Working Paper No. 343

On the Optimal Timing of Capital Taxes

John Hassler, Per Krusell, Kjetil Storesletten and Fabrizio Zilibotti

December 2007

On the Optimal Timing of Capital Taxes

John Hassler^{a,c}†, Per Krusell^{b,a,c,d}, Kjetil Storesletten^{f,c}, Fabrizio Zilibotti^{g,c}

Abstract

For many kinds of capital, depreciation rates change systematically with the age of the capital. Consider an example that captures essential aspects of human capital, both regarding its accumulation and its depreciation: a worker obtains knowledge in period 0, then uses this knowledge in production in periods 1 and 2, and thereafter retires. Here, depreciation accelerates: it occurs at a 100% rate after period 2, and at a lower (perhaps zero) rate before that. The present paper analyzes the implications of non-constant depreciation rates for the optimal timing of taxes on capital income. The main finding is that under natural assumptions, the path of tax rates over time must be oscillatory. Oscillatory tax rates are optimal when depreciation rates accelerate with the age of the capital (as in the above example), and provided that the government can commit to the path of future tax rates but cannot apply different tax rates in a given year to different vintages of capital.

Keywords: Asset depreciation, Human capital, Optimal taxation, Oscillations, State-contingent taxes, Tax dynamics.

JEL classification: D90, E61, E62, H21, H30

†Corresponding author: John Hassler, IIES Stockholm University, 10691 Stockholm, Sweden. Telephone number: +46 816-2070. Fax number: +46 816-1443. Email: john.hassler@iies.su.se

[a] Institute of International Economic Studies, Stockholm University; [b]Princeton University; [c] CEPR; [d] CAERP; [e] NBER; [f] University of Oslo; [g]IEW University of Zurich;

1.   Introduction

What is the optimal path of taxes for a benevolent government that needs to finance some essential public expenditures? We study this question in a setting where taxation must take the form of proportional levies on capital income and where depreciation rates may vary over time (i.e., non-geometric depreciation). We are particularly interested in the case where depreciation rates increase with the age of capital, since we believe that it captures a realistic feature of the depreciation of physical and, in particular, human capital. For example, when a worker leaves the workforce, large parts of her human capital depreciates. We find that in such economies, it is optimal for the government to commit (if it can) to an oscillating tax sequence.

A standard principle in public finance is that taxation should be designed so as to keep distortions smooth over time. This principle applies whenever the social cost of raising tax revenue is convex, a circumstance that is met in most settings. In models where taxes only distort static decisions (e.g., to labor supply), and where the relevant elasticities are constant over time, this implies that taxes should be as close to constant as possible and that shocks to expenditures should be absorbed by time-varying debt (see, e.g., Barro, 1979). However, if taxes distort accumulation decisions, new issues arise, since such decisions depend not only on a single tax rate but on the *present value* of taxes generated by each unit of investment. One much studied question is how much tax revenue should be raised from income arising from static decisions (say, labor income) and how much should be raised from taxing income from accumulated production factors (such as physical capital). The seminal papers by Chamley (1986) and Judd (1985) in this area show, in particular, that optimal taxation in general involves taxing both labor and capital but at very particular, time-varying rates: over time, the tax rate on the accumulated factor should go to zero. Thus, they should be "front-loaded" and, in a typical setting (see Atkeson et al., 1999), high only for a finite number

of periods and thereafter zero forever. In this paper we emphasize that the smoothing of *distortions* across vintages of investment does not, in general, imply the smoothing of tax rates. Rather, under non-geometric depreciation, oscillations in tax rates can turn out to be a way of smoothing distortions while also imposing higher present-value taxes on the inelastic supply of initial capital and on capital installed early on.[1]

We consider a modified version of the standard neoclassical growth model. In addition to a more generalized depreciation structure, we consider linear utility—in order to avoid tax effects on the interest rate—and a two-sector production structure.[2] Consumption is linear in the capital input (which could be human capital or physical capital) whereas the production of investment involves decreasing returns. An important assumption for our results is that at any point in time, all capital income has to be taxed at the same rate; i.e., the government cannot impose vintage-specific taxes. Moreover, the government cannot levy taxes or subsidies on investments (see Section 5 for further discussion of these assumptions).

When depreciation is geometric, our model reproduces the standard result that taxes on capital should be front-loaded. Suppose, as is standard in the literature, that the government cannot tax capital income in period zero (which would be non distortionary). Then the planner taxes capital income in period 1 at a very high rate so as to extract revenue from the part of the initial tax base that is inelastic (i.e., from those assets that were accumulated before the start of the planning horizon). Thereafter, the optimal tax rate drops to its steady-state level. Though standard, an interesting aspect of this result is that the distortions on asset accumulation generated by this tax sequence are far from smooth: the tax burden is borne entirely by the investments in the first period. This may seem surprising: shouldn't the

---

[1]Strong time variation of tax rates is a characteristic of the optimal policy also in Greulich and Marcet (2007). They emphasize that, while capital taxes are front-loaded, labor taxes have to be back-loaded to encourage early capital accumulation. Moreover, Hagedorn (2007) emphasizes that in the presence of search frictions, Ramsey problems can be non-convex and therefore generate optimal tax cycles. This mechanism is, however, quite difference from the one we emphasize here.

[2]With linear utility the government will not try to manipulate the interest rate; see Lucas and Stokey (1983).

planner shift some burden to future investments, so as to smooth distortions? In addition, after the first period (with high taxation), since capital depreciates geometrically, there is still inelastic capital left. Both these factors speak for a large tax rate in the second period. However, the fact that the initial investment is heavily distorted by the first-period tax makes it very costly to distort it further by a high second-period tax rate. This speaks for lower taxes in period two. It turns out that the opposing forces cancel exactly under geometric depreciation, so that taxes go to their steady-state level immediately, although in our model the steady-state capital-income tax is not zero for reasons that are related to the analysis of Correia (1996).[3,4]

If, on the other hand, capital depreciates at a time-varying rate (changing with the age of the capital), the planner can and will use the timing of taxation to smooth distortions. To establish the result in a transparent way, we focus on a simple deviation from geometric depreciation that we label "quasi-geometric": the depreciation rate in the first period is allowed to be different from that in subsequent periods. The presence of a distribution of capital vintages turns the timing of taxation into an additional instrument for enabling distortion smoothing. We stress the case in which the depreciation rate increases with the age of the asset, since this seems empirically relevant for most types of capital (see below for more discussion). In this case, the Ramsey allocation implies oscillatory tax dynamics. The case of human capital illuminates this point. Suppose that the asset is accumulated in period $t - 1$ and is fully productive in periods t and $t + 1$ but not thereafter. This is a particular case of quasi-geometric depreciation, where the depreciation increases with the asset age (depreciation is zero initially, and then 100%). At time t, a surprise occurs, which

---

[3]Intuitively, if the present-value tax revenue extracted from inelastic capital were held constant, then shifting capital taxation to later dates would be detrimental: it would not reduce the burden on time-zero investments, and it would distort future investment decisions unnecessarily.

[4]Here, as in Correia (1996), we assume that a production input (investment goods) cannot be taxed. Absent this restriction on taxation, some long-run taxation of capital income will be optimal, since such taxation would indirectly allow some taxation of the untaxed input.

increases the need for the government to raise funds (e.g., a war).[5] In this case, the planner wants to seize the opportunity to extract a large amount of tax revenue from the generation that made its investment before the war. This generation sunk its investment under the expectation of lower taxes, and this investment is, at t and $t + 1$, an inelastic tax base, calling for a high tax rate. The key insight is that this high tax rate can be counteracted by a lower tax rate in $t + 2$ so that investments in period t are not too distorted. The revenue from taxes paid by capital originating from investments done before the shock is not hurt by the reduction in period $t + 2$ taxes, since it is fully depreciated by then. Then, since the $t + 2$ tax rate is low, a higher $t + 3$ tax rate helps smooth investments, and so on. This oscillating plan features a smoother path of distortions than full front-loading would. At the same time, it allows the planner to exploit the lower elasticity of the tax base at t. This example is simple and intuitive because the asset (human capital) is only productive for two periods. However, we show that this intuition is robust to the case where assets are infinitely lived and depreciate smoothly but at rate that is increasing in the asset age.

In Section 2, we describe the basic setup from the perspective of standard Ramsey problems where the issue is that of how and when to finance an exogenous stream of government expenditures when the government can borrow and lend. Section 3 derives our main results. Section 4 introduces stochastic shocks to government spending needs. This extension shows that, if government debt is not state contingent, optimal tax oscillations can arise after a fiscal shock. Thus, the fluctuations in our examples are not necessarily mere memories of the initial-period capital stock. However, if debt is state-contingent, no new fluctuations occur: those that are present are indeed a memory of the initial period. Section 5 concludes. The appendix contains some proofs and technical derivations. Some additional proofs are

---

[5]The assumption in this example—that the change is a "surprise"—is made for simplicity. It can be interpreted as allowing the planner to make a commitment but then re-optimize after a zero-probability shock realization. We show in Section 4 that the argument is robust to assuming that the shock is the realization of a stochastic process of which agents know the probability distribution, and the government commits, ex ante, to a state-contingent plan.

contained in a technical appendix, available from the corresponding author's webpage.

2.   The model

In this section, we set up the basic model. We first discuss the main maintained assumption – the general structure of depreciation – and then describe the Ramsey problem facing a benevolent planner who must finance an exogenous stream of expenditures.

2.1.   Quasi-geometric human capital depreciation

The key new element we consider is variable depreciation rates of the stock of capital. To fix ideas, we will refer to human capital throughout, though we briefly argue in Section 5 that also many kinds of physical capital share this depreciation structure. Thus, let us subdivide the life of a unit of capital, which is now represented by a worker, into three stages: youth, young adulthood, and old adulthood. The conditional probability of death increases with age. More precisely, a young agent dies with probability zero, a young adult dies with probability $\delta\rho$ and an old adult dies with probability δ, where $\delta \in (0,1]$ and $\rho \in [0,1]$. Moreover, each period young agents are born so that the size of the population is constant. Youth and young adulthood last for at most one period: a surviving young agent turns into a young adult, whereas a surviving young adult turns into an old adult. Only young agents invest in human capital, e.g., through education. A unit of investment at time t leads to one unit of productive capital in period $t + 1$. Thereafter, human capital does not depreciate within the lifetime of an individual, but disappears when an agent dies.[6] Thus, the expected contribution to the future stock of human capital of a unit of investment at t is 1 unit in period $t+1$, $1-\rho\delta$ units in period $t+2$, and $(1-\rho\delta)(1-\delta)^k$ units in period $t+2+k$. We label this structure *quasi-geometric* depreciation. Note that $\rho = 1$, i.e., a constant mortality rate

---

[6]In addition, one could assume that the human capital (knowledge) of an individual decreases with age even conditionally on survival. This would yield the same qualitative dynamics of human capital depreciation as the mortality channel discussed in the text.

within the working population, yields a standard, geometric depreciation of human capital, whereas $\rho < 1$, i.e., an increasing mortality rate within the working population, yields a lower initial depreciation than in the geometric case. The case $\rho = 0$ and $\delta = 1$, on the other hand, describes the case where worker human capital survives for two periods without depreciation and then disappears. Figure 1 represents a case of accelerating depreciation, showing the fraction of investments made in period $t - 1$ that survives at t, $t + 1$, ..., etc.

< FIGURE 1 ABOUT HERE >

In order to derive implications for the optimal taxation of human capital, we consider a discrete-time, infinite-horizon model where agents age and die according to the description above. To abstract from mortality risk issues, which are orthogonal to our focus, we assume agents to be part of "large families". In particular, the economy is populated by a continuum of representative unitary households, each consisting of a continuum of agents of different ages.[7] The total size of the representative household is unity. The age distribution of each household is constant over time. As above, an agent born in period t builds up i_t units of human capital in the first period of her life and becomes productive as of period t+1. Thereafter, her human capital remains constant until her death.

The total stock of human capital of the household is the integral of the human capital of all its members. Because of the age-dependent mortality rates, in order to determine the total human capital of the household, it is necessary to distinguish between two kinds of human capital at time t: the capital of the old adults, for which we use the notation h_t^o, and that of the young adults, h_t^y. Clearly, $h_t^y = i_{t-1}$. The difference between these kinds of capital is not in their productivities—the total human capital input of the household at t, which we call h_t, equals $h_t^o + h_t^y$—but in their depreciation rates from t to $t + 1$. Thus, our

---

[7]In our unitary households, all utilities will be interpreted from the perspective of perfect altruism across generations: the representative agent is a "dynasty planner" who internalizes the effects of current choices on all future generations.

assumptions are summarized by the following laws of motion for the two types of human capital:

$$h^o_{t+1} = (1-\delta)h^o_t + (1-\rho\delta)i_{t-1}, \tag{1}$$

$$h^y_{t+1} = i_t.$$

These equations amount to a generalized version of the standard accumulation equation:

$$h_{t+1} = i_t + (1-\delta)h_t + \delta(1-\rho)i_{t-1}. \tag{2}$$

In this formulation, total human capital productive next period equals (i) the investment made this period plus (ii) total capital in use this period depreciated at rate δ, with (iii) an adjustment upward by $\delta(1-\rho)i_{t-1}$ due to the fact that not all capital in use today actually depreciates at a constant rate δ: part of it, i_{t-1}, depreciates at the lower rate $\rho\delta$.[8] Notice, in particular, that when $\rho = 1$ equation (2) reduces to the standard Blanchard-Yaari perpetual youth model which yields geometric depreciation.. Much of the analysis below will be conducted in terms of old capital, h^o, since it is a natural state variable, whereas h_t is not.

A standard three-period model where agents invest in their youth and work for two periods can be viewed as a particular case of the general quasi-geometric depreciation structure described above, where $\delta = 1$. [9] In this case, $\rho = 0$ means that productivity is constant throughout the life of an individual, whereas $\rho > 0$ would capture a downward-sloping age-earnings profile (the worker's knowledge depreciates with age). We will focus on this simple case in the analysis of stochastic shocks of Section 4.

---

[8]With $h_t - i_{t-1}$ depreciating at rate δ and i_{t-1} at rate $\rho\delta$, the new total capital in use becomes $i_t + (h_t - i_{t-1})(1-\delta) + i_{t-1}(1-\rho\delta)$, which delivers the right-hand side of equation (2).

[9]Its counterpart in the literature on physical capital depreciation is a *one-hoss shay* depreciation structure, where investment at t stays intact until $t+2$ but then depreciates fully.

2.2. Preferences and technology

Since our goal is to develop a tractable framework, we introduce two stark assumptions about technology and preferences. First, we assume the intertemporal preferences of the representative household to be time-additive and the intratemporal preferences to be linear in consumption and quadratic in "educational effort."[10] More formally,

$$U_0 = \sum_{t=0}^{\infty} \beta^t \left(c_t - i_t^2 \right),$$

implying that the gross interest will be $1/\beta$.[11] Second, we assume that the production function is linear in human capital. In particular, production at t is simply h_t: it equals total (old plus new) human capital. In this model, the issue is purely one of when income should be taxed; there is no choice between taxing different factors of production.[12]

The representative household chooses investment plans to maximize U_0. The optimal choice of investment must balance the marginal cost of investment $(2i_t)$ and the expected present discounted value (PDV) of the after-tax output generated by a marginal unit of human capital. Since the marginal product of human capital is unity by assumption, this value is given by

$$\beta \left(1 - \tau_{t+1} \right) + \left(1 - \rho\delta \right) \sum_{s=2}^{\infty} \beta^s \left(1 - \delta \right)^{s-2} \left(1 - \tau_{t+s} \right).$$

Defining

$$\kappa \equiv \beta + \left(1 - \rho\delta \right) \sum_{s=2}^{\infty} \beta^s \left(1 - \delta \right)^{s-2} = \beta \frac{1 + \beta\delta \left(1 - \rho \right)}{1 - \beta \left(1 - \delta \right)}$$

---

[10]¿From now on, all variables will be aggregated at the unitary household level. Note that, due to risk neutrality, our formulation is identical to one in which there is no unitary household and the planner is utilitarian, i.e., attaches the same weight to all living agents.

[11]It is possible to relax the assumption of quadratic investment costs and generalize it to any convex cost. Then, one can provide a characterization of the dynamics around a steady state which is qualitatively identical to the global solution we obtain.

[12]In spite of the linear technology, our model does not feature endogenous growth, due to the quadratic investment cost.

and

$$T_t \equiv \beta \tau_{t+1} + (1 - \rho \delta) \sum_{s=2}^{\infty} \beta^s (1 - \delta)^{s-2} \tau_{t+s}, \tag{3}$$

it follows immediately that we can write the household's optimal choice of investments in the following compact way;

$$i_t = i(T_t) \equiv \frac{1}{2} (\kappa - T_t) \tag{4}$$

where κ is the *effective duration* of new investment and T_t is the the *effective discounted sum of taxes* (which we label the "present-value tax") on period-t investments.

2.3. The Ramsey problem

The government must finance a given sequence $\{g_t\}_{t=0}^{\infty}$ of expenditures subject to an intertemporal budget constraint

$$b_0 + \sum_{t=0}^{\infty} \beta^t (g_t - \tau_t (h_t^o + i_{t-1})) \leq 0, \tag{5}$$

where pre-tax output equals $h_t = h_t^o + i_{t-1}$ and b_0 is initial government debt. Note that the only instrument available to the government is taxation of the return to human capital, which coincides here with output taxation.

The Ramsey problem can now be formulated as a planner choosing a tax sequence maximizing the representative household's utility subject to its budget constraint (5), and the restriction that the allocation be a competitive equilibrium. Due to risk neutrality, maximizing total utility of the representative household is equivalent to maximizing the PDV of after-tax output minus investment costs. Therefore, the Ramsey problem amounts to

$$\max_{\{\tau_t, i_t, h_{t+1}^o\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left((h_t^o + i_{t-1})(1 - \tau_t) - i_t^2 \right),$$

subject to the budget constraint, (5), the law of motion of (old) capital under quasi-geometric depreciation, (1), and the implementability constraint, (4). In addition, we impose that tax

rates are bounded and that $\tau_0 = 0$.[13]

Before turning to the analysis, it is useful to relate our model to existing results in the optimal capital taxation literature. First, using a model with geometric capital depreciation and a linear (as opposed to quadratic) investment cost, Chamley (1986) and Judd (1985) established that if the Ramsey tax sequence converges to a steady state, then the steady state must be zero.[14] However, the Chamley-Judd result does not apply to our model even in the particular case of geometric depreciation. In fact, we will show later that our model features positive taxation in the long run, due to the quadratic investment costs. This result, which is not the main focus of our analysis, is a particular case of the more general analysis by Correia (1996).[15]

3.   Analysis

Define λ as the Lagrangian multiplier associated with the government budget constraint. The Lagrange method then implies that the Ramsey problem can be expressed, after rearranging terms, as:

$$\max_{\{\tau_t, i_t, h^o_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t ((\tau_t (\lambda - 1) + 1)(h^o_t + i_{t-1}) - i_t^2) - \lambda \left(b_0 + \sum_{t=0}^{\infty} \beta^t g_t \right). \tag{6}$$

The solution to the problem in (6) depends on the Lagrangian multiplier, λ. The value of λ is determined by minimizing the objective in (6). It represents the shadow value of the government's budget constraint, (5), and is increasing in the government's needs to

---

[13]The tax τ_0 would be lump-sum as it is levied on predetermined human capital only. Therefore, if τ_0 were a choice variable, it would be set at its maximum feasible level, with no effect on any other choice. Hence, setting $\tau_0 = 0$ is without loss of generality.

[14]In the analysis of Chamley and Judd, the planner can tax both labor and capital. However, this is not important for the current discussion. Moreover, Atkeson et al. (1999) show that, under CRRA utility, the Ramsey solution features zero capital taxation for all $t \geq 2$.

[15]Correia's main insight is that untaxed input factors provide one channel through which capital taxation can be used beneficially, even in the long run. In our framework, the human-capital investment is a non-taxable household activity, subject to increasing marginal cost. There are therefore untaxed "profits" in this operation, which are equivalent to the untaxed factor income in Correia's analysis.

raise funds $\left(b_0 + \sum_{t=0}^{\infty} \beta^t g_t\right)$ and decreasing in $h_t^o + i_{t-1}$. In the rest of the paper, we will characterize the optimal sequence of taxes conditional on λ, bearing in mind that conditioning on λ is equivalent to conditioning on a set of initial conditions.[16]

The following Lemma is a useful step towards characterizing the solution to the Ramsey problem (the proof is simple algebra and is, therefore, omitted).

Lemma 1 *Setting $\tau_0 = 0$, the Ramsey problem, (6), subject to (1) and (4) is equivalent to the following program:*

$$\max_{\{\tau_t\}_{t=0}^{\infty}} (\lambda - 1) \left(\tau_0 \left(h_0^o + i_{-1}\right) + \hat{T}_0 h_1^o\right) + \sum_{t=0}^{\infty} \beta^t y\left(T_t\right) - \lambda(b_0 + \sum_{t=0}^{\infty} \beta^t g_t), \qquad (7)$$

where

$$\hat{T}_0 \equiv \beta \sum_{t=0}^{\infty} (\beta(1 - \delta))^t \tau_{t+1}, \qquad (8)$$

$$y(T_t) \equiv \lambda \kappa i\left(T_t\right) - i\left(T_t\right)^2 (2\lambda - 1), \qquad (9)$$

and T_t and $i\left(T_t\right)$ are defined as in (3) and (4), respectively.

The new functions $y\left(T_t\right)$ and $\hat{T}_0$ will be particularly useful in the analysis below. The function $y\left(T_t\right)$ is the contribution of the human-capital investment of generation t to the planner's discounted utility. Each such "vintage" investment contributes to the planner's utility via private consumption, $i_t\left(\kappa - T_t\right)$, the financing of government expenditure, $\lambda T_t i_t$, and the investment cost, $-i_t^2$. By using (4) to eliminate T_t, expression (9) follows immediately. Furthermore, $\hat{T}_0$ is the effective discounted sum of taxes levied on human capital investments made before the beginning of the planning horizon, and thus inelastic. With analogy to previous definitions, we label it the "present-value tax on inelastic capital". Taxes

---

[16]The Ramsey problem above admits an alternative interpretation whereby households derive utility from both private consumption and the consumption of a public good. The intratemporal utility is modified to $u(c, g, i) = c + \lambda g - i^2$, where λ in this case denotes the constant marginal utility agents derive from the consumption of the public good, and government revenue is entirely spent on the public good. See the working paper version of the present paper, Hassler et al. (2004), for details.

entering $\hat{T}_0$ are discounted at the rate $\beta(1 - \delta)$, reflecting the discount factor and the rate of depreciation of the initial human capital.

The objective function (7) is then the sum of the PDV of the contribution to the planner's utility of all investments from time zero onwards, $\sum_{t=0}^{\infty} \beta^t y(T_t)$, and the PDV of the tax revenue from pre-existing human capital. Ignoring irrelevant constants and predetermined variables and recalling that we rule out lump-sum taxes ($\tau_0 = 0$), the Ramsey problem simplifies to

$$\max_{\{\hat{T}_0, T_t\}_{t=0}^{\infty}} (\lambda - 1)\hat{T}_0 h_1^o + \sum_{t=0}^{\infty} \beta^t y(T_t), \tag{10}$$

where

$$\hat{T}_0 = \sum_{t=0}^{\infty} (-\delta\beta(1-\rho))^t T_t. \tag{11}$$

Expression (11) follows from equations (3) and (8) and is formally derived in the appendix. This expression implies that $\hat{T}_0$ cannot be chosen independently of the T_ts. Note that h_1^o is a key predetermined variable; its size will influence the dynamics of present-value taxes.

The program (10) pins down the optimal sequence $\{T_t\}_{t=0}^{\infty}$s and $\hat{T}_0$ rather than the tax sequence $\{\tau_t\}_{t=1}^{\infty}$. Since $i_t = (\kappa - T_t)/2$, this amounts to the planner choosing the investment sequence. Thus, (10) is a *primal* formulation, where the planner chooses an allocation directly, subject to the constraint that it is a competitive equilibrium. In the appendix, we prove formally that there is a one-to-one mapping between the primal and the dual formulations. Namely, a sequence of present-value tax rates (or investments) pins down uniquely a sequence of individual tax rates satisfying (3) and (8).[17]

---

[17]Intuitively, because tax rates are bounded and $\beta(1-\delta) < 1$, the present-value taxes must be bounded as well. Forward iterating on equation (3) leads to $T_{t+1} = \beta^{-1}(1-\delta)^{-1}(T_t - \beta\tau_{t+1})$. This difference equation can be solved for a unique feasible sequence of tax rates. Namely, given a sequence $\{T_t\}_{t=0}^{\infty}$, one can back out a unique sequence of tax rates $\{\tau_t\}_{t=1}^{\infty}$ which satisfies the boundedness condition. See Proposition 2 below and its proof in the appendix.

3.1. The case of geometric depreciation

We first analyze the benchmark case of constant mortality rate of the adult population, i.e., geometric depreciation ($\rho = 1$). In this case all the present-value taxes are geometric (as opposed to quasi-geometric) sums of future tax rates. In particular, $\hat{T}_0 = T_0$: the present-value tax on inelastic capital is identical to the present-value tax on investment in period zero, which is distortionary. Thus, we can rewrite the Ramsey problem of equation (10) as

$$\max_{\{T_t\}_{t=0}^{\infty}} (\lambda - 1)T_0 h_1^o + \sum_{t=0}^{\infty} \beta^t y(T_t).$$

Expression (11), which is formally derived in the appendix and follows from equations (3) and (8), implies that $\hat{T}_0$ cannot be chosen independently of the T_ts. Note that h_1^o is a key predetermined variable; its size will influence the dynamics of present-value taxes.

The program (10) pins down the optimal sequence $\{T_t\}_{t=1}^{\infty}$s and $\hat{T}_0$ rather than the tax sequence $\{\tau_t\}_{t=1}^{\infty}$. Since $i_t = (\kappa - T_t)/2$, this amounts to say that the planner chooses the investment sequence. Thus, (10) is a *primal* formulation, where the planner chooses an allocation directly, subject to the constraint that it is a competitive equilibrium. In the appendix, we prove formally that there is a one-to-one mapping between the primal and the dual formulation. Namely, a sequence of present-value tax rates (or investments) pins down uniquely a sequence of individual tax rates satisfying (3) and (8).[18]

In period zero, the problem is different: here T_0, which distorts i_0, also raises revenue from the taxation of the inelastic human capital, h_1^o. Thus, the optimal T_0, which we label T_0^*, satisfies

$$(\lambda - 1)h_1^o + y'(T_0^*) = 0.$$

---

[18]Intuitively, because tax rates are bounded and $\beta(1-\delta) < 1$, the present-value taxes must be bounded as well. Forward iterating on equation (3) leads to $T_{t+1} = \beta^{-1}(1-\delta)^{-1}(T_t - \beta\tau_{t+1})$. This difference equation can be solved for a unique feasible sequence of tax rates. Namely, given a sequence $\{T_t\}_{t=0}^{\infty}$, one can back out a unique sequence of tax rates $\{\tau_t\}_{t=1}^{\infty}$ which satisfies the boundedness condition. See Proposition 2 below and its proof in the appendix.

Clearly, $T_0^* > T^*$ which in turn implies that $\tau_1 > \tau^*$. The extent of the initial tax hike depends positively on h_1^o.

3.2. Quasi-geometric depreciation

In the general case with an increasing mortality rate (quasi-geometric human-capital depreciation), $\hat{T}_0$ is no longer equal to T_0. Since the inelastic capital, h_1^o, depreciates at a different rate from new investments, the timing of taxes can be used to improve efficiency. Now, the connection between $\hat{T}_0$ and the sequence of T_ts in equation (11) is key for understanding the oscillatory tax dynamics: if $\rho < 1$, the weights on the future present-value taxes T_t have alternate signs. Thus, every T_t will influence the taxation of inelastic capital, and whether T_t increases or decreases the present-value tax on inelastic capital depends on whether t is even or odd.

After eliminating $\hat{T}_0$, using (11), from (10), the Ramsey problem now reads

$$\max_{\{T_t\}_{t=1}^\infty} (\lambda - 1) \left(\sum_{t=0}^\infty (-\delta\beta(1-\rho))^t T_t \right) h_1^o + \sum_{t=0}^\infty \beta^t y(T_t).$$

The first-order condition with respect to T_t is

$$(\lambda - 1) h_1^o (-\delta(1-\rho))^t + y'(T_t) = 0. \tag{12}$$

The set of FOCs for $t \geq 0$ pins down uniquely the optimal present-value tax sequence $\{T_t\}_{t=1}^\infty$ and, hence, the optimal tax sequence $\{\tau_t\}_{t=1}^\infty$ (see the proof of Proposition 2). Note that the first-order condition for T_0 is the same as in the case of geometric depreciation. However, under geometric depreciation h_1^o only affects future present-value taxes via its effect on the Lagrange multiplier, λ. In contrast, under quasi-geometric depreciation h_1^o also affects directly the dynamics of the entire sequence of investments and taxes, as shown by equation (12). The solution can be summarized by our main proposition.[19]

---

[19]The assumption that $b_0 + \sum_{t=0}^\infty \beta^t g_t$ is not too large is meant to avoid uninteresting complications arising from corner solutions in the choice of taxes.

Proposition 2 *Assume that $\|\delta(1-\rho)\| \leq 1$ and that $b_0 + \sum_{t=0}^{\infty} \beta^t g_t$ is not too large. Then, the optimal (Ramsey) present-value tax sequence is given by*

$$T_t = \frac{\lambda - 1}{2\lambda - 1}\left(\kappa + 2h_1^o\left(-\delta(1-\rho)\right)^t\right) \ \ for \ t \geq 0. \tag{13}$$

The corresponding unique tax sequence that implements the Ramsey allocation is:

$$\tau_{t+1} = \tau^* - \delta(1-\rho)(\tau_t - \tau^*) \ \ for \ t \geq 1, \tag{14}$$

$$\tau_1 = \tau^*\left(1 + 2h_1^o\frac{1 + \beta\delta(1-\delta)(1-\rho)}{\beta\left(1 - \beta\delta^2(1-\rho)^2\right)}\right), \tag{15}$$

where $\tau^* \equiv (\lambda - 1)/(2\lambda - 1) < 1/2$, and λ guarantees that equation (5) is satisfied with equality, given the investment rule (4), the definition of T_t in (3), and the optimal tax sequence defined by (14)-(15). If $\delta(1-\rho) = 0$, then the tax sequence is constant after the first period. If $\delta(1-\rho) \in (0,1)$, then the tax sequence converges in an oscillatory fashion to τ^*. If $\delta(1-\rho) = 1$, then the optimal tax sequence is a two-period cycle.

Proof (sketch): The first-order condition (12), together with the definition of $y(T_t)$ as given in (9), yield the optimal present-value tax sequence, (13). The proof in the appendix amounts to showing that the tax sequence (14)-(15) is the unique sequence satisfying (13) and the tax constraint $\tau_t \leq 1$, given the definition of the T_t's as in (3).

Figure 2 shows the dynamics of tax rates (τ_t), present-value taxes (T_t), investments and net output, defined as $h_t^o + i_{t-1} - i_t^2$, in a case of quasi-geometric depreciation. Note that investments fluctuate less than taxes, an illustration of the fact that although taxes may fluctuate a lot over time, investments and distortions are smoother. Net output fluctuates around a geometric trend toward the steady state.[20]

<div align="center">< FIGURE 2 ABOUT HERE ></div>

---

[20]However, gross output, excluding investment costs, i.e., $h_t^o + i_{t-1}$, displays monotone convergence and is, in fact, constant in the case of $\delta = 1$. The proof and details about the calibration are in the technical appendix, available from the corresponding author's webpage.

3.3.   Interpretation

3.3.1.   A second-best benchmark: age-specific taxation

In order to understand the results of the previous section, it is useful to compare them with the case in which the planner has access to age-specific taxation, i.e., she can tax the income produced by different cohorts at different rates. The Ramsey sequence is then very simple: the planner taxes the human capital income of the initially old adults, h_1^o, at the highest possible rate every period, since these taxes are non-distortionary. All cohorts after period zero are then taxed at the constant rate τ^* such that $y'(T_s) = 0$ for all $s > 0$, where $T_s = T^* \equiv \beta(1 - \beta(1 - \delta))^{-1}\tau^*$. We will refer to this benchmark allocation as *second best*. This allocation achieves a perfectly smooth distortion of investments by smoothing perfectly the taxes affecting future investment vintages.

In contrast, when age-specific taxes are ruled out, the planner cannot separate taxation of output produced by inelastic human capital from distortionary taxation on output produced by later human-capital vintages.[21] Thus, a trade off arises between the objective of smoothing distortions and that of taxing inelastic human capital. Note, that the Ramsey tax sequence of Proposition 2 features perfect tax and investment smoothing only when $h_1^o = 0$: when there is no inelastic capital, the planner chooses constant taxes as she would do in the second best.

3.3.2.   Geometric depreciation ($\rho = 1$)

In the case of geometric depreciation, there are no oscillations, and taxes are smooth after one period. Investments, however, are far from smooth. In particular, since $\tau_1 > \tau^*$, while $\tau_t = \tau^*$ for all $t > 1$, all distortions generated to extract income from the inelastic capital are borne by the first cohort of young agents ($T_0 > T_t = T^*$, for all $t > 0$). This

---

[21]Hassler et al. (2007) analyzes the properties of the Ramsey allocation in a two-period version of this model when age-dependent taxation is allowed.

implies very low investments in period zero. Why does the planner not attempt to smooth distortions by taxing capital at later dates, thus reducing τ_1 so as to increase i_0?

First, given the present-value tax on inelastic capital, $\hat{T}_0$, it is impossible for the planner to use the timing of taxes to alleviate distortions on period-zero investments. This follows immediately from the fact that $\hat{T}_0 = T_0$. For instance, if the planner were to reduce τ_1 and increase τ_2 so as to keep $\hat{T}_0$ constant, investment in period zero would not change. Second, such tax reallocation would increase T_1 and distort it away from the second best level, T^*. The same argument applies to any other potential changes in the timing of taxation (e.g., the same experiment using τ_3 instead of τ_2 would increase both T_1 and T_2). In sum, it is optimal for the planner to "front-load" taxes in order not to distort investments after the first period.

Our results imply that taxes for periods $t > 1$ only depend on h_1^o via its effect on λ (a larger h_1^o increases the tax revenue all else equal, relaxing the government budget constraint, and implying lower λ and lower τ^*). To understand this result, note that along the optimal path, the marginal distortion of τ_s must be proportional to the marginal revenue generated by that tax. If h_1^o is increased, the marginal revenue raised by τ_1 increases, so τ_1 should then be increased, increasing the distortion on period zero investments i_0. What are the implications for the optimal choice of τ_2? The trade-off between distortions and revenue generation for τ_2 is affected in two ways. First, as for τ_1, the higher h_1^o affects the marginal revenue of τ_2 positively. Second, however, the higher distortion on period-zero investments increases the marginal distortionary cost of τ_2 since this tax affects i_0 (in addition to affecting i_1). Under geometric depreciation, these two effects exactly balance each other out and the increase in τ_1 caused by a higher h_1^o should not lead to any changes in τ_2 or, more generally, in any subsequent tax rates.

3.3.3. Quasi-geometric depreciation

We now move to the general case, where $\rho < 1$. According to Proposition 2, the Ramsey tax sequence is oscillating when $\rho \in [0, 1)$. We refer to this case as accelerating depreciation, since capital depreciates less in the first period than afterwards. In order to understand why oscillations arise, it is useful to start from a particular case.

A particular case: $\delta = 1$ The case of $\delta = 1$ has a feature that makes the analysis particularly intuitive: τ_1 is the only instrument the planner has available for taxing the inelastic capital. Taxes at later dates do not extract revenue from h_1^o, since this will have depreciated fully. Why, then, not set $\tau_t = \tau^*$ for $t > 1$, instead of producing an oscillating sequence after the initial tax hike? The reason is that, unlike in the case of geometric depreciation, the planner can now use the timing of taxes to smooth future distortions. Recall that, while an initial tax hike is attractive since it generates revenue from an inelastic base, it also distorts investments in period zero, i_0 (as in the case of geometric depreciation, the magnitude of such hike is increasing in the inelastic capital). These distortions can be mitigated, because investment decisions depend on both τ_1 and τ_2 (recall that, when $\delta = 1$, we have $T_t = \tau_{t+1} + \beta (1 - \rho) \tau_{t+2}$). Thus, the planner can alleviate the distortion on period zero investments by promising a low tax rate in period two. In turn, the low tax rate in period two stimulates investments in period one, and since it is optimal to keep distortions smooth, it is therefore useful to compensate the tax break in period two by another tax hike in period three, and so on.

In contrast to the case of geometric depreciation, taxes at dates $t > 1$ are now affected by the size of the stock of inelastic capital, h_1^o. To understand this, note that when h_1^o is higher, it is optimal to increase τ_1 (relative to future taxes). This increases the marginal distortion of τ_2 because i_0 is already distorted by a high τ_1. Moreover, τ_2 does not extract revenue from h_1^o since it is fully depreciated by period $t = 2$. Thus, it is optimal to reduce τ_2.

The parameter ρ is key for the size of the oscillations. Consider for instance the extreme case when $\rho = 0$: the *one-hoss shay* case. As Proposition 2 shows, in this case oscillations do not die out: the economy ends up in a two-period cycle. The reason is that the increase in the distortion entailed by τ_2 on i_0 is particularly large since i_0 has not depreciated at all by period $t = 2$. Equivalently, the effectiveness of counteracting a current tax hike by a next-period tax break is high. When $\rho > 0$, a larger share of the return on the investment is accrued in the first period of life than in the second. Therefore, reducing τ_2 will be a less effective instrument for counteracting distortions in period zero. Hence, oscillations are smaller and die out in the long run.

The general case with accelerating depreciation   We now turn to the general case of accelerating quasi-geometric depreciation: $\rho \in [0, 1)$ and $\delta \in (0, 1)$. As under geometric depreciation, human capital is never completely depleted, and the present-value tax on inelastic capital, $\hat{T}_0$, depends on the entire tax sequence. However, unlike in the case of geometric depreciation, the Ramsey tax sequence follows an oscillatory pattern. The general point is that since $\hat{T}_0 \neq T_0$, it is possible to use the timing of taxes to alter T_0 while leaving $\hat{T}_0$ unchanged. For instance, if we decrease τ_2 and increase τ_1 so as to keep $\hat{T}_0$ constant, T_0 will decrease, since taxes from period two and onwards have a larger impact on T_0 than on $\hat{T}_0$.[22] The planner can now use the timing of taxation as an imperfect substitute for the absence of age-specific taxes and achieve better distortion smoothing. Recall, in particular, that the hike in τ_1 distorts heavily i_0. Thus, distortion smoothing makes it desirable for the planner to use future taxes to reduce T_0. This is achieved by setting $\tau_2 < \tau^*$ (as in the $\delta = 1$ case). However, having done this, it is not optimal to set $\tau_t = \tau^*$ for $t > 2$, because such a sequence would imply a deviation from the second-best benchmark in the direction of too large investments in period one ($T_1 < T^*$), while all future investment levels would be set at

---

[22]The particular case of $\delta = 1$ provides an extreme example: by keeping τ_1 constant and reducing τ_2, one can decrease T_0 while keeping $\hat{T}_0$ constant.

the second-best level. Again, distortion smoothing suggests an increase in τ_3 so as to reduce i_1, and so on.

4. Stochastic government expenditure

Proposition 2 establishes conditions under which fluctuations in taxes and output are efficient. However, if $h_1^o = 0$ (i.e., no pre-installed capital at time zero), the optimal tax sequence is smooth; that is, the optimal tax oscillations implied by the model can be entirely traced back to an initial condition. The aim of this section is to show that when future expenditure needs are stochastic and markets are incomplete (no state-contingent debt can be issued), then the transitional dynamics of the optimal tax sequence feature oscillations even if there is no inelastic capital to begin with. However, if the government can issue state-contingent debt, no oscillations arise.

For simplicity, government-expenditure risk is limited to a one-time event only. More precisely, as of period 1 it is revealed whether spending requirements will be high (state h) or low (state l). However, in period zero the state is unknown, and $p \in (0, 1)$ denotes the probability that the state will be high. Again, for simplicity we focus on the case $\delta = 1$, i.e., the standard overlapping-generations case with no intergenerational human capital transmission, and assume that $i_{-1} = 0$.

4.1. Incomplete markets

In this section, we assume that the government cannot issue state-contingent debt. However, the government can set, with full commitment, state-contingent taxes sequences, except for τ_1. An interpretation of this assumption is that τ_1, as well all other tax rates, must be set one period in advance: there is an "implementation lag" of one period, implying that the tax rate in period one cannot depend on information revealed in period one, whereas the

subsequent taxes can depend on that information. Thus, at time zero, the planner sets τ_1 and a state-contingent tax plan, $\{\tau_{j,t}, \tau_{j,t}\}_{t=2}^{\infty}$ for $j \in \{l, h\}$. When the first-period investment, i_0, is chosen, only τ_1 is known with certainty, whereas agents do not know whether the tax rate will be $\tau_{h,2}$ or $\tau_{l,2}$ in period two. In contrast, all subsequent generations of investments (including i_1) are made under perfect information.

When uncertainty has unraveled, the sequence $\{\tau_{h,t}\}_{t=2}^{\infty}$ is implemented if spending requirements are high, whereas the sequence $\{\tau_{l,t}\}_{t=2}^{\infty}$ is implemented if these are low. The tax sequences must now satisfy one government budget constraint for each state $j \in \{l, h\}$:

$$b_0 + g_0 + \sum_{t=1}^{\infty} \beta^t g_{j,t} \equiv G_j = \beta \tau_1 i_0 + \sum_{t=2}^{\infty} \beta^t (\tau_{j,t}(h_{j,t}^o + i_{j,t-1})), \tag{16}$$

where $h_{j,t}^o$ and $i_{j,t}$ denote equilibrium stocks of old and new capital in state j at time t. The Lagrange multipliers of the two budget constraints are denoted by λ_h and λ_l. Clearly, $G_h > G_l > 0$ implies $\lambda_h > \lambda_l > 0$.[23]

We extend the definition (9) to the stochastic case:

$$y_j(T) \equiv \lambda_j \kappa i(T) - i(T)^2 (2\lambda_j - 1). \tag{17}$$

Thus, $y_j(T_{j,t})$ will be the *realized* contribution to the planner's utility of the human-capital investment of the generation born in period $t \geq 1$, conditional on state $j \in \{l, h\}$ and on the realized present-value tax $T_{j,t}$. The analysis of the investment of the generation born in period zero necessitates some new notation, as such

The investment i_0 is made under uncertainty and requires some new notation. We denote by

$$y_0^e(T_{l,0}, T_{h,0}) \equiv \left[i(T_0^e)(\kappa - T_0^e) - i(T_0^e)^2 \right] + (1-p)\lambda_l i(T_0^e) T_{l,0} + p\lambda_h i(T_0^e) T_{h,0} \tag{18}$$

---

[23] As above, we do not solve explicitly for λ_j as a function of G_j. However, we note that the optimal steady-state tax rate corresponding to a a particular value of λ_j is given by $\tau_j^* = (\lambda_j - 1)/(2\lambda_j - 1)$.

the *expected* contribution to the planner's utility of the generation born at zero. The *expected* (as opposed to realized) period-zero present-value tax is denoted:

$$T_0^e \equiv pT_{h,0} + (1-p)\,T_{l,0}, \tag{19}$$

where, again, $T_{l,0}, T_{h,0}$ are the *realized* present-value taxes on period-zero investments in the two states. Due to certainty equivalence, i_0 is fully determined by T_0^e. The right-hand side of equation (18) consists of two terms: the certainty equivalent utility from private consumption (in square brackets) and the expected value for the planner of the tax revenue levied on the investments of the generation born at zero.

The Ramsey plan can be formulated as

$$\max_{\{\tau_1,T_{h,t},T_{l,t}\}_{t=0}^{\infty}} y_0^e\left(T_0^e, T_{l,0}, T_{h,0}\right) + \sum_{t=1}^{\infty} \beta^t \left(p \cdot y_h\left(T_{h,t}\right) + (1-p) \cdot y_l\left(T_{l,t}\right)\right), \tag{20}$$

subject to (17), (18), (19), and the constraints that, for $j \in \{l, h\}$,

$$T_{j,0} = \beta\tau_1 + \beta\,(1-\rho) \sum_{t=0}^{\infty} \left(-\beta\,(1-\rho)\right)^t T_{j,t+1}, \tag{21}$$

which generalize equation (11).[24]

Here, we summarize the results. The details of the analytical derivations are provided in the technical appendix available from the corresponding author's webpage. Substituting the constraints (19) and (21) into (20), T_0^e, $T_{l,0}$ and $T_{h,0}$ can be eliminated from the objective function. The Ramsey program can then be formulated as an unconstrained maximization

---

[24]The expression in (21) is derived from the definition $T_{j,t} \equiv \beta\tau_{j,t+1} + (1-\rho)\,\beta^2\tau_{j,t+2}$, which implies

$$T_{j,t-1} - \beta\tau_{j,t} = \beta\,(1-\rho)\,T_{j,t} - \beta\,(1-\rho)\,(T_{j,t} - \beta\tau_{j,t+1})\,.$$

Forward substitution gives

$$\begin{aligned} T_{j,t-1} - \beta\tau_{j,t} &= \beta\,(1-\rho)\sum_{s=0}^{\infty}\left(-\beta\,(1-\rho)\right)^s T_{j,t+s} \\ &\quad - \beta\,(1-\rho)\lim_{T\to\infty}\left(-\beta\,(1-\rho)\right)^T \left(T_{j,t+T} - \beta\tau_{t+T+1}\right), \end{aligned}$$

where the last term is zero.

problem with choice variables τ_1 and $\{T_{h,t}, T_{l,t}\}_{t=1}^{\infty}$. The first-order conditions for this problem imply a linear system of equations yielding unique solutions for τ_1, $T_{h,0}$, and $T_{l,0}$ (with $T_{h,0} > T_{l,0}$) in terms of primitives and the shadow values of the two budget constraints, λ_h and λ_l. The the sequences of present-value taxes, $\{T_{h,t}, T_{l,t}\}_{t=1}^{\infty}$, can be shown to satisfy

$$T_{h,t} - T_h^* = -\left(T_h^* - T_0^e + \frac{\lambda_h - \lambda_l}{2\lambda_h - 1}(1-p)T_{l,0} \right)(-(1-\rho))^t, \tag{22}$$

$$T_{l,t} - T_l^* = -\left(T_l^* - T_0^e - \frac{\lambda_h - \lambda_l}{2\lambda_l - 1}pT_{h,0} \right)(-(1-\rho))^t. \tag{23}$$

Clearly, if $0 < \rho < 1$, the sequences $\{T_{h,t}, T_{l,t}\}_{t=1}^{\infty}$ converge in an *oscillatory*fashion to their respective limits T_h^* and T_l^*, where $T_j^* \equiv \kappa(\lambda_j - 1)/(2\lambda_j - 1)$ for $j \in \{h, l\}$. If $\rho = 1$ oscillations do not die out. The difference equations (22)-(23) and τ_1 provide a complete characterization of the optimal state-contingent present-value taxes. Given τ_1 and $\{T_{h,t}, T_{l,t}\}_{t=1}^{\infty}$, the tax sequences $\{\tau_{j,t}\}_{t=2}^{\infty}$ can be backed out using (recursively) the expression;

$$\tau_{j,t+1} = \frac{T_{j,t-1} - \beta\tau_{j,t}}{\beta^2(1-\rho)},$$

where $\tau_{h,1} = \tau_{l,1} = \tau_1$. This yields

$$\tau_{j,t+1} = \tau_j^* - (1-\rho)\left(\tau_{j,t} - \tau_j^*\right), t \geq 2, \tag{24}$$

where $\tau_h^* \equiv (\lambda_h - 1)/(2\lambda_h - 1)$ and $\tau_l^* \equiv (\lambda_l - 1)/(2\lambda_l - 1)$.

Figure 3 shows a numerical example. The upper panel shows that present-value taxes $(T_{j,t})$, and thus investments, oscillate in both states of nature. The right panel shows the actual tax sequence that implements the optimal allocation. Tax oscillations arise in both states of nature, even though there is no inelastic capital. Had the state of nature been known in advance, the planner would have chosen a constant sequence $\tau_h^* = 1/3$ in the high-spending state and $\tau_l^* = 0.3$ in the low-spending state, respectively. However, due to uncertainty, τ_1 must be set at a level producing an intermediate investment level in period zero. In fact, $T_0^e = 0.397$ and is thus in between the two steady-state levels of present-value

taxes (0.419 and 0.377, respectively). If the high-spending state is realized, the planner sets $\tau_{h,2} > \tau_h^*$. The reason is twofold: first, the government faces larger spending needs than expected; second, taxes have turned out to be lower than what agents born in period zero had expected and based their investment upon. In other words, i_0 is higher than the steady-state investments under large financing needs. Then, distortion smoothing requires this generation to be taxed more heavily in the second period. In contrast, if the low state is realized, the planner sets $\tau_{l,2} < \tau_l^*$ and the mirror image of the argument a\above applies. From period three and onwards, taxes continue to oscillate following the dynamics characterized in the deterministic case of Proposition 2. For instance, the high (low) $\tau_{h,2}$ ($\tau_{l,2}$) tends to distort the investment of the generation born in time one heavily (lightly). Thus, distortion smoothing requires a high (low) $\tau_{h,3}$ ($\tau_{l,3}$), and so on.

< FIGURE 3 ABOUT HERE >

4.2. Complete markets

The results in the previous subsection can be compared with those in an environment of complete markets, where there exist markets for state-contingent assets paying one unit of the consumption good conditionally on the realization of the high-spending or the low-spending state. Let period-one consumption be the *numéraire* and define $q_{j,t}$ as the Arrow-Debreu price of the consumption good in period t and state j. The assumption of a one-period implementation lag in taxes is maintained. The two government budget constraints given by (16) can now be consolidated into one constraint:

$$b_0 + g_0 + \sum_{t=1}^{\infty} \sum_{j \in \{h,l\}} q_{j,t} g_{j,t+1} = \beta \tau_1 i_0 + \sum_{t=1}^{\infty} \sum_{j \in \{h,l\}} q_{j,t} \tau_{j,t+1} \left(\tau_{j,t+1} (h_{j,t+1}^o + i_{j,t}) \right). \qquad (25)$$

Since individual utility is linear in consumption, it follows that the Arrow-Debreu prices must be given by the discounted probabilities, i.e., that $q_{h,t} = \beta^t p$ and $q_{l,t} = \beta^t (1-p)$.

Hence,

$$b_0 + g_0 + \sum_{t=1}^{\infty} \beta^t \left(g_{t+1}^e - \tau_{t+1}^e \left(\tau_{t+1}^e (h_{t+1}^{o,e} + i_t^e) \right) \right) = 0, \tag{26}$$

where variables with superscript e denote expected values: $x^e \equiv p x_h + (1 - p) x_l$. Since there is only one budget constraint, the Ramsey plan simplifies to

$$\max_{\{T_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t y_{CM} (T_t) = \sum_{t=0}^{\infty} \beta^t \left(\lambda_{CM} \kappa i (T_t) - i (T_t)^2 (2\lambda_{CM} - 1) \right),$$

where λ_{CM} denotes the multiplier associated with the complete-market budget constraint (26). Clearly, the solution features $y'_{CM} (T_t) = 0$ for all t, namely, constant present-value taxes and investment.

Intuitively, under complete markets the government can achieve perfect distortion smoothing by letting private agents bear all the spending risk. The resulting allocation is identical to one where the government efficiently collects just enough resources to satisfy its spending needs in expectation and then uses lump-sum taxes to cover additional needs in the high-spending state and to rebate the surplus to the private agents in the low-spending state. Note that the assumption of risk-neutrality ensures that agents are prepared to own a portfolio of state-contingent debt that makes them act as insurers of the government.

5. Final remarks

We have shown, using a modified neoclassical growth model, that a benevolent government can find it optimal to make the sequence of capital income tax rates oscillatory in order to finance a given stream of expenditures at a minimal cost to consumers. Three assumptions underlie this result. First, depreciation rates for capital are increasing in age, as opposed to constant. Second, the government cannot apply different tax rates to income from different vintages of capital. Third, the government has commitment to set future tax rates. We now make brief comments on each of these assumptions.

Our main argument rests on the assumption that depreciation rates increase as capital ages. We do believe that this captures essential features of the evolution of human capital: it tends to "die" with worker retirement, as well as to some extent when workers switch tasks (to the extent that human capital is task-specific). There is also substantial empirical evidence that the depreciation rate of many physical assets is increasing with age. A seminal study by Coen (1975) estimates capacity depreciation for equipment and structures in 21 industries and finds a predominant pattern of depreciation increasing with age. In many cases, capital depreciation is found to be of the *one-hoss shay* variety, i.e., capital maintains its full capacity until when it is scrapped. Similar results are obtained by Penson et al. (1977) and by Pakes and Griliches (1984), who find that the productive value of investments is actually increasing over the first three years and remains constant for the following four to five years. The evidence for increasing depreciation rates is particularly sharp in the case of IT technologies (see e.g. Whelan (2002), Geske, Ramey, and Shapiro (2003), and Dunn et al. (2004)).[25]

If the government could apply vintage-specific tax rates, the taxation problem would become trivial: the planner could expropriate pre-installed capital and attain perfect distortion smoothing on new investments. Such a conclusion follows independently of the depreciation structure. In particular, taxation in the standard Chamley-Judd framework would not feature any dynamics either.[26] The motivation for ruling out vintage-specific taxation by assumption is that we believe that it is difficult in practice to distinguish when existing

---

[25]In contrast, studies based on second-hand asset prices argue that geometric decay is a good model of economic depreciation (see Hulten and Wykoff, 1981). We believe, however, that the price of second-hand capital is a poor proxy for the internal productive capacity of installed capital (which is the relevant notion for our analysis), since this is affected by private information and adverse-selection issues. Moreover, there is some variation in results across studies using second-hand prices. For example, Oliner (1996) finds that economic depreciation for machine tools is significantly increasing with age.

[26]In this case, all revenue generated by pre-installed capital could always be fully captured by the government. Thus, in every period the government would have a separate tax rate for that income which originates in investment prior to period zero. This rate could be bounded at any point in time, but taxation of the initial base for capital income would then continue until it is exhausted.

capital was built. For human capital, the timing of education is observable, but the timing of later investments in human capital (on and off the job), and their importance relative to educational investments, are for the most part not observed. For physical capital, though initial investment amounts might be measured by tax authorities, later adjustments in the form of maintenance and upgrades are difficult to assess. Moreover, a feature of many forms of investments is that they have a consumption component. This is obvious for the case of education, but it is arguably the case also for many other investment activities. Thus, with substantial investment subsidies, the difficulty for fiscal authorities of sorting out the consumption component from true productive investments arguably make such subsidies quite imperfect tools. A more thorough treatment relying explicitly on information asymmetries would be an interesting extension to the present work.[27]

Finally, what if the government could not commit to its future tax rates? Then taxes would indeed be set differently, unless one could invoke reputational mechanisms: the commitment equilibrium is time inconsistent, for reasons standard to capital taxation problems. In a working paper version of the this paper, we show that the lack of commitment implies a natural tendency for taxes not to fluctuate or, at least, to fluctuate less (see also Hassler et al., 2005, for a similar result with a politico-economic interpretation). When there is no commitment or commitment is imperfect (as in Debortoli and Nunes, 2006), the government's trade-off between costs and benefits changes. As a general principle for both the case with and that without commitment, the excess value of government funds times the marginal revenue of taxes at period t is set equal to the marginal distortionary cost of taxes in period t. Under commitment, the marginal distortionary cost depends on a weighted sum of the wedges between first-best and actual investments levels prior to t, where the weights are determined by the depreciation structure. In contrast, if, due to a lack of commitment,

---

[27]The assumption that the government has no access to age-dependent taxes, upon which our results depend, has been adopted elsewhere in the literature; see, for instance, Erosa and Gervais (2002).

the government sets the tax for t no earlier than in period $t-1$, the marginal cost of taxes in period t depends only on the investment wedge in period $t-1$, since all previous investments are then sunk, thus making decisions in different periods more similar. In the smooth Markov-perfect (limit-of-finite-horizon) equilibrium we look at, this leads to a dampening, or complete elimination, of the fluctuations we find to be optimal under commitment. In conclusion, the policy implications can differ substantially depending on the extent of government commitment. Thus, our model suggests one avenue for testing the extent to which commitment is present, at least if the other maintained assumptions of our analysis are met.

6. References

Atkeson, A., Chari,V.V., Kehoe, P., 1999. Taxing capital income: A bad idea. Federal Reserve Bank of Minneapolis Quarterly Review 23, 3–17.

Barro, R. J., 1979. On the determination of public debt. Journal of Political Economy 87, 940–971.

Coen, R. M., 1975. Investment behavior, the measurement of depreciation, and tax policy. American Economic Review 65, 59–74.

Chamley, C., 1986. Optimal taxation of capital income in general equilibrium with infinite lives. Econometrica 54, 607–622.

Correia, I., 1996. Should capital income be taxed in the steady state? Journal of Public Economics 60, 147-151.

Debortoli, D., Nunes, R., 2006. Loose commitment. Mimeo, Universitat Pompeu Fabra.

Dunn, W., Doms, M., Oliner, S., Sichel, D., 2004. How fast do personal computers depreciate: Concepts and new estimates. NBER Working Paper No. 10521.

Erosa, A., Gervais, M., 2002. Optimal taxation in life-cycle economies. Journal of Economic Theory 105, 338–369.

Geske, M. J., Ramey, V., Shapiro, M. D., 2003. Why do computers depreciate? Mimeo, University of Michigan.

Greulich, K., Marcet, A., 2007. Optimal capital and labor taxes with heterogeneous agents: Making everybody happy. Mimeo, University of Zurich.

Hagedorn, M., 2007. Optimal Ramsey tax cycles. Mimeo, European Central bank.

Hassler, J., Krusell, P., Storesletten, K., Zilibotti, F., 2004. On the timing of optimal capital taxation. CEPR Working Paper No. 4731.

Hassler,J., Krusell, P., Storesletten, K., Zilibotti, F., 2005. The dynamics of government. Journal of Monetary Economics 52, 1331-1358.

Hassler, J., Storesletten, K., Zilibotti, F., 2007. Democratic public good provision. Journal of Economic Theory 133, 127-151.

Hulten, C. R., Wykoff, F.C., 1981. The estimation of economic depreciation using vintage asset prices: An application of the Box-Cox power transformation. Journal of Econometrics 15, 367–396.

Judd, K. L., 1985. Redistributive taxation in a simple perfect foresight model. Journal of Public Economics 28, 59–83.

Lucas, R. E., Stokey, N., 1983. Optimal monetary and fiscal policy in an economy without capital. Journal of Monetary Economics 12, 55–94.

Oliner, S. D., 1996. New evidence on the retirement and depreciation of machine tools. Economic Inquiry 34, 57–77.

Pakes, A., Griliches, Z., 1984. Estimating distributed lags in short panels with an application to the specification of depreciation patterns and capital stock constructs. Review of Economic Studies 51, 243-262.

Penson , J. B., Hughes, D. W., Nelson, G.L., 1977. Measurement of capacity depreciation based on engineering data. American Journal of Agricultural Economics 59, 321–329.

Whelan K., 2002. Computers, obsolescence and productivity. Review of Economics and

Statistics 84, 445–461.

7. Appendix

7.1. *Derivation of equation (11)*

From the definition

$$T_t \equiv \beta\tau_{t+1} + (1 - \rho\delta) \sum_{s=2}^{\infty} \beta^s (1 - \delta)^{s-2} \tau_{t+s}$$

it follows immediately that

$$T_{t-1} - \beta\tau_t + (T_t - \beta\tau_{t+1}) \beta\delta (1 - \rho) = \beta (1 - \rho\delta) T_t. \tag{27}$$

Forward substitution implies

$$
\begin{aligned}
T_{t-1} - \beta\tau_t &= \beta (1 - \rho\delta) \sum_{s=0}^{\infty} (-\beta\delta (1 - \rho))^s T_{t+s} + \\
&\quad \lim_{T \to \infty} (-\beta\delta (1 - \rho))^T (T_{t+T} - \beta\tau_{t+T}) \\
&= \beta (1 - \rho\delta) \sum_{s=0}^{\infty} (-\beta\delta (1 - \rho))^s T_{t+s},
\end{aligned}
$$

where $\lim_{T \to \infty} (-\beta\delta (1 - \rho))^T (T_{t+T} - \beta\tau_{t+T}) = 0$, since taxes are bounded, implying that their PDVs (in particular the T_ts) are also bounded. In particular, the expression above implies that

$$T_0 = \beta\tau_1 + \beta (1 - \rho\delta) \sum_{s=0}^{\infty} (-\beta\delta (1 - \rho))^s T_{s+1}. \tag{28}$$

Recall that, by definition, $T_0 \equiv \beta\tau_1 + (1 - \rho\delta) \sum_{s=2}^{\infty} \beta^s (1 - \delta)^{s-2} \tau_s$. This, together with equation (28), implies that

$$\sum_{s=2}^{\infty} \beta^s (1 - \delta)^{s-2} \tau_s = \beta \sum_{s=0}^{\infty} (-\beta\delta (1 - \rho))^s T_{s+1}. \tag{29}$$

Finally, rearranging the expressions for $\hat{T}_0$ and $\hat{T}_0 = \sum_{s=1}^{\infty} \beta^s (1 - \delta)^{s-1} \tau_s$ leads to

$$\hat{T}_0 = \beta\tau_1 + (1 - \rho\delta) \sum_{s=2}^{\infty} \beta^s (1 - \delta)^{s-2} \tau_s - \delta (1 - \rho) \sum_{s=2}^{\infty} \beta^s (1 - \delta)^{s-2} \tau_s,$$

which, in turn, can be rewritten, using (28)-(29), as

$$\hat{T}_0 = T_0 + \sum_{s=1}^{\infty} \left(-\beta\delta\left(1-\rho\right)\right)^s T_s,$$

which is expression (11) in the paper.

7.2. Details of the proof of Proposition 2

Solving (27) for τ_{t+1} yields

$$\tau_{t+1} = \frac{T_{t-1} - \beta\tau_t - T_t\beta\left(1-\delta\right)}{\beta^2\delta\left(1-\rho\right)}.$$

Using (13) and the expression for τ^* given in the text to replace T_{t-1} and T_t yields (for $t \geq 1$)

$$\tau_{t+1} = \tau\frac{1+\beta\delta\left(1-\rho\right)}{\beta\delta\left(1-\rho\right)} - \tau\frac{1+\beta\left(1-\delta\right)\delta\left(1-\rho\right)}{\delta^2\left(1-\rho\right)^2\beta^2}2h_1\left(-\delta\left(1-\rho\right)\right)^t - \frac{\tau_t}{\beta\delta\left(1-\rho\right)}.$$

The complete solution to this difference equation can be written

$$\tau_t = \tau^* + \frac{1+\beta\delta\left(1-\delta\right)\left(1-\rho\right)}{\beta\left(1-\beta\delta^2\left(1-\rho\right)^2\right)}2h_1\tau\left(-\delta\left(1-\rho\right)\right)^{t-1} + c\left(-\frac{1}{\beta\delta\left(1-\rho\right)}\right)^t,$$

where c is an arbitrary integration constant. The interpretation of the arbitrary c is that there is an infinite number of tax sequences that implement the optimal allocation. However, since the root of the homogeneous part, $-1/\left(\beta\delta\left(1-\rho\right)\right)$, is outside the unit circle, the constraint $\tau_t \in [0,1]$ is not satisfied for $c \neq 0$. Thus, the only feasible solution to (13) is determined by setting $c = 0$. Writing this solution recursively yields the solution in Proposition (2).

The non-diverging dynamics implies that it is sufficient that τ_1 be bounded for guaranteeing a uniformly bounded τ_t. Clearly, this condition is satisfied.
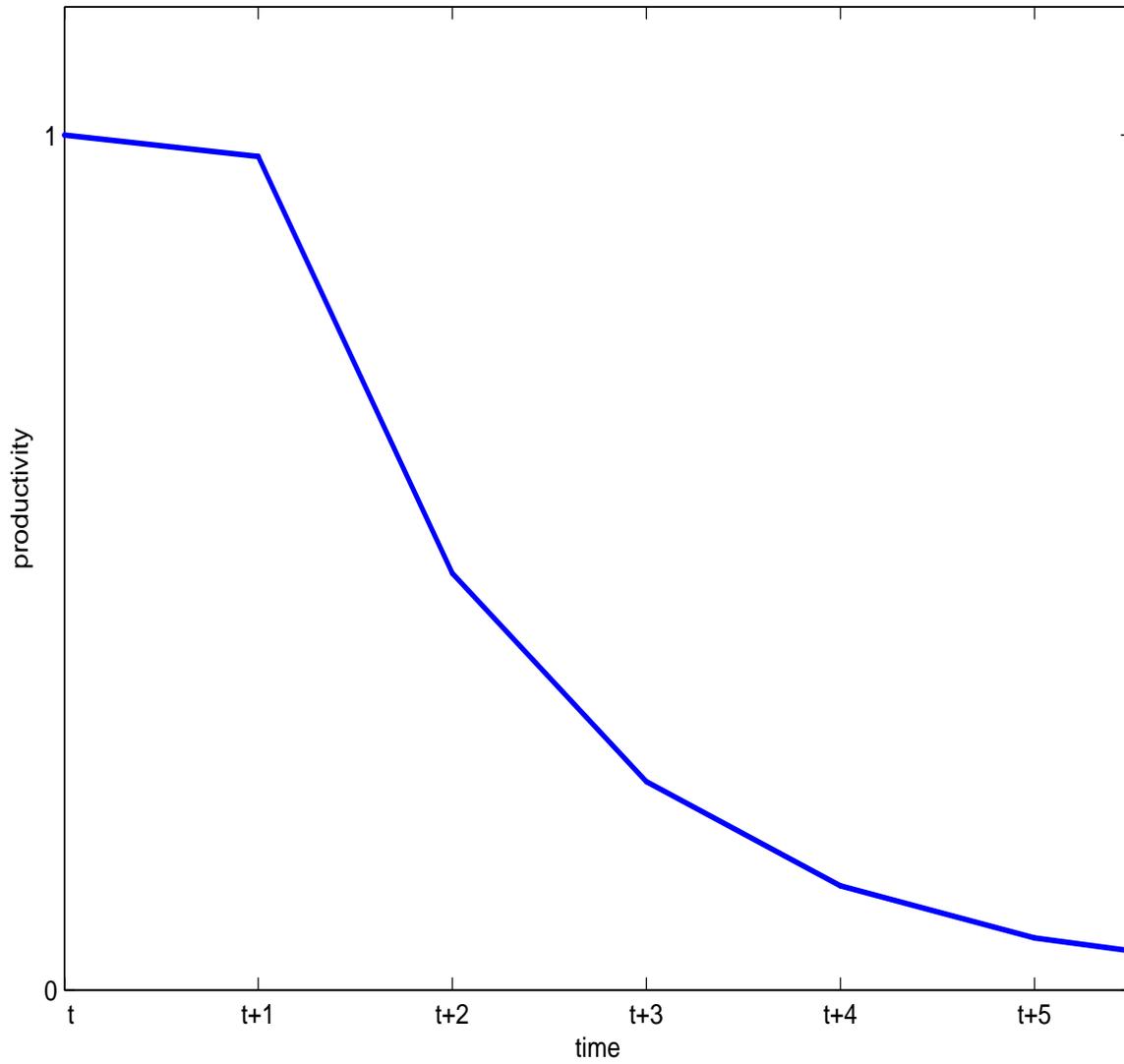
Figure 1: Remaining stock of capital installed in period $t - 1$ with quasi-geometric depreciation ($\rho \in (0, 1)$). The parameter values in the example are $\rho = 0.05$ and $\delta = 0.5$.
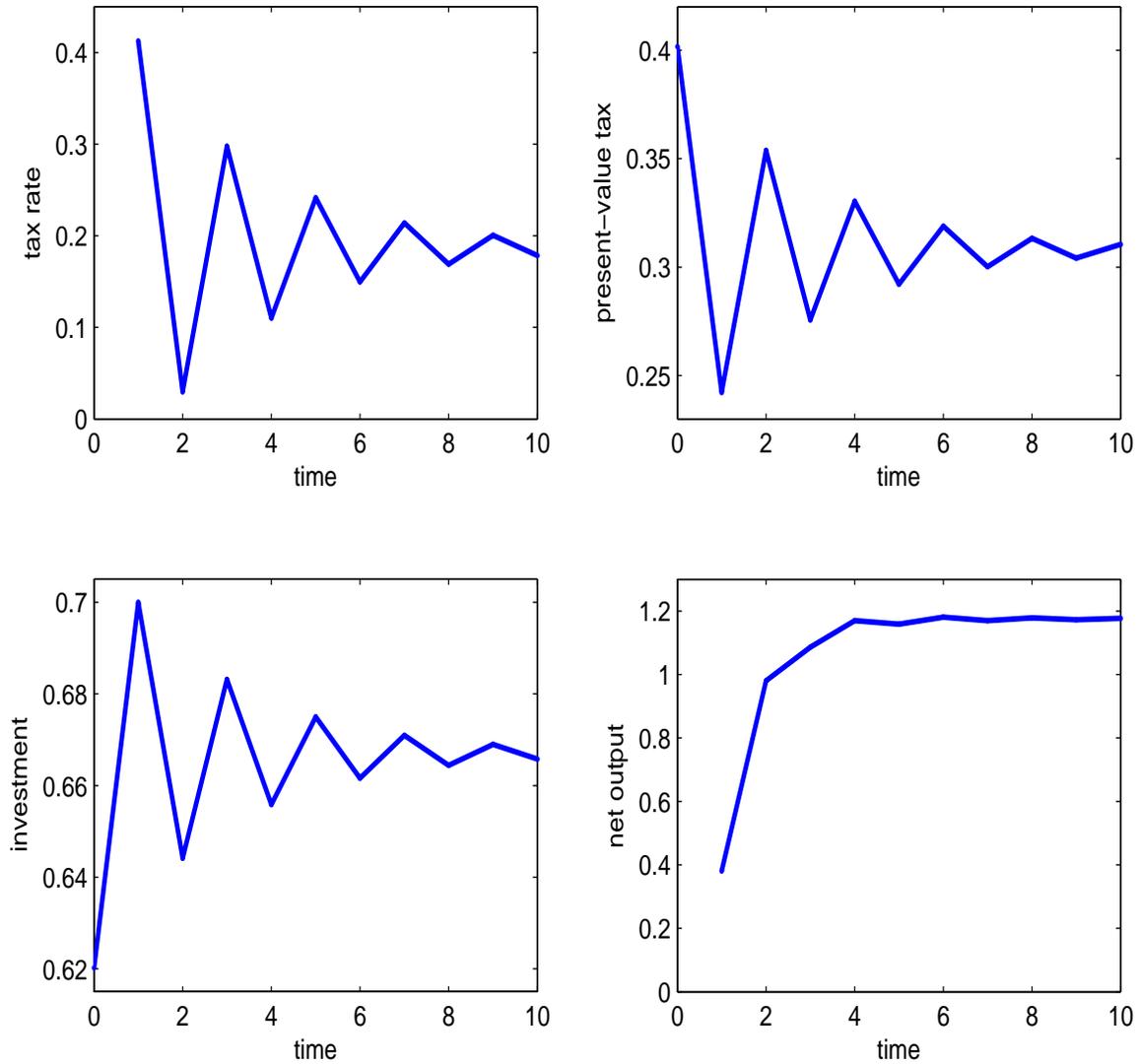
Figure 2: Ramsey dynamics – an example with accelerating depreciation. The figure displays the dynamic evolution of tax rates (τ_t), present-value taxes (T_t), investments (i_t), and net output (y_t) in the optimal Ramsey allocation of Proposition 1. The parameter values underlying the figures are $\delta = 0.7$, $\rho = 0$, $\beta = 0.8$, and an initial stock of installed capital of $h_1^o = 0.25$. Moreover, the government expenditure are such that the Lagrange multiplier is $\lambda = 1.3$.
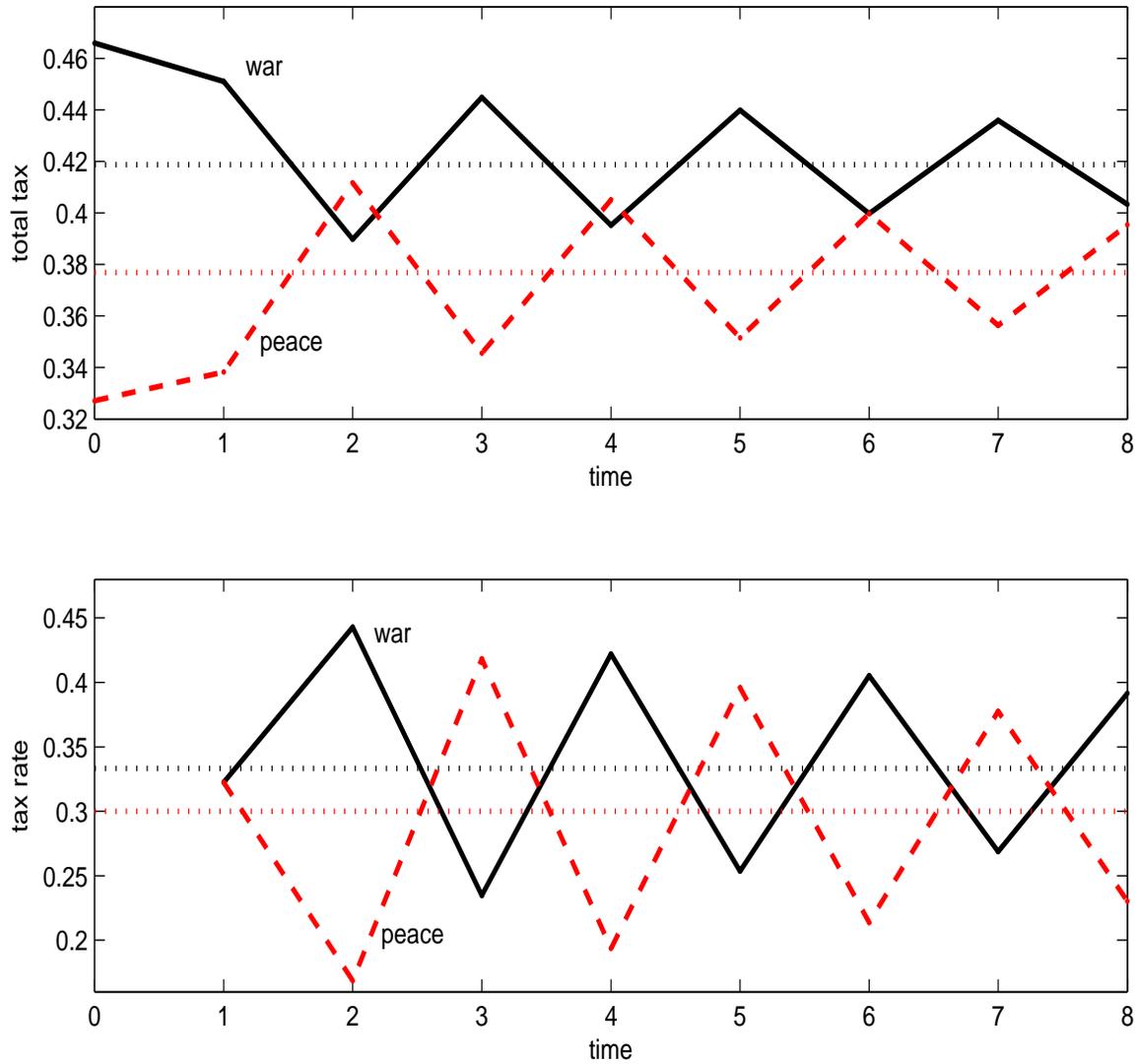
Figure 3: Upper panel displays the present-value taxes in case of high $T_{h,t}$ (black) and low $T_{l,t}$ (red) spending requirement. The lower panels displays the associated sequence of tax rates. The dotted lines represent steady-state values for taxes conditional on war and peace, respectively. The parameter values underlying this example are $\rho = 0.1$, $p = 0.5$, and $\beta = 0.75$. Moreover, the government expenditures associated with war and peace are chosen so that the Lagrange multipliers become $\lambda_h = 2$, and $\lambda_l = 1.75$, respectively.