms a coalition with time inconsistent individuals as to increase equilibrium capital taxes.