

ON THE OPTIMAL TIMING OF CAPITAL TAXES

John Hassler,^{*} Per Krusell,[†]

Kjetil Storesletten,[‡] and Fabrizio Zilibotti,[§]

October 2004

Abstract

This paper analyzes the optimal timing of taxes on capital income. We show that the celebrated result that taxes should be front-loaded with an initially high tax followed by a discrete jump to the steady state is knife-edge, hinging on capital having a constant depreciation rate. An empirically supported deviation from this case, involving depreciation rates that increase over the lifespan of the investment, implies that optimal taxes should oscillate. Furthermore, the optimality of fluctuating tax rates hinges on the government being able to commit to the path of future tax rates. Without commitment, optimal taxes may be smooth also under accelerating depreciation. In a calibrated example, we find that optimal taxes are oscillating under commitment and smooth without commitment.

JEL Classification Codes: D90, E61, E62, H21, H30.

Keywords: Optimal taxation, Tax dynamics, Time-consistency, Capital depreciation.

^{*}John Hassler: IIES and CEPR; e-mail: John.Hassler@iies.su.se.

[†]Per Krusell: Princeton University, IIES, CAERP, CEPR, and NBER; e-mail: pkrusell@princeton.edu.

[‡]Kjetil Storesletten: University of Oslo, IIES, Frisch Centre and CEPR; e-mail: kjetil.storesletten@econ.uio.no.

[§]Fabrizio Zilibotti: IIES, IFS and CEPR; e-mail: Fabrizio.Zilibotti@iies.su.se.

1 Introduction

What is the optimal path of taxes for a benevolent government that needs to finance some essential public expenditures? Regarding this classical question, we make two new points in this paper. The first is that, in very natural settings, the policy maximizing consumers' welfare in a non-stochastic economy involves fluctuations in taxes and expenditures. The second point is that the size of the fluctuations depends on the extent to which the government can commit to how to set future taxes: less commitment leads to a dampening of the fluctuations. What we mean by a "natural setting" is, primarily, that installed capital loses productivity at a rate that increases as capital ages, and, in addition, that the government has limited ability to differentiate taxes across different types (vintages) of capital.

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 (e.g., Barro, 1979). However, if taxes distort accumulation decisions, new issues arise. One important such issue 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. More precisely, consider a model where a benevolent "Ramsey planner" chooses how and when to finance a given stream of government expenditures (or a public good) using proportional taxes at different points in time. Under standard assumptions on preferences and technology, the optimal tax sequence then prescribes high taxes on capital income for a finite number of periods. I.e., capital taxes should be "front-loaded" and zero in the long run (see Atkeson et al., 1999).

In this paper, we consider a version of the standard neoclassical growth model. We generalize this model by allowing empirically supported deviations from geometric depreciation and convex costs of capital accumulation. To be able to fully characterize optimal tax-dynamics, we then assume preferences and technology to be linear and separable in capital and labor. When depreciation is geometric, our model reproduces the standard result that taxes on capital should be front-loaded. Suppose, for simplicity, that it takes one period for the government to implement a taxation decision, so that a decision to tax capital initial income will distort the investment in the first period. Then the result is that the planner taxes initial capital income 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 "jumps down" 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 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, and it turns out that the opposing forces cancel exactly so that taxes go to their steady-state level immediately.⁵

The first main finding in our paper is that the optimality of constant taxes after the first period is a knife-edge result, hinging on the specific (albeit commonly used) assumption about the depreciation of assets in the economy: geometric depreciation. If assets depreciate at a non-constant rate over time (i.e., depreciation deviates from a geometric pattern), 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

⁵Intuitively, if the PDV 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.

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. A simple example can be used to illustrate the intuition. Suppose that the asset is accumulated in period t and is fully productive in periods $t+1$ and $t+2$ 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%). To attach words to the assumption in this example, we can think of agents who live for three periods and in the first period make an educational choice whose benefits remain with the worker in the two following periods, i.e., until he exits the labor market. At time t , a surprise occurs, which increases the need for the government to raise funds.⁶ To fix ideas, let this be a surge in an external security threat (e.g., “terrorism”), causing an increase in the social value of defense. In this case, the planner wants to seize the opportunity to extract a large tax revenue from the generation who made its investment before the surge of the threat. This generation sunk its investment under the expectation of lower taxes, and this investment is, at t , an inelastic tax base. So a high tax rate is called for at t . This high tax rate can be counteracted by a lower tax rate in $t+1$ so that investments in period t are not too distorted. Then, since the $t+1$ tax rate is low, the government can afford a higher $t+2$ rate, 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 increasing with its age.

Throughout the paper, we maintain the assumption that the government cannot tax assets at rates that vary with the age of the capital. If this were possible (or if the government could offset distortions by investment subsidies), 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. The motivation for simply ruling out vintage-specific taxation by assumption is that we believe that it is difficult in practice to distinguish when existing 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. Formally, one could therefore defend the assumption that vintage-specific taxation is not allowed with a framework based on explicit information asymmetries between firms and the fiscal authorities. In order to make the analysis as focused as possible, we simply rule out vintage dependence. A more realistic framework would allow some vintage specificity and possibly derive the extent of such specificity from a microeconomic structure based on information asymmetries.

Having established that tax oscillations are a property of the Ramsey solution, we consider the effect of commitment on tax dynamics. Time-inconsistency is a standard feature of the optimal capital taxation literature which is also present in our model: the tax rate at a point in time t influences investment decisions prior to period t , and these effects should be taken into account when a plan for the time t tax rate is made; when period t arrives, however, these investment decisions are already made, thus causing the government to revise its tax plan, if it can. While the Ramsey planner in ideal model worlds are often assumed to have the ability to commit to a tax sequence, real-world institutions do not automatically have the power to commit. In particular, democratic rules naturally limit the government’s ability to commit to future policy choices: commitment, were it possible to arrange, would be a way of disenfranchising future voters. But more generally it is not at all clear how commitment to policy could be implemented in practice, a point

⁶The assumption in this example—that the change is a “surprise”—is made for simplicity. It can be thought of as allowing the planner to make a commitment but then re-optimize after a zero-probability shock realization. We show in section 6 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.

first made by Kydland and Prescott (1977). A number of “mechanisms” for solving commitment problems have been proposed in the literature. Some appeal to reputation equilibria, along the lines of Abreu, Pearce, and Stacchetti (1990) and Chari and Kehoe (1990); we rule out reputation mechanisms by focusing on finite-horizon equilibria and their limits—formally, we look at Markov-perfect equilibria.⁷ Others, such as Rogoff (1985), appeal to specific institutional structures involving delegation; we do not allow committing to delegation here. Finally, Lucas and Stokey (1983) point to sophisticated government portfolio design as a way of overcoming some forms of commitment problems; such instruments are not available in our setup. In our view, it is therefore important to explore what optimal taxation would look like if it took into account the restrictions implicit in not being able to commit.

We show that the lack of commitment implies a natural tendency for taxes not to fluctuate or, at least, to fluctuate less. When there is no commitment, 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, 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. Specifically, when capital is productive for two periods, the marginal distortionary cost of a tax in period $t > 1$ depends on the *sum* of investment wedges in periods $t - 1$ and $t - 2$. This sum should be proportional to the marginal revenue of the tax in period t , leaving open for the individual investment wedges to fluctuate. Under no commitment, the optimal solution instead requires each investment wedge to be proportional to the marginal revenue of the tax. This allows some movements in tax rates because of the initial sunk capital stock, but the fluctuations are minor.

Our main argument rests on the notion that depreciation rates are increasing in age, as opposed to constant. Clearly, some extent of accelerating depreciation seems justified just based on the plausible idea that there is a finite upper bound on the lifetime of any asset. More importantly, however, there is systematic empirical support for our assumption. A number of studies suggest that the productive capacities of many assets do not fall at a constant rate but, rather, decrease with age, which is the case we emphasize here. 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 1981) using engineering data to estimate capacity depreciation for farm tractors. Pakes and Griliches (1982) 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. This could be explained by learning-by-doing: capital is not used at its full potential until some time from its installation. In contrast, a number of studies based on cross-sectional studies of second-hand asset prices, most notably Hulten and Wykoff (1981), conclude that geometric decay is a good approximation for economic depreciation, and show that in some cases depreciation rates are actually *higher* for young than for old capital. In our view, however, the methodology used by Hulten and Wykoff to estimate depreciation is unlikely to capture the right notion of loss of productive capacity for our analysis. 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), and it is affected by private information and adverse-selection issues, as well as capital specificity (see, e.g., Ramey and Shapiro, 1998) and learning-by-doing. In addition, some recent studies using this methodology in fact find different results; e.g., Oliner (1996) finds that economic depreciation for machine tools is significantly increasing with age.⁸ The evidence that depreciation rates increase with age is particularly sharp in the case of IT technologies. Geske, Ramey, and Shapiro (2003) find that a large

⁷For a recent study of a class of reputation equilibria, see Fernandez-Villaverde and Tsyvinski (2004).

⁸This study is based on data on non-numerically controlled machines collected from a survey of dealers belonging to the Machinery Dealers National Association. The price-age profile is estimated by taking a weighted average of the observed price of a machine and the unobserved zero price, with the weights reflecting the probability of remaining in service at a given age. Allowing depreciation rates to change over time, as in Hulten and Wykoff (1981), Oliner (1996) finds that the yearly depreciation rate is 2.9% after ten years, 6.1% after 20 years, 11.1% after 30 years and 18.1% after 40 years. Imposing geometric depreciation would yield yearly depreciation rate equal to 9.5%, but the hypothesis that the depreciation rate is constant is strongly rejected in the data.

part of the loss of value of computers is due to technological obsolescence, while there is very little effect of loss of productive capacity due to age. In a recent study on the age-price profile of personal computers, Dunn et al. (2004) estimate the depreciation rates of personal computers to be close to zero in the first year, and then to increase to 17%, 25%, and 30%, respectively, in the three subsequent years.⁹ Whelan (2002) estimates a structural model where computer maintenance is costly and scrapping endogenous and finds even more compelling results: before scrapping, the productivity decay is insignificant, as in the “one-hoss shay” pattern, for most computer equipment, with the exception of printers for which geometric decay is not rejected. In summary, the assumption that depreciation rates increase with age seems to be in accordance with the bulk of the empirical evidence.

We are not aware of systematic studies on the depreciation of human capital. However, it seems clear that to the extent to which human capital accumulation is important, our case is significantly strengthened: acquired knowledge and skills embodied in live, working people is hard to think of as geometrically decaying, since they effectively vanish as the worker exits the labor market.

Our analysis relates to Barro’s (1979) result that tax smoothing is optimal. Barro looked at debt- versus tax-finance of a given stream of expenditures. Our model differs from Barro’s in one main respect: in Barro, the distortionary effects of taxation have a static (e.g., labor supply) rather than a dynamic nature. Our result that fluctuations in taxation can be optimal emphasizes an unexplored aspect of the general principle that the smoothing should occur for distortions, not for taxes. In particular, in our model, the distortions to an agent’s effort choice can be summarized by the *present value* of extra taxes incurred by the effort choice: whether to become educated (a higher-earning career, presumably) or not depends on what one thinks will happen with the taxes over the entire course of one’s working life. So a fluctuating tax rate on income is not bad per se and, as we show, is desirable in order to implement a higher taxation of already installed (more inelastic) sources of income. Another difference between our setup and Barro’s is that in our model, the structure of optimal taxation depends on whether or not there is commitment, whereas in Barro’s setup, because the effort decision is static and because interest rates are exogenous, the commitment outcome is time-consistent.¹⁰

Our paper has implications for political economy. In a related paper (Hassler et al., 2004), we build on the insights here to show that the time-consistent Ramsey solution can be given politico-economic micro-foundations: if agents vote over redistribution (or, like in part of this paper, on public good provision) with altruism towards future generations, and the political mechanism is represented by a probabilistic voting model a la Lindbeck and Weibull (1987), then the time-consistent Ramsey solution is indeed a political equilibrium. If we attach a politico-economic interpretation to the time-consistent solution of the model in this paper, we can state that, to the extent cycles are present, they are not “politically driven”, but rather counteracted by politics.

Our paper is also related to the recent study of public expenditure choice in Klein, Krusell, and Ríos-Rull (2003). Their model is a neoclassical growth setup where the government has no access to debt and has no commitment; like in this paper, they focus on Markov-perfect equilibria. The focus there is on (i) deriving and interpreting first-order conditions for the government and (ii) numerical methods and a quantitative evaluation. The present paper is different not mainly because it derives closed-form solutions but because it emphasizes conditions under which non-monotonic dynamics arise (and which are plausible). The neoclassical framework in Klein et al.’s work uses only physical capital, with geometric depreciation, thus ruling out the possibility of oscillations in the solution with commitment.

In Section 2, we describe the 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. Then in Sections 2.5 and 3 we describe and analyze a problem which is formally equivalent to the former one under the specific assumption of our model: that of choosing public expenditures subject to a balanced budget. We compare the commitment and no-commitment solutions in Section 4. In Section 5 we then interpret our findings on the basis of the idea of “distortion smoothing”. The issue of whether

⁹Our calculation is based on Table 7a in their paper. These figures do not include what the authors refer to as the “revaluation effect”, i.e., the fact that technical progress makes better computers available, causing a fall over time of the constant-quality price of a PC. Since we are interested in the productivity of the asset, this effect should not be included. Including revaluation continues to give an increasing depreciation rate, although the differences are less pronounced.

¹⁰A closely related setup to Barro’s, namely, that in Lucas and Stokey (1983) features time inconsistency since it features endogenous interest rates.

the fluctuations in our examples are all memories of the initial period is then dealt with in Section 6 in a very simple extension of the basic setup to uncertainty in the second period. There we show that, if government debt is not state-contingent, “new” fluctuations occur as a result of the shock whereas if it is state-contingent, no new fluctuations occur: those that are present are indeed a memory of the initial period. Section 7 concludes.

2 The model

In order to relate our results to the literature on optimal taxation (see, e.g., Chamley (1986), Judd (1985), or Atkeson, Chari, and Kehoe (1999)), we initially describe a very standard setup. We then both specialize and generalize this setup. The generalizations are the focus of our analysis; the key generalization involves a non-constant depreciation rate. We also consider “adjustment costs” to capital formation, but this specification is made mainly for reasons of tractability. The specializations are made only in order to make the analysis as tractable as possible. First, we consider linear utility in private consumption—in order to avoid issues of government manipulation of the interest rates in pursuing optimal taxation.¹¹ Second, our formulation is made in terms of an abstract capital good, which could be either physical or human capital; however, we do not model static labor supply (we assume that production is linear in capital). The reason for this is that our main focus is on the timing of taxes on accumulable factors, and the presence of a tax on purely statically provided input has no direct impact on this analysis.

An additional motivation for our modelling choices is that, under our assumptions, the model is observationally equivalent to one where the government chooses public expenditures over time. More precisely, the typical problem of choosing taxes under an exogenous stream of public expenditures and an intertemporal government budget without restrictions on borrowing is identical to one where public expenditures are endogenous and the government’s budget must balance in every period. The analysis in the subsequent sections is conducted under this latter interpretation, since this makes it easier to extend the analysis to the case when the government cannot commit to future policy, as we discuss further in the end of this section and in section 4.

In the following, we present our model, beginning with the standard formulation, which we then generalize and specialize as discussed above.

2.1 Choosing how to finance an exogenous expenditure stream

Time is discrete and infinite, and there is an infinitely lived household endowed with standard time-additive preferences:

$$\sum_{t=0}^{\infty} \beta^t u(c_t, 1 - n_t).$$

Production is of the standard, neoclassical variety:

$$c_t + i_t = F(k_t, n_t), \tag{1}$$

where F has constant returns to scale. The stock of capital accumulates according to

$$k_{t+1} = i_t + (1 - \delta)k_t. \tag{2}$$

In a decentralized environment where the consumer accumulates capital and rents capital and labor services to firms and where income is taxed at proportional rates, the budget constraint reads

$$c_t + k_{t+1} + b_{t+1} = (1 - \delta)k_t + r_t k_t (1 - \tau_{kt}) + b_t R_t + w_t n_t (1 - \tau_{nt}),$$

with obvious notation; b_{t+1} represents lending to the government in period t and R_t denotes the gross interest rate, which has to satisfy

$$R_t = 1 - \delta + r_t (1 - \tau_{kt}) \tag{3}$$

¹¹See Lucas and Stokey (1983) and Krusell, Martin, and Ríos-Rull (2004).

in equilibrium since there are no arbitrage opportunities then.

Firms maximize profits taking input prices as given, yielding standard marginal-product-pricing formulae:

$$w_t = F_2(k_t, n_t) \text{ and } r_t = F_1(k_t, n_t). \quad (4)$$

The government has to finance a given sequence $\{g_t\}_{t=0}^{\infty}$ of government expenditures. Its period budget constraint reads

$$g_t + b_{t+1} = R_t b_t + \tau_{kt} r_t k_t + \tau_{nt} w_t n_t$$

or, with a no-Ponzi-game restriction imposed,

$$b_0 + \sum_{t=0}^{\infty} \frac{1}{R_0 R_1 \cdots R_t} (g_t - \tau_{kt} r_t k_t - \tau_{nt} w_t n_t) = 0. \quad (5)$$

The Ramsey problem, where a planner chooses tax sequences in order to maximize representative-agent utility subject to its budget constraint and subject to the restriction that the allocation be a competitive equilibrium allocation, can be stated as follows.

$$\max \sum_{t=0}^{\infty} \beta^t u(c_t, 1 - n_t) \text{ subject to (1), (2), (3), (4), (5),}$$

$$\begin{aligned} u_{l,t} &= u_{c,t} \cdot w_t (1 - \tau_{nt}), \\ u_{c,t} &= \beta u_{c,t+1} \cdot (r_t (1 - \tau_{kt}) + 1 - \delta), \end{aligned} \quad (6)$$

and an assumption that τ_{k0} is bounded above by some low enough number that the taxation problem is still nontrivial. This is an entirely standard problem whose solution has been discussed extensively in the literature; it implies, among other things, that taxes on capital income converge to zero over time. With more specific assumptions on utility—constant relative risk aversion—it also is easy to see that these taxes reach zero in finite time: monotone, quick convergence. We will make a partial statement of these results in the next subsection. We will now successively generalize and specialize this setup.

2.2 Generalizing the setup

The key new element we consider is variable depreciation of capital. For clarity, we will make a marginal generalization to the standard setup; later we will discuss some substantive applications. Suppose that a unit of investment at time t leads to one unit of productive capital 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 depreciation structure *quasi-geometric*.¹² Obviously, $\rho = 1$ refers to the standard, geometric depreciation case, whereas $\rho < (>)1$ captures lower (higher) initial depreciation than in the geometric case. Figure 1 represents a case of “accelerating” depreciation ($\rho \in (0, 1)$), showing the fraction of investments made in period $t-1$ that survives at $t, t+1, \dots$, etc..

Because of the non-geometric depreciation structure, it is now necessary to distinguish two kinds of capital at time t : “old” capital, for which we use the notation k_t^o , which was built at $t-2$ or earlier, and “new” capital, k_t^n , which was built at $t-1$, and which therefore equals i_{t-1} . The difference between these kinds of capital is not in their productivities—the total input of capital into production at t , which we still call k_t , equals $k_t^o + k_t^n$ —but in their depreciation rates from t to $t+1$. Thus, our assumptions are summarized by the following laws of motion for the two types of capital:

$$\begin{aligned} k_{t+1}^o &= (1 - \delta)k_t^o + (1 - \rho\delta)k_t^n, \\ k_{t+1}^n &= i_t. \end{aligned}$$

¹²One could alternatively assume that investments at t become productive already at t , a specification that was adopted in an earlier version of this paper. No result changes in this alternative setup. We choose the current specification for consistency with the standard capital taxation literature.

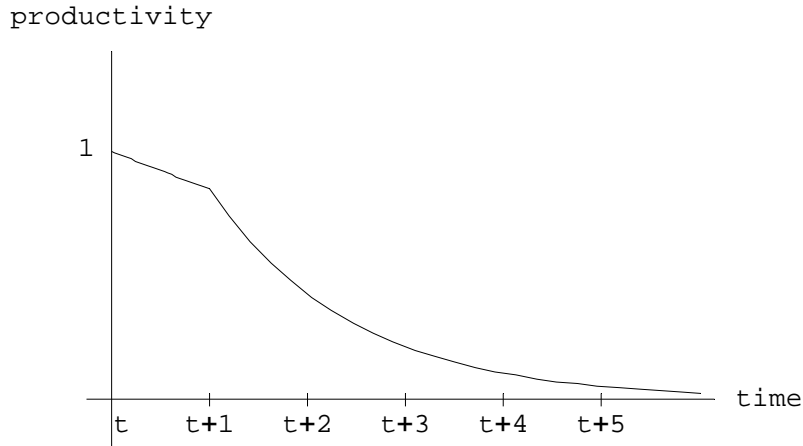


Figure 1: Remaining stock (“productivity”) of capital installed in period $t-1$ with quasi-geometric “accelerating” depreciation ($\rho \in (0, 1)$).

Put in terms of standard notation—capital in use, $k_t = k_t^o + k_t^n$ —these equations can be rewritten as a version of the standard capital accumulation equation:

$$k_{t+1} = i_t + (1 - \delta)k_t + \delta(1 - \rho)i_{t-1}. \quad (7)$$

In this formulation, total capital in use 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 $\rho\delta$.¹³ Notice, in particular, that when $\rho = 1$ equation (7) reduces to the standard case of geometric depreciation. Much of the analysis below will be conducted in terms of old capital, k^o , since it is a natural state variable, whereas k_t is not.

A one-hoss-shay depreciation structure—where investment at t stays intact until $t+2$ but then depreciates fully—is a particular case of the quasi-geometric setup: set $\delta = 1$ and $\rho = 0$ (in terms of figure 1, the segment between t and $t+1$ is perfectly flat, and, then, the curve falls vertically to zero).¹⁴ We will stress this special case, because it has an alternative, and important, interpretation: that of human capital. An investment in human capital has the basic feature that (at least a large part of) what is learned disappears when the person dies/leaves the labor force.¹⁵ So when $\delta = 1$ and $\rho = 0$ our model can be interpreted as one of overlapping generations of agents, each of whom invests in human capital and then works for two periods. Labor productivity is proportional to the amount invested and given this amount, labor supply is inelastic. One can also consider more general values for ρ : $\rho > 0$ corresponds to a downward-sloping age-earnings profile (the worker’s knowledge depreciates with age) and $\rho < 0$ corresponds to an upward-sloping one (workers learn without cost in the first period of working life). Our assumptions on utility are then 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. The version of our model with a human capital interpretation is useful in several ways. It depicts a form of capital for which perpetual, constant depreciation is an especially unnatural assumption. Moreover, it allows us to highlight the intuition behind

¹³With $k_t - i_{t-1}$ depreciating at rate δ and i_{t-1} at rate $\rho\delta$, the new total capital in use becomes $i_t + (k_t - i_{t-1})(1 - \delta) + i_{t-1}(1 - \rho\delta)$, which delivers the right-hand side of equation (7).

¹⁴A one-hoss-shay structure where capital lasts for $n > 2$ periods of life can be described as well, but not as a special case of the present setup. The quasi-geometric setup, however, can be generalized to allow for n distinct, initial depreciation rates followed by a final rate δ that applies for the rest of time.

¹⁵Of course, there can be transmission of human capital across workers or within families, in which case one can think of human capital as potentially living forever even though workers do not. Nevertheless, it seems plausible that a significant part of human capital is lost when a person departs.

the main result. Finally, it is a case of independent interest, particularly for political-economy applications where there is conflict of interest between generations: the assumption of perfect altruism is straightforward to relax (see Hassler et al., 2004).

Our second generalization is to consider a form of “adjustment cost”: we assume that consumption and investment are not perfect substitutes. In particular, we assume that the resource requirement in terms of consumption goods of producing i_t units of investment goods at time t is $G(i_t)$. So the resource constraint is now

$$c_t + G(i_t) = F(k_t, n_t), \quad (8)$$

where G is increasing and convex. The reason for this assumption is mainly technical, and we will return to it later. We assume, in our decentralization, that the adjustment cost is borne on the level of the consumer accumulating the capital. Moreover, we maintain the assumption that consumption and investment cannot be taxed at separate rates. This can be justified by informational constraints: only income is verifiable and can be taxed at the rate τ_t .

It is useful for what follows to state a preliminary set of results as a background. They pertain to the case with standard geometric depreciation but with no adjustment costs.

Proposition 1 *Assume that $\rho = 1$ and that $G(i)$ is linear. Then if the Ramsey problem above has a solution where the tax rates converge to a limit, it has to have $\tau_{kt} \rightarrow 0$. Moreover, in the special case where $u(c, n) = \frac{c^{1-\sigma}-1}{1-\sigma} + v(n)$, the Ramsey solution has $\tau_{kt} = 0$ for all $t \geq 2$.*

The proof can be found in Atkeson, Chari, and Kehoe (1999). The idea that taxes on capital converge to zero is well known and goes back to Chamley (1986) and Judd (1985). That taxes reach steady state in finite time is a less-discussed feature, and it sets a useful benchmark for the analysis in Section 3.1 below. In particular, we show there that the transitional dynamics are richer and qualitatively different when capital depreciates at a non-geometric rate ($\rho \neq 1$).

Next, we state a result concerning the effect of adjustment costs on long-run taxation.

Proposition 2 *Assume that $\rho = 1$ and that $G(i)$ is strictly convex. Then the Ramsey problem above with equation (1) replaced by equation (8) does not have a solution in which $\tau_{kt} \rightarrow 0$.*

Even though the long-run properties of capital taxation are not the main focus of our analysis, this result is useful. The Ramsey allocation in our model features positive taxation in the long-run, and Proposition 2 simply clarifies that this is due to the presence of adjustment costs. The proof of Proposition 2 is straightforward and is available upon request. Its closest relative in the literature is Correia (1996), who shows that the presence of non-taxed factor leads to nonzero limit taxation of capital income. 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 the framework we study here, the production of investment goods is not, unlike in the standard neoclassical model, just a linear function of output. We model the choice of investment as a household choice—say, through “home production”—and there are therefore untaxed “profits” in this operation. These profits are the equivalent of the untaxed factor income in Correia’s analysis.

2.3 Specializing the setup

We make two key simplifying assumptions. First, we assume that utility is linear in private consumption: $u(c, n)$ is quasi-linear. This is helpful because it pins down the interest rate. The interest rate would otherwise be subject to government “manipulation”, as in Lucas and Stokey’s 1983 paper: it is in the government’s interest to influence it so as to make the distortions associated with servicing government debt as small as possible. Moreover, a recent paper (Krusell, Martin, and Ríos-Rull, 2004) has shown that interest-rate manipulation in the case of no commitment, a case we have particular interest in studying in this paper (see Section 4), is substantially more complex than in the commitment case. Thus, linear utility of private consumption simplifies our analysis considerably and allows a sharper focus.

Second, we assume that the production function is linear in capital, with labor being completely unproductive: we assume “AK” production. In particular, we shall assume that production at t is simply k_t : it equals total (old plus new) capital. Thus, the issue here is now purely one of when capital income should

be taxed; there is no choice between taxing capital and taxing labor. The incorporation of labor into the analysis would be possible, but it would not deliver additional insights into the main issue studied here.

We can now clarify the role of the adjustment costs in our economy. Given the lack of curvature in utility and technology, the equilibrium allocation would have bang-bang properties. Convex adjustment costs prevent this unattractive feature. Incidentally, the strict convexity of G rules out long-run growth in spite of the AK technology (for this reason we can assume with no loss of generality that $A = 1$). We also make a specific assumption on G , which makes closed-form solutions possible, namely that $G(i) = i^2$. It is, of course, possible to relax this assumption and provide a characterization of the dynamics around a steady state which is qualitatively identical to the global solution we obtain. However, the solution of the case where there is no commitment is substantially less mathematically demanding in the case when G is quadratic than in the general case; see Klein, Krusell, and Ríos-Rull (2003).

2.4 The Ramsey problem

Using the simplifying notation $\tau_{kt} = \tau_t$, we arrive at the main object of analysis in this paper: the Ramsey problem. It corresponds to the following maximization problem:

$$\max_{\{\tau_t, i_t, k_{t+1}^o\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t (k_t^o + i_{t-1} - g_t - i_t^2)$$

subject to the government's budget,

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

the law of motion of (old) capital under quasi-geometric depreciation,

$$k_{t+1}^o = (1 - \delta) k_t^o + (1 - \delta\rho) i_{t-1}, \quad (9)$$

where k_0^o and i_{-1} are predetermined, and to

$$2i_t = \beta (1 - \tau_{t+1} + \beta\delta(1 - \rho)(1 - \tau_{t+2}) + 2(1 - \delta)i_{t+1}), \quad (10)$$

which is the first-order condition for the consumer's optimal investment choice, namely the analogue of the Euler equation (6) with adjustment costs and $G'(i) = 2i$. In addition, taxes are bounded between zero and one. In Section 3 we characterize the solution of this problem. Section 4 looks at the situation where there is no commitment. In this case, the outcomes cannot be characterized using control theory alone: the problem must be viewed as a game between successive dynastic planners each solving a control-theory problem.

2.5 The public-goods interpretation of the Ramsey problem

We will now present an alternative interpretation of the Ramsey problem—the public-goods interpretation—which is useful in ways that will be explained momentarily. To see how the alternative interpretation is arrived at, note that the Ramsey problem can be analyzed with standard Lagrange multiplier methods. Thus, define Ψ as the Lagrangian multiplier associated with the government budget constraint. Given this multiplier, the Lagrange method then implies that the Ramsey problem can be expressed as follows:

$$\max_{\{\tau_t, i_t, k_{t+1}^o\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t ((\tau_t (\Psi - 1) + 1)(k_t^o + i_{t-1}) - i_t^2), \quad (11)$$

subject, as before, to (9) and (10). This formulation is arrived at by stating the problem in the usual way and rearranging terms; there is a remainder, $-\Psi(b_0 + \sum_{t=0}^{\infty} \beta^t g_t)$, which contains no choice variables, so it can be ignored when choosing taxes and investments. The solution to the problem in (11) depends on Ψ ; the value of Ψ , in turn, is determined by minimizing the objective in (11), less $-\Psi(b_0 + \sum_{t=0}^{\infty} \beta^t g_t)$: it represents the shadow value of the government's budget constraint. The shadow value of this constraint depends on the

present value of government expenditures, initial debt, and installed capital at time zero, which determine the government's needs to raise funds and the elasticities of the tax bases in this problem.

The public-goods interpretation follows from observing (11). Suppose, in particular, that the government's expenditure sequence is not exogenous but subject to choice. Suppose, in addition, that private consumers derive utility from the public good: the period utility flow is $u(c, g) = c + Ag$, with $A > 1$.¹⁶ Absent uncertainty, access to capital markets is, under this interpretation (in particular, since the marginal utility of the public good is constant), a redundant policy instrument for the planner: the same utility and allocation is attained irrespective of whether the government can or cannot save and issue debt. It therefore entails no loss of generality to assume that public good provision is subject to a balanced-budget condition: $g_t = \tau_t(k_t^o + i_{t-1})$. All other assumptions are unchanged. After substituting the government budget constraint in the objective function, the Ramsey problem can now be stated as follows:

$$\max_{\{\tau_t, i_t, k_{t+1}^o\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t ((\tau_t(A-1) + 1)(k_t^o + i_{t-1}) - i_t^2), \quad (12)$$

subject, as before, to (9) and (10). Clearly, the tax and investment sequences generated by this problem is the same as in the case of exogenous government expenditure, (11), as long as $\Psi = A$. Since, as explained above, Ψ is determined by exogenous factors, the two models are isomorphic. One can interpret A as either the marginal utility agents derive from the public good or as the shadow value of government funds when an intertemporal budget constraint has to be satisfied.

In the rest of the analysis, we stress the public-goods interpretation; this means that we have to ask the reader to keep in mind that our main result—that the Ramsey tax sequence features oscillations when $\rho \in (0, 1)$ —also applies to the standard optimal-taxation problem with exogenous government expenditures. We view the public-goods formulation as being of independent interest—the choice of when to consume public goods is one which has been studied in the literature—and close relatives of this formulation are useful models of political economy.¹⁷ In the political-economy settings, which involve more elaborate population structures (e.g., overlapping generations, lucky and unlucky workers, etc.), the expenditure can take the form of transfers between groups or public goods.¹⁸ In addition, the balanced-budget assumption is useful for our study of how the results in the main proposition are changed if the government does not have access to commitment; in political-economy settings, especially, the role of debt raises separate issues that we wish to abstract from here.

From the public-goods perspective, two issues are worth noting. First, we assume linearity also in the utility of public goods. The key, however, is that the marginal utility of public goods is higher than that of private goods. With access to lump-sum taxes, thus, good policy would be represented by consuming only public goods, whereas in a more general case there would be a non-degenerate mix of public and private goods in the unrestricted optimum (characterized by $u_c = u_g$). With distortions, there will be a nontrivial trade-off between the marginal benefits of public goods and the marginal costs of raising the funds to finance these goods. Linearity simplifies the analysis and is important for establishing equivalence between this framework and that in the previous section. Second, the flow utility from consuming public goods is stationary: A does not depend on time. This will mean that the choice between g_t and g_{t+1} is not an interesting one: there will be indifference, just like there is for private consumption. It will also mean that the second new assumption that we introduce here, namely the balanced-budget assumption for the government, is not binding. Thus, the focus here is on the timing of when to finance the public good as opposed to on the timing of when to consume it.

¹⁶For our two-period-life overlapping-generations interpretation, the public good can be thought of as consumed simultaneously by all living agents, providing each living individual with a marginal utility equal to $A/3$, thus delivering a total marginal utility to the dynasty of A .

¹⁷See, e.g., Klein, Krusell, and Ríos-Rull (2003), which studies the optimal provision of public expenditures in the long run, with special emphasis on the role of commitment.

¹⁸For examples based on models related to the present one, see Hassler, Rodríguez-Mora, Storesletten, and Zilibotti (2003), Hassler, Storesletten, and Zilibotti (2003a and 2003b), and Hassler, Krusell, Storesletten, and Zilibotti (2004).

3 Analysis

The Ramsey problem described above admits an analytical solution. To see how, it is convenient to express the constraints of the Ramsey problem differently. First, note that

$$k_{t+1}^o = k_1^o (1 - \delta)^t + (1 - \rho\delta) \sum_{s=0}^{t-1} (1 - \delta)^{t-1-s} i_s. \quad (13)$$

Second, solving forward the Euler condition for investment, (10), assuming that the investment sequence does not blow up (i.e., $\lim_{t \rightarrow \infty} \beta^t i_{t+1} = 0$), one arrives at

$$i_t = i(T_t) = \frac{1}{2} (\kappa - T_t) \quad (14)$$

for $t \geq 0$, where

$$T_t \equiv \beta \tau_{t+1} + (1 - \rho\delta) \sum_{s=2}^{\infty} \beta^s (1 - \delta)^{s-2} \tau_{t+s}, \quad (15)$$

denotes the effective discounted sum of taxes (that we label the “total tax”) in period t , and

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

is the effective duration of new investment. Thus, the distortions on investment in period t can be captured entirely by the total tax T_t . This is a very useful simplification of our setup: present and future tax rates entering T_t are perfect substitutes for investment decisions. Another useful feature implicit in the investment demand equation (14) is that the investment in vintage t does not interact with investment in other vintages. This feature is due to utility being linear in consumption.

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

Lemma 3 *The Ramsey problem (12) subject to (9) and (10) is equivalent to the following program:*

$$\max_{\{\tau_t\}_{t=0}^{\infty}} (A - 1) \left(\tau_0 (k_0^o + i_{-1}) + \hat{T}_0 k_1^o \right) + \sum_{t=0}^{\infty} \beta^t y(T_t), \quad (16)$$

where

$$\hat{T}_0 \equiv \beta \sum_{t=0}^{\infty} \beta^t (1 - \delta)^t \tau_{t+1}, \quad (17)$$

$$y(T_t) \equiv A\kappa i(T_t) - i(T_t)^2 (2A - 1), \quad (18)$$

and T_t and $i(T_t)$ are defined as in (15) and (14).

The new functions $y(T_t)$ and \hat{T}_0 will be particularly useful in the analysis below. The function $y(T_t)$ is the present value of the contribution to the planner’s utility of the investment vintage installed at t . Each vintage contributes to the planner’s utility via private consumption, $i_t (\kappa - T_t)$, the provision of financing to the public good, $AT_t i_t$, and the investment cost, $-i_t^2$. Furthermore, \hat{T}_0 is the effective discounted sum of taxes levied on capital installed before the beginning of the planning horizon, and thus inelastic. With analogy to previous definitions, we label it the “total tax on inelastic capital”. Taxes entering \hat{T}_0 are discounted at the rate $\beta(1 - \delta)$, reflecting the interest rate and the rate of depreciation of the pre-installed capital stock.

The objective function (16) is then the sum of the present discounted value (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-installed capital (the expression ignores constant terms representing the PDV of output

from pre-installed capital). Inspecting the latter term, one sees that the tax τ_0 is entirely lump sum: it enters additively with respect to other tax rates and it multiplies a predetermined term, $k_0^o + i_{-1}$. Therefore, if τ_0 were a choice variable, it would always be set at its maximum feasible level, with no effect on any other choice. Hence, without loss of generality, we can actually assume that the tax rate at time zero is exogenous. In particular, following Klein and Ríos-Rull (2003) we set $\tau_0 = 0$ as if there were a one-period implementation lag.¹⁹ Ignoring irrelevant constants and predetermined variables, the Ramsey problem simplifies to:

$$\max_{\{\hat{T}_0, T_t\}_{t=0}^{\infty}} (A-1)\hat{T}_0 k_1^o + \sum_{t=0}^{\infty} \beta^t y(T_t). \quad (19)$$

Here, we have defined the choice variables of the problem to be the T_t 's and \hat{T}_0 rather than the tax sequence $\{\tau_t\}_{t=1}^{\infty}$. This can be regarded as a *primal* formulation, where the planner chooses an allocation directly, subject to the constraint that it is a competitive equilibrium (recall that $i_t = (\kappa - T_t)/2$ and, hence, choosing the T_t is equivalent to choosing investments). In addition, it has to be verified that there is a one-to-one mapping between sequences of individual tax rates and sequences of present-value taxes all satisfying (15) and (17). Because we require tax rates to be bounded between zero and one, the present-value taxes are bounded as well, since $\beta(1-\delta) < 1$. Given a choice of present-value taxes, $\{T_t\}_{t=0}^{\infty}$, one can then back out a unique sequence of tax rates $\{\tau_t\}_{t=1}^{\infty}$ which satisfies the boundedness condition.²⁰ Recall, finally, that k_1^o is a key predetermined variable; its size will influence the dynamics of taxes.

3.1 The case of geometric depreciation

We first look at the case of geometric depreciation, i.e., $\rho = 1$. The key feature to note is that in this case all the present-value tax expressions are geometric (as opposed to quasi-geometric) sums of future tax rates. In particular, $\hat{T}_0 = T_0$: the total tax on inelastic capital is identical to the total tax on investment in period zero, which is distortionary. Thus, rewriting equation (19), the Ramsey planner faces the following problem:

$$\max_{\{T_t\}_{t=1}^{\infty}} (A-1)T_0 k_1^o + \sum_{t=0}^{\infty} \beta^t y(T_t).$$

The solution to this Ramsey problem is simple and striking: the problem is separable in the T_t 's so these variables can be chosen independently—one by one. Moreover, the choice problem for T_t looks identical for all t except for $t = 0$. This observation, together with the fact that each such problem is, by definition, strictly concave (since y is strictly concave), immediately implies that the total taxes will be constant—say, $T_t = T^*$ —from $t = 1$ and onwards. That is, we reach a steady state after one period. It is given by

$$y'(T^*) = 0$$

Backing out tax rates, this implies that tax rates are also constant after one period: $\tau_t = \tau^*$ from $t = 2$ and onwards.

In period zero, the choice problem is different: here the choice of T_0 , which distorts i_0 , also raises revenue and generates surplus from the taxation of the inelastic capital, k_1^o . Thus, the optimal T_0 , which we label T_0^* , will satisfy

$$(A-1)k_1^o + y'(T_0^*) = 0,$$

and thus it will exceed T^* . In turn, this implies that $\tau_1 > \tau^*$. The extent of the initial tax hike depends on k_1^o : τ_1 turns out to be linearly increasing in this variable. However, k_1^o has no effect on any variable after period one.

¹⁹ While the assumption of a one-period implementation lag is immaterial in the case of commitment, it becomes important in the case when the government has no commitment technology. In such case, it rules out a trivial solution where the government in every period would set the current tax to its maximum.

²⁰ Forward iterating on equation (15) 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. See the details of the proof of Proposition 4 in the appendix for a formal argument.

3.2 Quasi-geometric depreciation

In the general case with quasi-geometric depreciation, \hat{T}_0 is no longer equal to T_0 : as indicated above, and as we will explain in more detail below, because the inelastic capital, k_1^o , depreciates at a different rate from new investments, the timing of taxes can be used to exploit these differences. Formally, one can use equations (15) and (17) to obtain, after repeated substitutions (proof in the appendix):

$$\hat{T}_0 = \sum_{t=0}^{\infty} (-\delta\beta(1-\rho))^t T_t. \quad (20)$$

The connection between \hat{T}_0 and the sequence of T_t s is key for understanding the oscillatory tax dynamics. Note that, if $\rho < 1$ (i.e., if capital depreciates less in the earlier part of its productive life), the weights on the future total taxes T_t are oscillatory. Thus, every T_t will influence the taxation of inelastic capital, and whether it increases or decreases the total tax on inelastic capital depends on whether t is odd or even.

Rewriting (19), the Ramsey problem now reads

$$\max_{\{T_t\}_{t=1}^{\infty}} (A-1) \left(\sum_{t=0}^{\infty} (-\delta\beta(1-\rho))^t T_t \right) k_1^o + \sum_{t=0}^{\infty} \beta^t y(T_t),$$

and the first-order condition for the T_t 's is

$$(A-1)k_1^o(-\delta(1-\rho))^t + y'(T_t) = 0, \quad (21)$$

which uniquely pins down the T_t 's, in turn pinning down the unique feasible sequence of tax rates (see, again, the details of the proof of Proposition 4). The first-order condition for T_0 is the same as in the geometric case. However, now the inelastic capital, k_1^o , affects the entire sequence of investments and taxes.

The solution can be summarized as follows:

Proposition 4 *Assume that $\|\delta(1-\rho)\| \leq 1$ and that $(A-1)k_1^o$ is not too large. Then, the unique tax sequence that implements the Ramsey allocation is given by*

$$\tau_{t+1} = \tau^* - \delta(1-\rho)(\tau_t - \tau^*) \text{ for } t \geq 1, \quad (22)$$

$$\tau_1 = \tau^* \left(1 + 2k_1^o \frac{1 + \beta\delta(1-\delta)(1-\rho)}{\beta(1 - \beta\delta^2(1-\rho)^2)} \right), \quad (23)$$

where $\tau^* \equiv (A-1)/(2A-1) < \frac{1}{2}$. 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-cycle. If $\delta(1-\rho) < 0$, the tax sequence converges monotonically to τ^* .

Proof. The first-order condition (21), together with the definition of $y(T_t)$ as given in (18), yield the following sequence:

$$T_t = \left(\frac{A-1}{2A-1} \right) \left(\kappa + 2k_1^o(-\delta(1-\rho))^t \right). \quad (24)$$

The proof amounts to showing that the tax sequence (22)-(23) is the unique sequence satisfying (24) and the boundary condition $\tau_t \in [0, 1]$, given the definition of the T_t 's as in (15). The details are in the appendix. ■

Figure 2 shows the dynamics of total taxes (T_t), investments and private output, defined as $k_t^o + i_{t-1} - i_t^2$, in a case of accelerated 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. Private output fluctuates around a geometric trend toward the steady state.²¹

²¹However, private output, excluding adjustment costs, i.e., $k_t^o + i_{t-1}$, displays monotone convergence and is, in fact, constant in the human capital case ($\delta = 1$). The proof is available upon request.

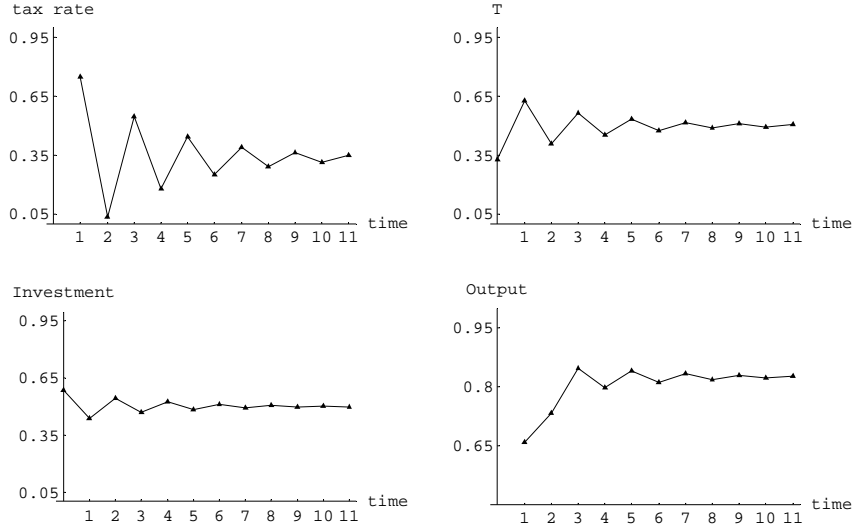


Figure 2: Ramsey dynamics—an example with “accelerated” quasi-geometric depreciation ($\delta = 0.9$ and $\rho = 0.1$).

3.3 Interpretation

3.3.1 A second-best benchmark: vintage-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 vintage-specific taxation. Suppose that the planner can tax the income produced by different investment cohorts at different rates. The Ramsey sequence, then, is very simple: the planner taxes the returns to the inelastic capital, k_1^o , at the highest possible rate every period, since these taxes are non-distortionary. All vintages after period zero are 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. The fact that taxes and investments are constant is a manifestation of the planner’s desire to smooth distortions. When vintage-specific tax instruments are available, there is no reason to deviate from perfectly smooth (i.e., constant) taxes and investments.

In contrast, in the economy analyzed in the previous sections, vintage-specific taxes are ruled out, and the only way the planner can extract tax revenue from inelastic capital is by distorting new investments away from the second-best.²² In this economy, as we will see, there is a trade-off between the objective of smoothing distortions and that of taxing inelastic capital. Note, in particular, that the Ramsey tax sequence of Proposition 4 features perfect tax and investment smoothing when $k_1^o = 0$: when there is no inelastic capital from which the planner can extract revenue, she chooses taxes that are constant and 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. It is important to note, however, that investments 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 vintage of investments ($T_0 > T_t = T^*$, for all $t > 0$).²³ This 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 ?

²²Hassler et al. (2004) analyzes the properties of the Ramsey allocation in a two-period version of this model when age-dependent taxation is allowed.

²³The absence of vintage-specific instruments reduces public good provision. In the second-best allocation, the revenue from inelastic capital is completely confiscated, whereas here it is taxed at rate τ^* after one period.

First, we note that given the tax burden on the inelastic tax base, \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 tried to reduce τ_1 and increase τ_2 so as to keep \hat{T}_0 constant, investment in period zero would also remain unaffected. In addition, 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 $t > 1$ are independent of the amount of initial inelastic capital. To understand this result, we begin by noting that along the optimal path of taxes, the marginal distortionary cost associated with each tax τ_s must be constant relative to the marginal revenue generated by that tax.²⁴ If k_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 k_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 k_1^o should not lead to any changes in τ_2 or, more generally, in any subsequent tax rates.²⁵

In the more traditional case when the government has to finance an exogenous stream of expenditures, the argument is identical, with one qualification. While the marginal excess value of public funds is exogenous at $A - 1$ under the public good interpretation, an increase in k_1^o reduces the Lagrange multiplier on the budget constraint Ψ in the expenditure financing case. This implies that while τ_1 increases, all other tax rates must fall: for all $s > 1$, $\tau_s = \tau^* = \frac{\Psi}{1+2\Psi}$.

3.3.3 Quasi-geometric depreciation

We now move to the general case, where $\rho \neq 1$. According to Proposition 4, 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: the human capital model discussed in Section 2.2.

A particular case: human capital ($\delta = 1$). The human capital case has a feature that makes the analysis particularly intuitive: τ_1 is the only instrument the planner has available to tax the inelastic capital. Taxes at later dates do not extract revenue from k_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 (as in the case of geometric depreciation, the magnitude of such hike is increasing in the inelastic capital), it also distorts investments in period zero, i_0 . 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 taxes in period two stimulate 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, k_1^o . To understand this we note that as in the geometric case, τ_1 is optimally increased when k_1^o is higher causing the the marginal distortionary cost of tax τ_2 to increase. However, this marginal cost increase is *not* balanced by higher marginal revenues of τ_2 since the higher stock of k_1^o is fully depreciated in period 2. Thus, τ_2 should be reduced.

The parameter ρ is key for the strength of the oscillations. Consider for instance the extreme case where $\rho = 0$: the *one-hoss shay* case. As Proposition 4 shows, in this case oscillations are particularly potent and

²⁴We will return to the issue of distortion smoothing below.

²⁵The formal analysis behind these arguments is available in the Appendix.

do not die out: the economy ends up in a two-period cycle. The reason is that in this case, the increase in the distortionary cost of τ_2 due to the reduction in period zero investments is particularly large. Equivalently, the effectiveness of counteracting a current tax hike by a next-period tax break is high. When $\rho > 0$, more of the capital's income is accrued in the first period of life than in the second and the effectiveness of using a reduction in τ_2 to counteract distortions in period zero is lower. Oscillations are therefore weaker 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, and in contrast to the human capital case, capital lives forever, and the total 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 key difference now is that $\hat{T}_0 \neq T_0$: 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 .²⁶ Unlike in the geometric depreciation case, the planner can now use the timing of tax rates as an imperfect substitute for the missing vintage-specific tax instrument to achieve better distortion smoothing. Recall, in particular, that the tax hike in the first period causes investment to be particularly low. Thus, distortion smoothing makes it desirable for the planner to reduce T_0 and increase initial investments. In close parallel to the human capital case, the planner achieves that goal by setting $\tau_2 < \tau^*$. However, having done this, it is not optimal to set $\tau_t = \tau^*$ for $t > 2$. 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 the second-best level. Again, distortion smoothing suggests an increase in τ_3 so as to reduce i_1 . In turn, this creates a motive to smooth the distortion on capital invested in period two, which is best done by reducing τ_4 below τ^* , and so on.

Again in contrast to the geometric case, taxes at period $t > 1$ are affected by the amount of initial inelastic capital k_1^o . Whenever $\delta < 1$, some of k_1^o remains in period 2, implying that an increase in k_1^o increases the marginal revenue of τ_2 . The marginal distortionary cost of τ_2 also increases. However, the distortion induced by τ_2 increases *more* than the marginal revenue since when $\rho < 1$, period zero investments depreciate less than k_1^o between periods 1 and 2. This implies that τ_2 should be reduced relative to τ^* , which reduces the distortion from τ_3 relative to the marginal revenue it generates, implying that τ_3 should be increased, and so on.

The case with decelerating depreciation. When $\rho > 1$ the tax sequence converges monotonically to τ^* . The intuition for the different dynamics is the mirror image of the previous argument. If capital depreciates faster during its first period of life, the planner can still exploit the different depreciation rates to reduce T_0 while keeping \hat{T}_0 constant. Here, this is achieved by *reducing* τ_1 and *increasing* τ_2 . An increase in k_1^o leads to an increase in τ_1 as in all previous cases. However, the increased marginal distortionary cost from τ_2 now increases *less* than the increase in marginal revenues from τ_2 since the inelastic tax base depreciates more slowly than do new investments. As a consequence, τ_3 should also be increased relative to τ^* , and so on. As long as $\delta(1 - \rho) > 1$, the dynamics converge monotonically to a steady state.

4 Lack of commitment

So far, we have allowed the planner to determine taxes for all future dates under full commitment. The purpose of this section is to characterize the optimal time-consistent allocation, namely, the allocation that is chosen by a benevolent planner without access to a commitment technology. Complete lack of commitment would lead to a trivial outcome: if the planner sets the tax rate τ_t at time t , the optimal tax rate would be 100% in every period. Instead, as discussed in section 3, we assume that the planner has access to commitment for one period: there is a short implementation lag. Investments at t are made after period

²⁶The particular case of human capital provides an extreme example: by keeping τ_1 constant and reducing τ_2 , one can decrease T_0 while keeping \hat{T}_0 constant.

$t + 1$ taxes are set, and this means that there is no “free lunch”: any taxation is distortionary at the time the tax rate is chosen.

4.1 Formalizing the lack of commitment

When there is no commitment, one needs to study a dynamic game between successive governments, with the private sector “moving second” in each period. More precisely, in the beginning of period t , the government sets the tax rate to be applied in period $t + 1$. Based on this knowledge, the private sector decides on current investment, i_t . Both the government and the private sector make decisions given expectations of how future tax rates, $\{\tau_{t+2}, \tau_{t+3}, \dots\}$, will be set, period by period, by future governments. We will look at limits of finite-horizon equilibria. These equilibria are the most natural contrasts to equilibria under commitment: they do not allow reputation mechanisms to partly or fully replicate the commitment allocation.²⁷ Moreover, finite-horizon equilibria exist and are unique in our model. This means that the equilibria we focus on are the only equilibria which are robust in the sense that they exist under very long time horizons, whether finite or infinite. In practice, because finite-horizon equilibria are first-order Markov with linear decision rules, we operationalize the equilibrium concept for the infinite-horizon economy by restricting attention to linear, first-order Markov-perfect equilibria.²⁸ In conclusion, the Markov equilibrium that we focus on in the following will (linearly) map the state variable as of the beginning of a period into a policy choice and an implied investment choice.

What is the relevant state variable here? Three variables of significance are predetermined at time t : the tax rate to be implemented this period, τ_t , the amount of old capital, k_t^o , and the amount of new capital (i.e., investment last period), i_{t-1} . These variables all matter for utility. However, only a one-dimensional summary of these variables will matter for the choices of the current government, namely, k_{t+1}^o , which equals $(1 - \delta)k_t^o + (1 - \rho\delta)i_{t-1}$: this is next period’s tax base, to which τ_{t+1} applies directly. The current tax rate will not matter because it is a lump-sum tax that does not interact with future taxes; this follows because from the additive separability due to our preference and technology specifications. Similarly, new and old current capital also only matter separately for current utility flows but since they are additively separable in future tax rates, they do not matter for the choice of τ_{t+1} . Thus, the key equilibrium objects sought here are (i) a tax function, \mathcal{T} , delivering the current government choice, $\tau_{t+1} = \mathcal{T}(k_{t+1}^o)$, and (ii) an investment function, \mathcal{I} , delivering current investment, $i_t = \mathcal{I}(k_{t+1}^o)$. These functions are time-invariant, because we study a model with an infinite horizon.

We will now state the government’s problem as a dynamic programming problem. We will study the problem at time $t - 1$, when the government jointly selects τ_t and i_{t-1} subject to an implementability constraint, namely, that the first-order condition for investment holds. First, define the period $t - 1$ felicity of the planner by

$$F(i_{t-1}, k_t^o, \tau_t) \equiv \beta(1 + (A - 1)\tau_t)(k_t^o + i_{t-1}) - i_{t-1}^2.$$

When the planner announces the tax rate τ_t , k_t^o is predetermined. The tax τ_t instead affects investment at $t - 1$ (since $i_{t-1} = i(T_{t-1})$).

Expectations about future taxes and investment decisions are rational: they are set according to the time-invariant functions $\tau_{t+s-1} = \mathcal{T}(k_{t+s-1}^o)$ and $i_{t+s-2} = \mathcal{I}(k_{t+s-1}^o)$, respectively, where the future capital levels are given recursively from

$$k_{t+s}^o = \mathcal{H}(k_{t+s-1}^o) \equiv (1 - \delta)k_{t+s-1}^o + (1 - \rho\delta)\mathcal{I}(k_{t+s-1}^o), \quad s > 1$$

and $k_{t+1}^o = (1 - \delta)k_t^o + (1 - \rho\delta)i_{t-1}$. The Ramsey-Markov problem without commitment thus leads to the following equilibrium definition:

Definition 5 *A time-consistent (Markov) allocation without commitment is defined as a set of functions $\langle W, \mathcal{T}, \mathcal{I} \rangle$, where W is a bounded planner value function, $\mathcal{T} : [0, \infty) \rightarrow [0, 1]$ is a public policy rule $\tau_t = \mathcal{T}(k_t^o)$, and $\mathcal{I} : [0, \infty) \rightarrow [0, \infty)$ is a private investment rule $i_{t-1} = \mathcal{I}(k_t^o)$ such that:*

²⁷It is well known that in economies with a literally infinite time horizon and no commitment, under certain conditions on primitives, it may be possible to use reputation mechanisms to support the commitment allocation as an equilibrium.

²⁸We believe that there are other first-order Markov-perfect equilibria as well in this model, as for example found in Krusell and Smith (2003), but these other equilibria are not limits of finite-horizon equilibria.

1. W solves the following recursive problem

$$W(k_t^o) = \max_{\tau_t, i_{t-1}} \{F(i_{t-1}, k_t^o, \tau_t) + \beta W(k_{t+1}^o)\} \quad (25)$$

subject to

$$k_{t+1}^o = (1 - \delta) k_t^o + (1 - \rho\delta) i_{t-1} \quad (26)$$

$$i_{t-1} - \beta(1 - \delta)\mathcal{I}(k_{t+1}^o) = \frac{\kappa}{2}(1 - \beta(1 - \delta)) - \frac{\beta}{2}\tau_t - \frac{\beta^2}{2}\delta(1 - \rho)\mathcal{T}(k_{t+1}^o) \quad (27)$$

and

2. the tax and investment rules are optimal, i.e.,

$$\{\mathcal{T}(k_t^o), \mathcal{I}(k_t^o)\} = \arg \max_{\tau_t, i_{t-1}} \{F(i_{t-1}, k_t^o, \tau_t) + \beta W(k_{t+1}^o)\} \quad (28)$$

subject to (26)-(27)

Notice that this is a fixed-point problem in the decision rules \mathcal{T} and \mathcal{I} : based on these rules determining the expectations of how τ_{t+1} and i_t will be set, respectively, these same rules have to be the maximizing choices of τ_t and i_{t-1} for all values of the argument of these rules, k_t^o .

4.2 Finding the Markov equilibrium

Finding the linear Markov equilibrium is not difficult, given the linear-quadratic nature of the objective and constraints. More precisely, one proceeds by conjecturing that the functions \mathcal{T} and \mathcal{I} are both linear:

$$\mathcal{T}(k_t^o) = \alpha_{01} + \alpha_{11}k_t^o \quad (29)$$

and

$$\mathcal{I}(k_t^o) = \alpha_{02} + \alpha_{12}k_t^o. \quad (30)$$

Then, using these guesses one can solve the dynamic programming problem of the government, which is now a linear-quadratic problem whose solution will depend on the four parameters $\{\alpha_{01}, \alpha_{11}, \alpha_{02}, \alpha_{12}\} \equiv \alpha$. The solutions for the implied decision rules are linear; we can express them as a vector of two intercept-slope pairs, $\{\hat{\alpha}_{01}(\alpha), \hat{\alpha}_{11}(\alpha), \hat{\alpha}_{02}(\alpha), \hat{\alpha}_{12}(\alpha)\} \equiv \hat{\alpha}(\alpha)$, with obvious notation: these decision-rule coefficients depend on α . The function $\hat{\alpha}(\alpha)$ is nonlinear. The fixed-point problem is thus the nonlinear four-equation system $\alpha = \hat{\alpha}(\alpha)$ in the four unknown parameters.

In order for a Markov equilibrium to exist, the dynamics of τ_t must be non-explosive (else, the boundaries on taxes are reached in finite time, and the solution cannot be linear). Observe that

$$\begin{aligned} k_{t+1}^o &= (1 - \delta)k_t^o + (1 - \rho\delta)i_{t-1} \\ &= (1 - \delta + (1 - \rho\delta)\alpha_{12})k_t^o + (1 - \rho\delta)\alpha_{02} \end{aligned}$$

so that capital converges to a steady-state as long as $|1 - \delta + (1 - \rho\delta)\alpha_{12}| < 1$. Due to the linear dynamics, the same condition guarantees the convergence of the Markov tax sequence: $\tau_t = \alpha_{01} + \alpha_{11}k_t^o$. In particular, this implies that

$$\begin{aligned} k_t^o - k^{**} &= (1 - \delta + (1 - \rho\delta)\alpha_{01})(k_{t-1}^o - k^{**}) \\ \tau_t - \tau^{**} &= (1 - \delta + (1 - \rho\delta)\alpha_{01})(\tau_{t-1} - \tau^{**}) \end{aligned}$$

where k^{**} and τ^{**} are the steady-state levels of k^o and τ , respectively. In general, the steady state without commitment is different from the steady state with commitment.

We proceed as above in order to characterize the solution without commitment. We start with the special case of human capital, which is especially convenient since it allows a closed-form characterization and, hence, an analytical comparison with the problem with commitment. Thereafter, we study the general framework, for which we have not found a closed-form solution for α . We calibrate the general setup and thereby obtain insights into both the qualitative and quantitative comparisons between the optimal solutions with and without commitment.

4.2.1 The case of human capital

We proceed directly to the solution.

Proposition 6 *Assume that $\delta = 1$, $\rho \in [0, 1)$ and that $(A - 1)k_1^o$ is not too large. Then, there exists a Markov equilibrium such that*

$$\begin{aligned}\tau_{t+1} &= \tau^{**} - \lambda(\tau_t - \tau^{**}) \text{ for } t > 1, \\ \tau_1 &= \alpha_{01} + \alpha_{11}k_1^o\end{aligned}$$

where $\tau^{**} > \tau^*$, $\lambda \in (0, 1 - \rho)$, $\alpha_{01} > 0$, and $\alpha_{11} > 0$.

The proof as well as the exact expressions of τ^{**} , λ , α_{01} and α_{11} are provided in the appendix. The main findings embodied in this result are:

1. The allocation without commitment implies higher steady-state taxation ($\tau^{**} > \tau^*$), along with lower long-run output and investment levels, than does the allocation under commitment.
2. The steady-state Markov tax rate, τ^{**} , can exceed $1/2$, i.e., it can be larger than the constant value of taxes that maximizes tax revenues and public good provision.²⁹
3. The allocation without commitment implies oscillations, but these oscillations are more dampened (i.e., the tax sequence is smoother) than in the allocation with commitment: $0 < \lambda < (1 - \rho)$.³⁰

The first finding, thus, is that the lack of commitment induces the planner to systematically “over-tax” capital *ex post*. Agents anticipate that and respond by decreasing their investment. As a result, output is lower. The second finding means that it is possible in this model that a benevolent planner chooses a long-run tax rate that is on the “wrong” side of the Laffer curve. This is an extreme manifestation of the lack of commitment; if $\tau^{**} > 1/2$, the planner would clearly like to reduce the steady-state tax rate. However, the planner can only control next period’s tax rate and a one-period reduction of τ_{t+1} would lead to even higher taxes in the following period, resulting in an overall reduction of the current welfare.

The third finding concerns the dynamics, which is our main focus here. It states that, again due to a lack of commitment, the Markov planner chooses, along a transition, an inefficiently smooth tax sequence. The intuition runs as follows. As discussed above, when the Ramsey planner can commit, she compensates high taxation in period one by promising low taxation in period two in order to smooth distortions. The Markov planner cannot credibly promise as low future taxes as those committed upon by the Ramsey planner. At time zero, agents thus expect that period two taxes will be relatively high, so that for any τ_1 , the distortions are large. The optimal behavior of a Markov planner in period zero is therefore to set τ_1 *lower* than the Ramsey planner would, because such a choice counteracts to some extent the unavoidable fact that the next government will choose too high a tax: it keeps initial investment from being too low. The same logic applies to later periods. So the tax sequence tends to be smoother. For completeness, in Section 5 we will supplement this analysis with a formal treatment of the idea of “distortion smoothing”, which allows an alternative way of comparing the cases with and without commitment.

The upper panel of Figure 3 clearly illustrates these features for a simulated economy. The Markov tax sequence (the dashed line) starts below the Ramsey tax sequence (the continuous line), is smoother, and converges to a steady-state with higher taxation.

4.2.2 The general case: calibration

Here, we simply solve the four-equations-four-unknowns nonlinear equation system that characterizes the Markov solution numerically for parameter values that are an attempt to provide a rough calibration of our economy. The period length is set to four years, which is the length of a U.S. electoral cycle. We assume that the U.S. data is generated by a Ramsey planner without commitment who plays the Markov game

²⁹Specifically, this occurs whenever $A > (2 - \rho)(3 - \rho)(1 - \rho)^{-1}(3 - \rho + \beta(1 - \rho))^{-1}$.

³⁰Interestingly, λ is increasing in A and decreasing in ρ . Namely, oscillations are more pronounced when A is larger and less pronounced when investment have less persistent effects. Details of these comparative statics are available upon request.

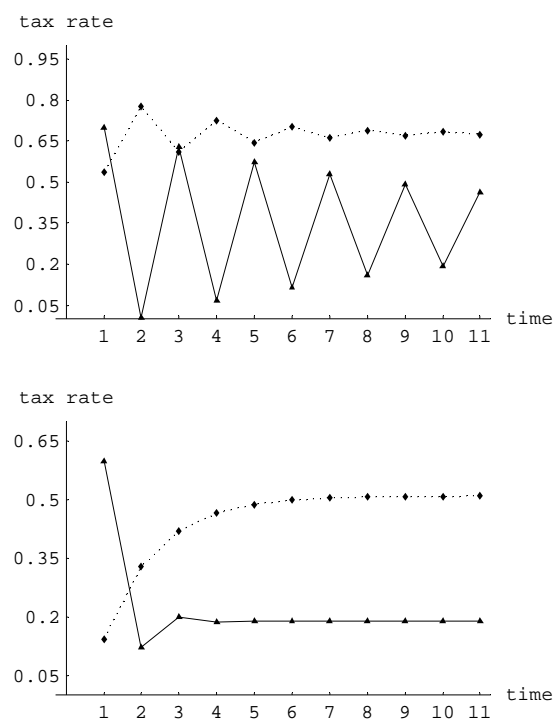


Figure 3: Comparison between Ramsey (the continuous line) and Markov (the dashed line) tax sequences in two simulated economies. Upper panel: human capital case with $\delta = 1$, and $\rho = 0.1$. Lower panel: calibrated economy with $\delta = 0.33$, and $\rho = 0.5$.

analyzed above. Long-run capital taxation is given by $\tau^{**} = 51\%$, as in Klein and Ríos-Rull (2003). We take average depreciation rate to be 10% per year. However, capital depreciates at a lower rate in the first period: we assume $\rho = 0.5$. This implies $\delta = 0.33$. Annual interest rates are set to 5%. Our assumption on τ^{**} implies that $A = 1.3$ (so that the distortionary cost of taxation under commitment equals 30%), and that the investment-to-output ratio is around 0.4, which is about one third larger than it is in the data.

The findings (see the lower panel of Figure 3 for a geometric illustration) are as follows.

1. In the economy with commitment, the steady-state tax rate is $\tau^* = 22\%$, i.e., significantly below the 50% rate under lack of commitment.
2. In the economy without commitment, the persistence of taxes is $-\lambda = 0.4$ (on a 4-year basis): taxes are highly persistent!
3. In the economy with commitment, the persistence of taxes is $-\delta(1 - \rho) = -0.165$ (again, on a 4-year basis): taxes oscillate.

These results are rather striking. First, the lack of commitment implies significantly higher long-run tax rates than if commitment were available. Second, though taxes are highly persistent under lack of commitment, the situation would be very different if commitment tools were available: tax rates would oscillate. Thus, both in terms of long-run levels and the dynamics towards the long-run level, there are important differences between the cases of commitment and no commitment. The dynamics are qualitatively different and the magnitudes of the differences are quite large.

Of course, the quantitative (and qualitative) results in this section depend both on the specific parameter values we used and on some of the functional-form assumptions in our analysis, such as the quasi-linearity of consumption and a production function which is linear in the accumulated input. Future work will determine how sensitive our calibration results are to these assumptions. If the sharp differences between the commitment and no-commitment outcomes are robust, however, it should prove fruitful to use our findings in order to assess, using estimation of the model based on time series and cross-section data, whether governments seem to have fiscal commitment.

5 Distortion smoothing with and without commitment

In this section we compare the determinants of the allocations with and without commitment. A common ground well suited for this builds on the idea, referred to above but not made formally explicit, of “distortion smoothing”. We explain how and why governments with and without the ability to commit optimize by smoothing distortions. This analysis is based on first-order conditions of the government’s problem that set marginal benefits of raising more revenue equal to marginal costs, and these first-order conditions have a different nature depending on whether or not there is commitment.

Our analysis of distortion smoothing also allows us to make comparisons with the principle of tax smoothing proposed in Barro (1979). Barro’s recipe built on a framework that can be thought of as one with quasi-linear utility (linear in consumption, nonlinear in effort) and a constant interest rate—features we have adopted as well—but with effort giving only a static payoff; here, effort gives a long-lived payoff, and the rate at which the payoff goes to zero is increasing (the quasi-geometric part). Thus, present values of tax liabilities due to building capital are what matter here, whereas in Barro’s model, each investment decision was only influenced by one tax rate.

We adopt the human capital version of our model here because it makes illustrations simpler. In it, the period utility function reads

$$(1 - \rho)i_{t-2} + i_{t-1} - g_t + Ag_t - i_t^2.$$

Ignoring any constraints, the derivative of this utility function with respect to g_{t+1} from the perspective of time t (recall the one period implementation lag) is $\beta(A - 1)$; this amount is strictly positive—on the margin, public consumption is worth more than private consumption—and it represents the marginal benefit of raising revenue. We define this “gap” as $\gamma_{g,t+1}$. Similarly, the marginal value of increasing investment at t is defined as $\gamma_{i,t} = -2i_t + \beta + \beta^2(1 - \rho)$: this is the difference between the marginal social value and cost

of investment, which is also positive in equilibrium since taxes distort investment downward. Whether the government has commitment or not, its optimal behavior dictates trading these gaps off against each other: raising revenue in order to gain $\gamma_{g,t+1}$ on the margin involves raising taxes, which distort investment and the marginal effect of this involves the γ_i 's. We will now see how this works in more detail.

It is useful to express policy as a function of government expenditure and private investment levels instead of using the tax rate as the choice variable. From the government's budget, the tax rate satisfies $\tau_t = g_t / ((1 - \rho)i_{t-2} + i_{t-1})$, i.e., the tax at t is a function of the expenditure at t and the investment levels at $t - 2$ and at $t - 1$, because the latter are the income base. We can now express the agent's first-order condition for investment at t by the following general condition

$$\eta(i_{t-1}, i_t, i_{t+1}, g_{t+1}, g_{t+2}) = 0. \quad (31)$$

This equation comes from the intertemporal condition (10)—where the choice variable investment in period t , i_t , is related to i_{t+1} and to the two tax rates τ_{t+1} and τ_{t+2} —replacing the tax rates as functions of government expenditures and investment levels, respectively. The constraint in (31) makes explicit the “budget externalities” in this model: private agents ignore the fact that increased investment, via the balanced government budget, indirectly raises the level of public expenditures. This is a positive externality since public goods are under-provided relative to the first best.

Thus, the Ramsey problem can be expressed as the maximization of $\sum_{t=0}^{\infty} \beta^t ((1 - \rho)i_{t-2} + i_{t-1} - g_t + Ag_t - i_t^2)$ subject to the sequence of constraints $\eta(i_{t-1}, i_t, i_{t+1}, g_{t+1}, g_{t+2}) = 0$ for $t \geq 0$, with i_{-2} , i_{-1} , and g_0 given. It is possible to think of the government maximization problem under lack of commitment in a parallel way: the objective is the same but at time t , when choosing i_t and g_{t+1} , the constraint is different: it is $\eta(i_{t-1}, i_t, (i_t), g_{t+1}, (i_t)) = 0$, where (i) and (i) are the mappings from the state variable—last period's investment in the human capital model—to the equilibrium choices for investment and public expenditures. This difference highlights the key role played by commitment: without it, the government is forced to view the future choices of i and g as beyond their current choice and instead determined by the “reaction functions” of future governments. Thus, when there is no commitment the current government can only influence future choices indirectly, by influencing the future state variable determining them.

Looking first at the problem under commitment, one can derive the following first-order condition for the choice at $t > 0$ of g_{t+1} :

$$\gamma_{g,t+1} = \frac{\partial \eta_t}{\partial g_{t+1}} \cdot \frac{di_t}{d\eta_t} \cdot \frac{dW_t}{di_t} + \frac{\partial \eta_{t-1}}{\partial g_{t+1}} \cdot \frac{di_{t-1}}{d\eta_{t-1}} \cdot \frac{1}{\beta} \cdot \frac{dW_t}{di_{t-1}}. \quad (32)$$

The left-hand side is the marginal benefit of raising revenue at t .³¹ The right-hand side measures the cost of increasing g_{t+1} on the margin, and this cost has several components. Here, $di_t/d\eta_t$ denotes the amount by which a marginal tightening of the constraint at time t forces investment in that period to fall, and dW_t/di_t is the total effect on utility as of t of increasing i_t . Since an increase in g_{t+1} tightens the constraints both at t and at $t - 1$, there are two terms on the right-hand side.

A first, basic observation is that the left-hand side of this equation is constant in this model: it equals $\beta(A - 1)$, since the net marginal benefit of transforming private into public consumption is constant. Thus, because the marginal benefits have to be constant over time, the marginal costs have to be constant: *distortion smoothing*. In the equivalent setup where the goal is to finance an exogenous expenditure stream by taxation at different points in time, the marginal benefit of raising revenue is also constant if interest rates are not affected by the policy, which by assumption they are not here (and they were not in Barro's setup).

Second, to see how the different gaps are traded off against each other, note that the effect on utility involves direct as well as indirect effects: it satisfies

$$\frac{dW_t}{di_t} \equiv \gamma_{i,t} - \left. \frac{dg_{t+1}}{di_t} \right|_{\eta_{t-1}=0} \cdot \beta \gamma_{g,t+1} - \left. \frac{dg_{t+2}}{di_t} \right|_{\eta_{t+1}=0} \cdot \beta^2 \gamma_{g,t+2}.$$

In words, there is a direct cost of decreasing i_t —it lowers utility by $\gamma_{i,t}$ —but there are also indirect costs due to the budget externalities: a decrease in i_t lowers the tax base and thus government expenditures at the adjacent dates $t - 1$ and $t + 1$, an effect which is ignored by the private agents.

³¹The derivations of the first-order conditions displayed in this section are contained in the Appendix.

At period 0, the first-order condition is simpler: it just has one cost term, so it reads

$$\gamma_{g,1} = \frac{\partial \eta_0}{\partial g_1} \cdot \frac{di_0}{d\eta_0} \cdot \frac{dW_0}{di_0}, \quad (33)$$

where dW_0/di_0 is now simpler as well: it equals $\gamma_{i,0} - \frac{dg_2}{di_0} \Big|_{\eta_1=0} \cdot \beta^2 \gamma_{g,2}$. Of course, these simplifications all capture how decisions made at time -1 cannot be affected by the policy choice in period 0.

The discrepancies between the first-order conditions at time 0 and at any time $t > 0$ also capture the time-inconsistency of the commitment solution. Turning to the first-order condition for g_{t+1} in an arbitrary period of the model without commitment, we obtain something very similar to equation (33):

$$\gamma_{g,t+1} = \frac{\partial \eta_t}{\partial g_{t+1}} \cdot \left(\frac{di_t}{d\eta_t} \right)^* \cdot \left(\frac{dW_t}{di_t} \right)^*. \quad (34)$$

The asterisks here represent the expressions being different under lack of commitment. Thus, under lack of commitment, there is only one cost term—that of lowering investment at t . And as in the first period of the commitment problem, the effect on utility, $(dW_t/di_t)^*$, does not involve past budget externalities: it equals $\gamma_{i,t} - \frac{dg_{t+2}}{di_t} \Big|_{\eta_{t+1}=0} \cdot \beta^2 \gamma_{g,t+2}$.³² Finally, the expression $(di_t/d\eta_t)^*$ is also different. This expression takes into account that a relaxation of the constraint at time t allows a direct increase in i_t but also has indirect effects on the other variables appearing in this constraint, and how these indirect effects play out depends on whether there is commitment or not, since the effects of future variables on the current constraint work differently in the two setups; for details, see the Appendix.

Comparing the first-order conditions with and without commitment—equations (32) and (34)—there is one key difference: there are two cost terms in the former case (for i_t and i_{t-1}) and one in the latter (for i_t). Thus, first of all, since distortion smoothing in the case with commitment takes more distortions into account, it leads to lower overall levels of distortions. Second, it does not make the marginal cost of raising revenues *on each investment decision* smooth, since a high distortion on i_t from g_{t+1} can be counteracted by a low one from g_{t+2} . In particular, a high initial distortion on i_0 does not imply a high distortion on i_1 , i_2 , and so on, since the investment gap $\gamma_{i,t}$ appears in two consecutive first-order conditions. Without commitment, the situation is different. Now optimal distortion smoothing imposes a high marginal cost period by period: each of the investment gaps appears only in one first-order condition leading to higher, and more persistent, distortions.

Finally, it may be useful to observe that in the special version of our model that collapses into Barro's framework—the case where $\delta = 1$ and $\rho = 1$ where effort gives only static returns—the solutions with and without commitment coincide.³³ Thus, in a framework where (i) effort gives static returns and (ii) the interest rate is exogenous, whether or not the government can commit is immaterial.

6 Stochastic government expenditure

Proposition 4 establishes conditions under which fluctuations in taxes and output are efficient. However, if there is no capital inheritance from the past (i.e., if $k_1^o = 0$), the optimal tax sequence is smooth; that is, the optimal tax oscillations that are produced in the model can all be traced back to an initial condition. This section discusses an extension where the result that the transitional dynamics of the optimal tax sequence have an oscillatory nature applies to economies where there is zero inelastic capital.

For simplicity, we restrict the analysis to overlapping generations economies ($\delta = 1$), and to make the point sharp we assume that $i_{-1} = 0$. We do not deal with lack of commitment here; the planner is assumed to have full commitment power over future policy choices, which in this case includes choosing a state-contingent tax plan. We introduce time variation and uncertainty in the marginal value of public goods; this choice is made for convenience only.

³²The expression $\frac{dg_{t+2}}{di_t} \Big|_{\eta_{t+1}=0}$ equals $\eta_{3,t+1}/\eta_{5,t+1}$ here as well as in the commitment case: the budget effect is the same in the two cases.

³³This is easy to verify as η no longer depends on the first and fifth arguments i_{t-1} and g_{t+2} , essentially making the second term in (32) cancel.

The uncertainty that we consider is limited to a one-time event only. More precisely, suppose that in periods zero and one, the value of the public good is $A = A_l$. In the beginning of period two, with probability p , A jumps to $A_h > A_l$ and stays there forever; with probability $1 - p$, A remains at A_l forever. To fix ideas, we interpret the shock as the start of a (permanent) war that makes public expenditures more socially valuable. This elementary stochastic process allows us to focus on the endogenous dynamics of taxes, as opposed to on dynamics that are driven by an exogenous stochastic process. The role of uncertainty is limited to triggering the transitional dynamics: we study, in a sense, the impulse-response function of taxes and government expenditure following a unique shock.

6.1 State-contingent taxes with balanced budget

At period zero, the planner sets τ_1 and a state-contingent tax plan, $\{\tau_{h,t}, \tau_{l,t}\}_{t=2}^{\infty}$. The sequence $\{\tau_{h,t}\}_{t=2}^{\infty}$ is implemented if $A = A_h$, whereas the sequence $\{\tau_{l,t}\}_{t=2}^{\infty}$ is implemented if $A = A_l$. The Ramsey problem can be formulated as follows.³⁴

$$\begin{aligned} & \max_{T_0^e, \tau_{h,2}, \{T_{h,t}, T_{l,t}\}_{t>0}} y_l(T_0^e) + \beta^2 (1 - \rho) p (A_h - A_l) \tau_{h,2} i(T_0^e) \\ & + \sum_{t=1}^{\infty} \beta^{t+1} (p \cdot y_h(T_{h,t}) + (1 - p) \cdot y_l(T_{l,t})), \end{aligned} \quad (35)$$

where

$$T_0^e = \beta \tau_1 + \beta^2 (1 - \rho) (p \tau_{h,2} + (1 - p) \tau_{l,2}), \quad (36)$$

$$T_{j,t} = \beta \tau_{j,t+1} + \beta^2 (1 - \rho) \tau_{j,t+2} \text{ for } t > 0, \quad (37)$$

$$y_j(T) = A_j \beta (1 + \beta (1 - \rho)) i(T) - i(T)^2 (2A_j - 1) \quad (38)$$

The term $\beta^2 (1 - \rho) p (A_h - A_l) \tau_{h,2} i(T_0^e)$ in the objective function is key for understanding the oscillations. If this term were zero (e.g., if either $\rho = 1$, or $p = 0$, or $A_h = A_l$), the Ramsey tax sequence would be smooth in each state of nature. This term captures the planner's incentive to over-tax (via $\tau_{h,2}$) the period-zero investment, $i(T_0^e)$, to which we will return later. It is convenient to replace $\tau_{h,2}$ by an expression in terms of the total tax sequence. Iterating forward on equation (36) yields

$$\beta \tau_{h,2} = \sum_{t=1}^{\infty} [-\beta (1 - \rho)]^{t-1} T_{h,t}.$$

Substituting away $\tau_{h,2}$, finally, yields the following primal formulation of the Ramsey problem:

$$\begin{aligned} & \max_{T_0^e, \{T_{h,t}, T_{l,t}\}_{t>0}} y_l(T_0^e) + \beta (1 - \rho) p (A_h - A_l) \sum_{t=1}^{\infty} [-\beta (1 - \rho)]^{t-1} T_{h,t} \cdot i(T_0^e) \\ & + \sum_{t=1}^{\infty} \beta^{t+1} (p \cdot y_h(T_{h,t}) + (1 - p) \cdot y_l(T_{l,t})). \end{aligned} \quad (39)$$

Differentiating (39) with respect to $T_{l,t}$ and $T_{h,t}$ yields

$$y'_l(T_{l,t}) = 0, \quad (40)$$

$$y'_h(T_{h,t}) = [-1]^{t-1} (1 - \rho)^t (A_h - A_l) i(T_0^e), \quad (41)$$

³⁴To understand the first term, note that the contribution to the planner's utility of the generation born at zero is given by

$$\beta i(T_0^e) ((1 + (A_l - 1) \tau_1) + \beta (1 - \rho) (p (1 + (A_h - 1) \tau_{h,2}) + (1 - p) (1 + (A_l - 1) \tau_{l,2}))) - (i(T_0^e))^2.$$

Rearranging terms and using the definitions of the function $y_l(T_0^e)$ and T_0^e given in equations (36) and (38) yields the first line of expression (35).

implying that $T_{l,t}$, as well the implied tax sequence, is constant and identical to the case in which there is no risk of war.³⁵ In case of war, instead, the sequence $T_{h,t}$ converges in an oscillatory fashion to a steady state such that $y'_l(T_h^*) = 0$. Notably, $T_{h,2} > T_h^*$, implying that $\tau_{h,2} > \tau_h^* \equiv (A_h - 1) / (2A_h - 1)$. Thus, conditionally on war, there is a tax hike in period two, followed by dampened oscillations. The implied tax sequence after period two follows the tax dynamics of Proposition 4, i.e., $\tau_{h,t+1} = \tau_h^* - (1 - \rho)(\tau_{h,t} - \tau_h^*)$. Of course, a higher value of public expenditures also leads to higher long-run tax rates: $\tau_h^* > \tau_l^* \equiv (A_l - 1) / (2A_l - 1)$.

Why oscillations can arise in case of war, even in the absence of inelastic capital, can be explained as follows. Conditionally on war, the generation born at time zero had invested more than future generations, since (i) taxation had been set lower in period one because public good provision was less valuable back then and (ii) agents had invested while attaching some probability to peace and low taxes being realized in future. Specifically, $i_0 = i_l + \beta(1 - \rho) \frac{A_h - A_l}{2A_l - 1} \frac{p}{2} \tau_{h,2}$, where i_l is investment in case of peace. Thus, the marginal cost of raising tax revenue for the Ramsey planner is lower in period two than in steady state. Even if she cannot “surprise” agents (as she is committed to a state-contingent tax plan), the planner promises high taxes in period two conditionally on war. Oscillations occur thereafter from the same mechanism as described in the benchmark model.

Two particular cases are worth emphasizing. First, if $\rho = 1$, i.e., agents only work in the first period and taxes distort their static labor supply, there is no scope for inducing oscillations after the war starts. In this case $\tau_{h,2} = \tau_h^*$ and the Ramsey tax sequence is perfectly smooth after the first-period upward jump. Second, suppose that $p = 1$, i.e., that war is perfectly anticipated. In this case, there are still fluctuations, even though there is no uncertainty. The reason is that even if agents anticipate the increase in future taxation, the government has an incentive to spend less in period one, since the marginal utility of the public good is low then and the government cannot save. Thus, in period two, a large inelastic tax base is inherited, and the planner has an incentive to initiate the oscillations.

Finally, τ_1 is also affected by the probability of a war. The comparative statics here are somewhat involved. However, numerical analysis suggests that increasing the probability of a war decreases τ_1 . As the war becomes more likely, the planner becomes more eager to induce large investments in period zero to increase the future tax base. Since she cannot accumulate assets, she attains this goal by reducing taxation in period one: a form of public savings.

To illustrate the effects, we display a numerical example. We set $\beta = p = 0.5$, $\rho = 0$, $A_l = 1.6$, and $A_h = 2$. Since $\rho = 0$, fluctuations do not dampen over time. In this case (see Figure 4), the Ramsey sequence implies $\tau_1 = 0.154$, $\tau_{h,2} = 0.632$, and $\tau_{l,2} = 0.273$. If there is war, taxes fluctuate between 0.63 and 0.04. If there is no war, the tax rate is constant at 0.27.

6.2 Allowing government borrowing and lending

The constraint that the government cannot borrow or lend becomes important when the value of government spending changes over time. In the previous section, the planner would have liked not to spend its budget in period zero, but rather save for the event of a war. The equivalence between the standard capital taxation problem with exogenous government expenditure discussed in section 2 and the public good provision problem under a balanced budget ceases to hold: access to capital markets gives the government a useful policy instrument for smoothing distortions. In this section we allow this additional instrument. The main finding is that while this affects the Ramsey solution, it does not eliminate the oscillations.

We rule out Ponzi schemes by assuming that from period two onwards, i.e., after uncertainty is resolved, an intertemporal budget constraint must hold. We continue to assume that the government cannot tax at time zero and that $\beta = (1 + r)^{-1}$. The expected future marginal value of the public good, $pA_h + (1 - p)A_l$, exceeds A_l , the marginal value of the public good in period one. Because of risk neutrality, the government will therefore spend no revenue on public goods in the first period and instead accumulate a budget surplus. Hence, assuming that the government starts off with zero assets, we have $b_2 = -\frac{1}{\beta}\tau_1 i_0$.

The main results are as follows. In the case of certain war ($p = 1$), the solution features $\tau_1 = \tau_{h,2} = \tau_h^*$, implying no dynamics: a perfectly anticipated war does not, alone, induce any fluctuations, as long as the

³⁵Closed-form expressions for $\tau_{h,2}$ and $\tau_{l,2}$ can be backed out from (40)-(41) using the definitions (36)-(38). It is also possible to back out τ_1 . The resulting expressions are provided in the appendix. The result that the no-war outcome is entirely smooth depends on the particular specification chosen where one of the two realization of the stochastic process coincides with the productivity of the public good at time zero.

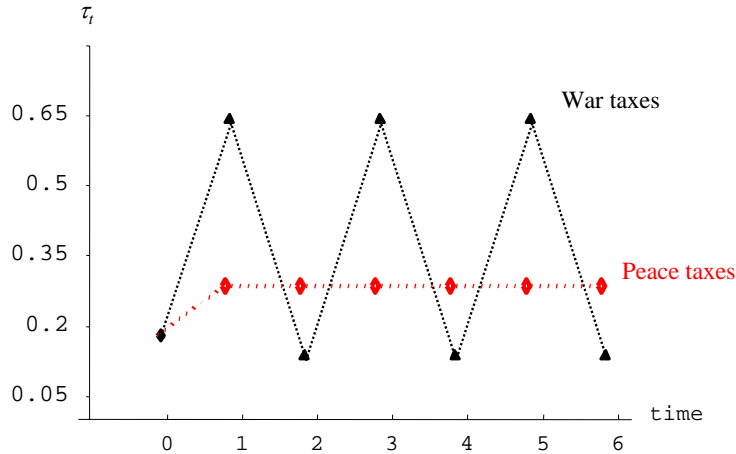


Figure 4: Taxes under war and peace, no government debt, $\beta = p = 0.5$, $\rho = 0$, $A_l = 1.6$, and $A_h = 2$

government can save or borrow. In the general case where $p \in (0, 1)$, however, $\tau_{h,2} > \tau_h^*$ and uncertainty triggers dynamics. Interestingly, oscillations arise in this case even if the war does not materialize. The tax sequence follows the dynamics of Proposition 4 under both war and peace, although, naturally, both the initial conditions $(\tau_{h,2}, \tau_{l,2})$ and the steady states (τ_h^*, τ_l^*) are different. The details of the analysis are available upon request.

Consider the same numerical example as before. The Ramsey sequence now implies $\tau_1 = 0.32$, $\tau_{h,2} = 0.47$, and $\tau_{l,2} = 0.08$. Conditionally on war, the tax rate fluctuates between $\tau_{h,t} = 0.47$ and $\tau_{h,t} = 0.19$. Conditionally on peace, the tax rate fluctuates between $\tau_{l,t} = 0.08$ and $\tau_{l,t} = 0.46$. Government savings in the first period amount to $-b_2 = 0.33$.

Moreover, taxation in period one is higher than in the case where the government had no access to capital markets: the government self-insures against the event of a war. This is in contrast with the case without government asset accumulation, when the only (less efficient) way the planner could prepare for a war was by encouraging human capital accumulation through initially low taxes. Taxes are now smoother conditionally on war and more volatile conditionally on peace.

Note that when $p \in (0, 1)$, a market for safe lending and borrowing does not span all states of the world—financial markets are still incomplete. As we will see in the next subsection, this incompleteness is crucial for the existence of tax fluctuations.

6.3 Allowing state-contingent government debt

Suppose, finally, that in the first period there are two state-contingent assets paying one unit of the consumption good in the state of war (peace) and zero in the state of peace (war). 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 . We continue to assume a one-period implementation lag. Thus, both taxes and government assets with time indices equal to zero can be ignored. The government budget constraints can now be consolidated into one constraint:

$$g_1 + \sum_{t=2}^{\infty} \sum_{j \in \{h,l\}} q_{j,t} g_{j,t} = \tau_1 i_0 + \sum_{t=2}^{\infty} \sum_{j \in \{h,l\}} q_{j,t} \tau_{j,t} ((1 - \rho) i_{j,t-2} + i_{j,t-1}). \quad (42)$$

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 $q_{l,t} = \beta^t (1 - p)$. Since, in addition, preferences over

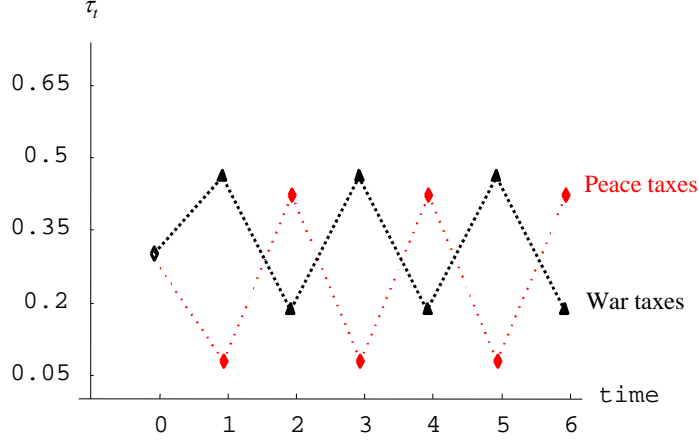


Figure 5: Taxes under war and peace, with government debt, $\beta = p = 0.5$, $\rho = 0$, $A_l = 1.6$, and $A_h = 2$.

public-good provision are linear, the planner will choose zero public good provision in the case of peace, and concentrate all spending in the state of war. Hence, $g_1 = g_{l,t} = 0$, for all $t \geq 2$.

Using the obvious notation $p_h = p = 1 - p_l$, the planner's time-zero objective function reads

$$\beta(1 - \tau_1) i_0 - i_0^2 + \beta A_l g_1 + \sum_{t=2}^{\infty} \sum_{j \in \{h,l\}} \beta^t p_j ((1 - \tau_{j,t}) ((1 - \rho) i_{j,t-2} + i_{j,t-1}) + A_j g_{j,t} - \beta^{-1} i_{j,t-1}^2). \quad (43)$$

Using the budget constraint, (42), and the facts, established above, that $q_{j,t} = \beta^t p_j$ and $g_1 = g_{l,t} = 0$ for all t , we can eliminate terms involving g 's and rewrite the planner's objective function, (43), as

$$\beta(1 + (A_h - 1) \tau_1) i_0 - i_0^2 + \sum_{t=2}^{\infty} \sum_{j \in \{h,l\}} \beta^t p_j ((1 + (A_h - 1) \tau_{j,t}) ((1 - \rho) i_{j,t-2} + i_{j,t-1}) - \beta^{-1} i_{j,t-1}^2). \quad (44)$$

As expression (44) shows, the marginal value of tax revenue is A_h in all states of nature. Hence, the optimal tax sequence is identical to the case in which war occurs with probability one discussed in section 6.2. Namely, when the government has access to state-contingent asset markets, then, $\tau_{j,t} = \tau_h^*$ for all j and t : no tax fluctuations arise as long as $i_{-1} = 0$. Taxes in all periods and all states are used to finance high provision in the state of war, while in the state of peace the government expenditure is zero. Specifically, if the public good provision is g_h in the case of a sure war, then the government will provide g_h/p in the stochastic case. Clearly this stark result hinges on the assumption that the marginal value of public expenditure is constant, leading to a corner solution (all revenue is spent in one state of nature). If the marginal value of public expenditure were decreasing, the government would spend some revenue in both states of nature, although public good provision would still be higher in the case of war than in that of peace. The main insights would generalize to this case. In particular, tax distortions would still be equalized across states.

7 Conclusions

This paper analyzes the optimal timing of taxes on capital. Our point of departure from a standard Chamley-Judd setup is to relax a common and seemingly innocent assumption in the standard optimal taxation litera-

ture, namely, that capital depreciation is geometric. Our specification is motivated by empirically estimated depreciation schedules, which feature accelerating depreciation rates as capital ages. Under this specification, the optimal path of capital taxes involves oscillating tax rates. We also show that the standard celebrated front-loading result is a knife-edge case, hinging on the rate of depreciation being constant over time. Furthermore, the optimality of fluctuating tax rates relies on the government being able to commit to the path of future tax rates. We contrast this case with a Markov equilibrium where the planner cannot commit. We find that without commitment the equilibrium tax sequence is smoother than under commitment, i.e., that tax dynamics are more persistent.

Though our model has typical neoclassical features, the analysis is simplified by assuming linear utility and quadratic investment costs. The implied time additivity and constant interest rates allow an especially illuminating analysis of the effects of taxation: investment decisions depend on present values of tax rates, independently of the consumption path followed. It should be important to extend our results quantitatively to the more standard neoclassical setup with endogenous interest rates.

If vintage-specific taxation were allowed in standard models like those in Chamley (1986) and Judd (1985), the taxation problem would become trivial: 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 0. 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. A special case of interest is that where the depreciation rate is zero, because in this case, the tax rate specific to time-zero capital would remain positive even in the long run. Thus, in general the front-loading result in the standard literature is to a large extent driven by assuming that different vintages cannot be taxed at different rates.

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). They show that when age-dependent taxation is not allowed and life-cycle motives are considered, the optimal capital taxation is not zero in the long run. Similarly, and perhaps more closely related to our findings, Correia (1996) shows that long-run taxes on capital income are not necessarily zero if there is another factor input which cannot be taxed. In this case, it would generally be optimal to tax this other factor, and taxes on capital income can be useful as an imperfect replacement for such a tax.

In our analysis, we interpret the Markov equilibrium as a lower bound on public commitment. The degree of commitment actual governments have could, in principle, be anything from full commitment to no commitment. Some commitment could be due to the existence of some explicit commitment technology. Alternatively, it could be sustained with reputation within the context of dynamic game between consecutive governments (as in e.g., Abreu et al., 1990). Empirically, it is an open issue to what extent governments actually have the ability to commit. The fact that our benchmarks with and without commitment yield such starkly different predictions, both qualitatively and quantitatively, provides us with a potential test for the degree of commitment of actual governments. To make this test sharper, more elaborate models with concave utility and more general production functions should be constructed and solved numerically. The degree to which governments can use investment credits to partly substitute for cohort specific taxes should also be estimated. We leave these tasks for future research.

8 References

- Abreu, Dilip, David Pearce and Ennio Stacchetti, 1990, "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Journal of Economic Theory* 56, 383–396.
- Atkeson, Andrew, V.V. Chari and Patrick Kehoe, 1999, "Taxing Capital Income: A Bad Idea", *Federal Reserve Bank of Minneapolis Quarterly Review*, Summer, Vol. 23 no. 3, pp. 3–17.
- Barro, Robert J., 1979, "On the Determination of Public Debt", *Journal of Political Economy* 87, 940–971.
- Chari, V.V. and Patrick J. Kehoe, 1990, "Sustainable Plans," *Journal of Political Economy* 98, 783–802.
- Coen, Robert M., 1975, "Investment Behavior, The Measurement of Depreciation, and Tax Policy" *American Economic Review*, 65, 59–74.

- Chamley, Christophe, 1986, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives", *Econometrica* 54:3, 607–622.
- Chari, V.V., and Patrick J. Kehoe, 1999. "Optimal Fiscal and Monetary Policy," NBER Working Papers 6891.
- Correia, Isabel, 1996, "Should Capital Income Be Taxed in the Steady State?", *Journal of Public Economics*, 60, 147-151.
- Dunn, Wendy, Mark Doms, Stephen Oliner, and Daniel Sichel, 2004, "How Fast Do Personal Computers Depreciate? Concepts and New Estimates", NBER Working Paper no. 10521.
- Erosa, Andres, and Martin Gervais, 2002, "Optimal Taxation in Life-Cycle Economies," with Andrés Erosa, *Journal of Economic Theory*, 105, 338–369.
- Fernandez Villaverde, Jesus and Aleh Tsyvinski, 2004, "Optimal Fiscal Policy in a Business Cycle Model without Commitment", Mimeo, University of Pennsylvania.
- Geske, Michael J., Valerie Ramey, and Matthew D. Shapiro, 2003, "Why Do Computers Depreciate?" Mimeo, University of Michigan.
- Hassler, John, P. Krusell, Kjetil Storesletten and Fabrizio Zilibotti, 2004, "The Dynamics of Government", Mimeo IIES.
- Hassler, John, J.-V. Rodriguez Mora, Kjetil Storesletten and Fabrizio Zilibotti, 2003, "The Survival of the Welfare State", *American Economic Review*, 93, 87–112.
- Hassler, John, Kjetil Storesletten and Fabrizio Zilibotti, 2003a, "Dynamic Political Choice in Macroeconomics", *Journal of the European Economic Association*, 1, 543–552.
- Hassler, J., Kjetil Storesletten and Fabrizio Zilibotti, 2003b, "Democratic Public Good Provision", CEPR Discussion Paper no. 4044.
- Hulten, Charles R., and Frank C. Wykoff, 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, Kenneth L., 1985, "Redistributive Taxation in a Simple Perfect Foresight Model", *Journal of Public Economics* 28, 59–83.
- Klein, Paul, Per Krusell and J.V. Ríos-Rull, 2004, : "Time-consistent Public Expenditures", Mimeo, University of Pennsylvania.
- Klein, Paul and J.V. Ríos-Rull, 2003, "Time-Consistent Optimal Fiscal Policy", *International Economic Review*, 44, 1217–1245.
- Krusell, Per and Anthony A. Smith, Jr., 2003, "Consumption-Savings Decisions with Quasi-Geometric Discounting", *Econometrica*, 71, 365–375.
- Krusell, Per and Jose V. Ríos-Rull, 1999, "On the Size of Government: Political Economy in the Neoclassical Growth Model", *American Economic Review*, 89:5, 1156–1181.
- Krusell, Per, Fernando Martin and Jose V. Ríos-Rull, 2004, "On the Determination of Government Debt", Mimeo, University of Rochester.
- Kydland, Finn and Edward C. Prescott, 1977, "Rules rather than discretion: The inconsistency of optimal plans", *Journal of Political Economy*, 85, 473–490.
- Lindbeck, Assar and Jörgen Weibull, 1987, "Balanced-budget redistribution as the outcome of political competition", *Public Choice* 52, 273–297.
- Lucas, Robert E., Jr. and Nancy Stokey, 1983, "Optimal Monetary and Fiscal Policy in an Economy Without Capital", *Journal of Monetary Economics* 12, 55–94.
- Oliner, Stephen D., 1996, "New Evidence on the Retirement and Depreciation of Machine Tools," *Economic Inquiry*, 34, 57–77.
- Pakes, Ariel and Zvi Griliches, 1983, Estimated Distributed Lags with Short Panels with an Application to the Specification of Depreciation Patterns and Capital Stock Constructs, NBER Working Paper no. 933.
- Penson, John B. Jr., Dean W. Hughes and Glenn L. Nelson, 1977, "Measurement of Capacity Depreciation Based on Engineering Data" *American Journal of Agricultural Economics*, 59, 321–329.
- Penson, John B. Jr., Robert Romain, and Dean W. Hughes, 1981, "Net Investment in Farm Tractors: An Econometric Analysis" *American Journal of Agricultural Economics*, 63, 629–635.
- Rogoff, Kenneth, 1985, "The optimal degree of precommitment to an intermediate monetary target", *Journal of International Economics*, 18, 1169–1190.

9 Appendix

9.1 Derivation of (20)

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

$$T_{t-1} - \beta\tau_t + (T_t - \beta\tau_{t+1})\beta\delta(1 - \rho) = \beta(1 - \rho\delta)T_t. \quad (45)$$

Forward substitution implies

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

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

$$T_0 = \beta\tau_1 + \beta(1 - \rho\delta) \sum_{s=0}^{\infty} (-\beta\delta(1 - \rho))^s T_{s+1}. \quad (46)$$

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 (46), implies that

$$\sum_{s=2}^{\infty} \beta^s (1 - \delta)^{s-2} \tau_s = \beta \sum_{s=0}^{\infty} (-\beta\delta(1 - \rho))^s T_{s+1}. \quad (47)$$

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 (46)-(47), as

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

which is expression (20) in the paper.

9.2 Details of the proof of Proposition 4

Solving (45) for τ_{t+1} yields

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

Using (24) and the expression for τ^* given in the text to replace T_{t-1} and T_t yields (for $t \geq 1$)

$$\tau_{t+1} = \tau \frac{1 + \beta\delta(1-\rho)}{\beta\delta(1-\rho)} - \tau \frac{1 + \beta(1-\delta)\delta(1-\rho)}{\delta^2(1-\rho)^2\beta^2} 2k_1(-\delta(1-\rho))^t - \frac{\tau_t}{\beta\delta(1-\rho)}.$$

The complete solution to this difference equation can be written

$$\tau_t = \tau^* + \frac{1 + \beta\delta(1-\delta)(1-\rho)}{\beta(1-\beta\delta^2(1-\rho)^2)} 2k_1\tau(-\delta(1-\rho))^{t-1} + c \left(-\frac{1}{\beta\delta(1-\rho)} \right)^t,$$

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, $-1/(\beta\delta(1-\rho))$, 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 (24) is determined by setting $c = 0$. Writing this solution recursively yields the solution in Proposition (4).

The non-diverging dynamics implies that it is sufficient that $\tau_1 < 1$ and $\tau_2 > 0$ to ensure that the bound τ_t in $[0, 1]$ never binds. It is immediate to verify that this is the case for a positive range of $(A-1)k_1^o$ sufficiently close to zero.

9.3 Details of the proof of Proposition 6

First, we provide the expressions of τ^{**} , λ , α_{01} , and α_{11} :

$$\tau^{**} = \tau^* + \frac{2A(1-\rho)(1-\beta(1-\rho)\phi)\tau^*}{2(A-1)(1-\rho) + 2A - 1 + \left(\sqrt{(2A-1)^2 - 4A(A-1)\beta(1-\rho)^2} \right)} \quad (48)$$

$$\lambda = -(1-\rho)\phi \quad (49)$$

$$\alpha_{01} = \frac{A\beta\phi(1-\rho)^2 + A - 1}{2A - 1 + A(\beta\phi(1-\rho)^2)} > 0, \quad (50)$$

$$\alpha_{11} = \frac{2\phi}{\beta(1 + \beta\phi(1-\rho)^2)} > 0, \quad (51)$$

where

$$\phi \equiv -\frac{2A - 1 - \sqrt{(2A-1)^2 - 4A(A-1)\beta(1-\rho)^2}}{2A\beta(1-\rho)^2} \in (-1, 0). \quad (52)$$

That $\phi \in (-1, 0)$ and, hence, that $\alpha_{01} > 0$ and $\alpha_{11} > 0$ can be established by straightforward algebraic manipulations. From $\phi \in (-1, 0)$, it follows immediately from (48) that $\tau^{**} > \tau^*$.

Next, we prove the proposition. Consider equation (27) in Definition 5. Using the linear guesses (29) and (30), we can rewrite (27) as

$$i_{t-1} = \frac{\kappa}{2} - \frac{\beta}{2}\tau_t - \frac{\beta^2}{2}(1-\rho)(\alpha_{01} + \alpha_{11}k_{t+1}^o). \quad (53)$$

Now using the second guess to eliminate k_{t+1}^o from the law of motion of capital, (26), and solving (53) for $\beta\tau_t$ yields

$$\beta\tau_t = Q_0 + Q_1 i_{t-1}, \quad (54)$$

where

$$\begin{aligned} Q_0 &= \beta(1 + \beta(1-\rho)(1 - \alpha_{01})), \\ Q_1 &= -((1-\rho)\beta^2(1-\rho)\alpha_{11} + 2). \end{aligned}$$

Next, we use (54) to eliminate τ_t in the planner's recursive objective function, (28), obtaining

$$W(k_t^o) = \max_{i_{t-1}} \{(\beta + (A-1)(Q_0 + Q_1 i_{t-1}))(k_t^o + i_{t-1}) - i_{t-1}^2 + \beta W(k_{t+1}^o)\}.$$

The first-order and envelope conditions for this problem are

$$\beta + (A-1)((Q_0 + Q_1(2i_{t-1} + k_t^o)) - 2i_{t-1} + \beta(1-\rho)W'(k_{t+1}^o)) = 0 \quad (55)$$

and

$$W'(k_t^o) = \beta + (A-1)(Q_0 + Q_1 i_{t-1}), \quad (56)$$

respectively.³⁶ Leading (56) by one period and using it to eliminate $W'(k_{t+1}^o)$ from (55) allows us to establish

$$W_0 + W_1 i_{t-1} + W_2 k_t^o = Z_1 i_t, \quad (57)$$

where

$$\begin{aligned} W_0 &= -(1 + \beta(1-\rho))(\beta + (A-1)Q_0), \\ W_1 &= -2((A-1)Q_1 - 1), \\ W_2 &= -(A-1)Q_1, \\ Z_1 &= \beta(1-\rho)(A-1)Q_1. \end{aligned}$$

Next, we use (30) to eliminate i_t from (57), and rearrange terms to obtain

$$i_{t-1} = \frac{-(Z_1 \alpha_{02} - W_0) + W_2 k_t^o}{Z_1 \alpha_{12}(1-\rho) - W_1},$$

which is a linear (affine) function of k_t^o , in line with our guess (30). We can now equate coefficients to obtain

$$\alpha_{02} = \frac{W_0 - Z_1 \alpha_{02}}{Z_1 \alpha_{12}(1-\rho) - W_1}, \quad (58)$$

$$\alpha_{12} = \frac{W_2}{Z_1 \alpha_{12}(1-\rho) - W_1}. \quad (59)$$

To verify our other guess, (29), we return to the condition (54). Using the guess (30) to eliminate i_{t-1} yields

$$\beta \tau_t = Q_0 + Q_1 \alpha_{02} + Q_1 \alpha_{12} k_t^o,$$

which is also linear (affine), consistently with our guess. Equating coefficients leads to

$$\beta \alpha_{11} = Q_1 \alpha_{12}, \quad (60)$$

$$\beta \alpha_{01} = Q_0 + Q_1 \alpha_{02}. \quad (61)$$

Equations (58)-(61) define a system of non-linear equations in the four unknown coefficients α_{01} , α_{11} , α_{02} , and α_{12} . Its solution is given by (50), (51), and

$$\begin{aligned} \alpha_{02} &= \frac{1}{2} A \beta \frac{1 + \beta(1-\rho)(1 + \phi(1-\rho)(1 + \beta(1-\rho)))}{2A - 1 + A\beta\phi(1-\rho)^2} \\ \alpha_{12} &= \phi, \end{aligned} \quad (62)$$

where ϕ is a root of

$$A\beta(1-\rho)^2 \phi^2 + (2A-1)\phi + A-1 = 0. \quad (63)$$

Equation (63) has two roots but one of them can be ruled out since it implies explosive dynamics. The stable root is given by (52).

³⁶The linear-quadratic nature of the problem implies trivially that second-order conditions are satisfied. The assumption on $(A-1)k_1^o$ not being too large implies interior optima.

The state variable has the following equilibrium law of motion:

$$k_{t+1}^o = (1 - \delta)k_t^o + (1 - \rho\delta)i_{t-1} = (1 - \rho)(\alpha_{02} + \alpha_{12}k_t^o).$$

Hence, the steady-state value for capital is $k^o = (1 - \rho)\alpha_{02}/(1 - \alpha_{12}(1 - \rho))$, and the steady-state tax rate is $\tau^{**} = \alpha_{01} + \alpha_{11}k^o$. Inserting the expressions for α_{01} and α_{11} delivers the expression in (48).

The decision rules for capital and taxes can be expressed as

$$\begin{aligned} k_{t+1}^o - k^o &= (1 - \rho)\alpha_{12}(k_t^o - k^o) \\ \tau_{t+1} - \tau^{**} &= (1 - \rho)\alpha_{12}(\tau_t - \tau^{**}), \end{aligned}$$

and setting $-\lambda = (1 - \rho)\alpha_{12}$ using (62) then leads to the expression for λ in (49).

9.4 Details of the analysis of Section 5

Denoting the multiplier for the private first-order constraint at t by $\beta^t \lambda_t$, the government's Lagrangian can be written as

$$\sum_{t=0}^{\infty} \beta^t \left((1 - \rho)i_{t-2} + i_{t-1} - g_t + Ag_t - i_t^2 - \lambda_t \eta(i_{t-1}, i_t, i_{t+1}, g_{t+1}, g_{t+2}) \right).$$

Letting $\eta_{t,j}$ denote the j th partial of $\eta(i_{t-1}, i_t, i_{t+1}, g_{t+1}, g_{t+2})$, the first-order conditions for the choices of i_t and g_t become, for $t > 0$,

$$\gamma_{i,t} - \beta^{-1} \lambda_{t-1} \eta_{t-1,3} - \lambda_t \eta_{t,2} - \beta \lambda_{t+1} \eta_{t+1,1} = 0, \quad (64)$$

and

$$\gamma_{g,t} - \beta^{-1} \lambda_{t-2} \eta_{t-2,5} - \lambda_{t-1} \eta_{t-1,4} = 0, \quad (65)$$

where $\gamma_{g,t}$ and $\gamma_{i,t}$ are defined as in the text. To find the first-order condition stated in terms of these wedges, it is necessary to eliminate the multipliers in the first-order conditions above. Thus, use equation (64) for period t and equation (65) for periods $t+1$ and $t+2$, since then we have three equations with the unknowns λ_{t-1} , λ_t , and λ_{t+1} . Thus, λ_t can be solved for as a function of $\gamma_{i,t}$, $\gamma_{g,t+1}$, and $\gamma_{g,t+2}$. Substitute the solutions for λ_t and λ_{t-1} into equation (65) at time $t+1$ and we obtain the final expression for the first-order condition for the government's policy choice:

$$\begin{aligned} \gamma_{g,t+1} &= \beta^{-1} \eta_{t-1,5} D_{t-1} \left(\gamma_{i,t-1} - \frac{\eta_{t-2,3}}{\eta_{t-2,5}} \gamma_{g,t} - \beta \frac{\eta_{t,1}}{\eta_{t,4}} \gamma_{g,t+1} \right) \\ &\quad + \eta_{t,4} D_t \left(\gamma_{i,t} - \frac{\eta_{t-1,3}}{\eta_{t-1,5}} \gamma_{g,t+1} - \beta \frac{\eta_{t+1,1}}{\eta_{t+1,4}} \gamma_{g,t+2} \right), \end{aligned}$$

where D_t , the measure of how much a unit increase of the constraint at t is worth in terms of effort i_t , is calculated as

$$D_t \equiv \frac{1}{\eta_{t,2} - \eta_{t,4} \frac{\eta_{t-1,3}}{\eta_{t-1,5}} - \eta_{t,5} \frac{\eta_{t+1,1}}{\eta_{t+1,4}}}.$$

This is equation (32) in the text.

Now consider the wedges at the beginning of time. Assuming that g_0 (and τ_0 , indirectly) is set at the beginning of time, and following similar steps to those above, the first-order condition for g_1 can be written

$$\gamma_{g,1} = \eta_{0,4} D_0 \left(\gamma_{i,0} - \beta \frac{\eta_{1,1}}{\eta_{1,4}} \gamma_{g,2} \right),$$

where

$$D_0 \equiv \frac{1}{\eta_{0,2} - \eta_{0,5} \frac{\eta_{1,1}}{\eta_{1,4}}}.$$

This is equation (33) in the text.

Consider, finally, the determination of the first-order condition under lack of commitment. To this end, first state the government's problem as a dynamic program with the previous investment choice, i_{t-1} , as state variable (recall that since $\delta = 1$ and $\rho = 0$ we have $k_{t+1}^o = i_{t-1}$):

$$W(i_{t-1}) = \max_{i_t, g_{t+1}} -i_t^2 + \beta((1-\rho)i_{t-1} + i_t - g_{t+1} + Ag_{t+1}) + \beta^2 W(i_t)$$

subject to

$$\eta(i_{t-1}, i_t, \mathcal{I}(i_t), g_{t+1}, \mathcal{G}(i_t)) = 0.$$

Here, $\mathcal{I}(i_{t-1}) = i_t$ is the policy rule for investment and $\mathcal{G}(i_{t-1}) = g_{t+1}$ is the policy rule for public expenditures. Taking first-order conditions, we obtain

$$\beta - 2i_t + \beta^2 W'_{i_t} = \lambda_t(\eta_{t,2} + \mathcal{I}'(i_t)\eta_{t,3} + \mathcal{G}'(i_t)\eta_{t,5})$$

for the choice of i_t and $\gamma_{g,t+1} = \lambda_t \eta_{t,4}$ for the choice of g_{t+1} . Solving for λ_t from the first-order condition for i_t , we obtain

$$\gamma_{g,t+1} = D_t \eta_{t,4} (\beta - 2i_t + \beta^2 W'(i_t)),$$

where

$$D_t \equiv \frac{1}{\eta_{t,2} + \mathcal{I}'(i_t)\eta_{t,3} + \mathcal{G}'(i_t)\eta_{t,5}}.$$

Since the envelope theorem gives

$$W'(i_t) = \beta(1-\rho) - \lambda_t \eta_{t,1},$$

evaluated the following period this expression and the first-order condition for g_{t+1} deliver the “distortion-smoothing” condition specifying how trade-offs between wedges occur in the model without commitment. It reads

$$\gamma_{g,t+1} = \eta_{t,4} D_t (\gamma_{i,t} - \beta \frac{\eta_{t+1,1}}{\eta_{t+1,4}} \gamma_{g,t+2}).$$

This is equation (34) in the text. It differs from the period-0 first-order condition from the commitment problem, (33), only in how D_t is determined. The expression D_t determines a key component of how the change in g_{t+1} influences i_t , via the implementability constraint. Here, an increase in i_t changes η_t exactly by $1/D_t$, and this expression includes the total effect on how a change in i_t would influence the future government behavior that feeds back to the current constraint (i_{t+1} and g_{t+2}). In D_0 of the commitment problem, in contrast, the current government can control future decisions and the effects of future government behavior on the current constraint are partial—they are derived keeping future constraints constant. Thus, whereas we have $\mathcal{I}'(i_t)\eta_{t,3} + \mathcal{G}'(i_t)\eta_{t,5}$ in D_t here, in D_0 we just have $\frac{\eta_{1,1}}{\eta_{1,4}}$.

9.5 Details of the analysis of Section 6

In the case in which the government has no access to capital markets (section 6.1), the state-contingent tax plan in the first and second period prescribes:

$$\begin{aligned} \tau_1 &= \frac{A_l - 1}{2A_l - 1} - p \frac{\beta(1-\rho)}{2A_l - 1} ((A_h + A_l - 1)\tau_h - (A_l - 1)), \\ \tau_{l,2} &= \frac{A_l - 1}{2A_l - 1}, \\ \tau_{h,2} &= \frac{A_h - 1}{2A_h - 1} + \frac{(1-\rho)(A_h - A_l)}{2A_h - 1} \cdot \\ &\quad \frac{A_l(2A_h - 1)(1 + \beta(1-\rho)(1-p)) + \beta(1-\rho)pA_h(A_h + A_l - 1)}{(2A_l - 1)(2A_h - 1)(1 - \beta(1-\rho)^2) - \beta(1-\rho)^2 p(A_h - A_l)^2}, \end{aligned}$$

where the second term in the expression of $\tau_{h,2}$ is positive, implying that $\tau_{h,2} > \tau_h^* = \frac{A_h - 1}{2A_h - 1}$.³⁷

³⁷The denominator of the last term on the right-hand side term of the expression can be negative. However, recall that we are restricting attention to the region of the parameter space where taxes are strictly inside the unit interval at all times and for all realizations. When this restriction is taken into account, the denominator is unambiguously positive, implying that $\tau_{h,2} > \tau_h^*$.

10 Appendix B (not for publication)

10.1 Dynamics of gross output

To characterize private output dynamics, we note that output gross of adjustment costs is

$$Y_t = k_t^o + i_{t-1}.$$

Using $i_t = \frac{1}{2}(\kappa - T_t)$ and $T_t = \tau^* \left(\kappa + 2k_1^o (-\delta(1-\rho))^t \right)$, implying $i_t = \frac{1}{2}\kappa(1-\tau^*) - \tau^*k_1^o(-\delta(1-\rho))^t$ in (13) we obtain

$$\begin{aligned} k_{t+1}^o &= k_1^o(1-\delta)^t + (1-\rho\delta) \sum_{s=0}^{t-1} (1-\delta)^{t-1-s} i_s \\ &= k_1^o(1-\delta)^t + \frac{1-\rho\delta}{2} \kappa(1-\tau^*) \sum_{s=0}^{t-1} (1-\delta)^{t-1-s} - (1-\rho\delta) \tau^* k_1^o \sum_{s=0}^{t-1} (1-\delta)^{t-1-s} (-\delta(1-\rho))^s \\ &= \frac{1-\rho\delta}{2} \kappa(1-\tau^*) \frac{1-(1-\delta)^t}{\delta} + k_1^o(1-\tau^*)(1-\delta)^t + k_1^o \tau (-\delta(1-\rho))^t. \end{aligned}$$

Gross private output in period $t+1$ is

$$\begin{aligned} k_{t+1}^o + i_t &= \frac{1-\rho\delta}{2} \kappa(1-\tau^*) \frac{1-(1-\delta)^t}{\delta} + k_1^o(1-\tau^*)(1-\delta)^t + k_1^o \tau (-\delta(1-\rho))^t \\ &\quad + \frac{1}{2} \kappa(1-\tau^*) - \tau^* k_1^o (-\delta(1-\rho))^t \\ &= (1-\tau^*) \left(\kappa \frac{(1+\delta(1-\rho))}{2\delta} + \left(k_1^o - \frac{\kappa}{2\delta} (1-\rho\delta) \right) (1-\delta)^t \right), \end{aligned}$$

implying monotone dynamics and constant output when $\delta = 1$. Adjustment costs are given by

$$\begin{aligned} i_t^2 &= \left(\frac{1}{2} \kappa(1-\tau) \right)^2 + (\tau k_1)^2 (\delta(1-\rho))^{2t} \\ &\quad - \kappa(1-\tau) \tau k_1 (-\delta(1-\rho))^t, \end{aligned}$$

which oscillate.

10.2 Generalized proof of Proposition 6: a basis for the calibration

Consider equation (27) in Definition 5. Using the linear guesses $\tau_t = \mathcal{T}(k_t^o) = \alpha_{01} + \alpha_{11}k_t^o$ and $i_t = \mathcal{I}(k_t^o) = \alpha_{02} + \alpha_{12}k_t^o$, we can rewrite (27) as

$$i_{t-1} - \beta(1-\delta)(\alpha_{02} + \alpha_{12}k_{t+1}^o) = \frac{\kappa}{2}(1-\beta(1-\delta)) - \frac{\beta}{2}\tau_t - \frac{\beta^2}{2}\delta(1-\rho)(\alpha_{01} + \alpha_{11}k_{t+1}^o). \quad (66)$$

Next, using the law of motion of capital, (26), to eliminate k_{t+1}^o and solving (66) for $\beta\tau_t$ one arrives at

$$\beta\tau_t = Q_0 + Q_1 i_{t-1} + Q_2 k_t^o \equiv Q(i_{t-1}, k_t^o), \quad (67)$$

where

$$\begin{aligned} Q_0 &= \beta(1+\beta\delta(1-\rho)(1-\alpha_{01}) + 2\alpha_{02}(1-\delta)), \\ Q_1 &= -((1-\delta\rho)\beta(\beta\delta(1-\rho)\alpha_{11} - 2\alpha_{12}(1-\delta)) + 2), \\ Q_2 &= -(1-\delta)\beta(\beta\delta(1-\rho)\alpha_{11} - 2\alpha_{12}(1-\delta)). \end{aligned}$$

Next, we can use (67) to eliminate τ_t in the planner's recursive objective function, (28) to obtain

$$W(k_t^o) = \max_{i_{t-1}} \{(\beta + (A-1)Q(i_{t-1}, k_t^o))(k_t^o + i_{t-1}) - i_{t-1}^2 + \beta W(k_{t+1}^o)\}.$$

The first-order and envelope conditions for this problem are

$$(\beta + (A-1)Q(i_{t-1}, k_t^o)) + (A-1)Q_1(k_t^o + i_{t-1}) - 2i_{t-1} + \beta(1-\rho\delta)W'(k_{t+1}^o) = 0 \quad (68)$$

and

$$W'(k_t^o) = (\beta + (A-1)Q(i_{t-1}, k_t^o)) + (A-1)Q_2(k_t^o + i_{t-1}) + \beta(1-\delta)W'(k_{t+1}^o),$$

respectively. Using the first-order condition to eliminate $W'(k_{t+1}^o)$ in the envelope condition yields

$$\begin{aligned} W'(k_t^o) &= (\beta + (A-1)Q(i_{t-1}, k_t^o)) + (A-1)Q_2(k_t^o + i_{t-1}) \\ &\quad - \left(\frac{1-\delta}{1-\rho\delta}\right) ((\beta + (A-1)Q(i_{t-1}, k_t^o)) + (A-1)Q_1(k_t^o + i_{t-1}) - 2i_{t-1}). \end{aligned}$$

Leading this expression by one period allows us to eliminate $W'(k_{t+1}^o)$ in (68). Then, replacing $Q(i_{t-1}, k_t^o)$ by its expression allows us to rewrite (68) as

$$W_0 + W_1 i_{t-1} + W_2 k_t^o = Z_1 i_t + Z_2 k_{t+1}^o, \quad (69)$$

where

$$\begin{aligned} W_0 &= (\beta(1-\delta)(\beta + (A-1)Q_0) - \beta - (A-1)Q_0 - \beta(1-\rho\delta)(\beta + (A-1)Q_0)) \\ W_1 &= -(2(A-1)Q_1 - 2) \\ W_2 &= -(A-1)(Q_1 + Q_2) \\ Z_1 &= (-\beta(1-\delta)(2(A-1)Q_1 - 2) + \beta(1-\rho\delta)(A-1)Q_1 + \beta(1-\rho\delta)(A-1)Q_2) \\ Z_2 &= (-\beta(1-\delta)((A-1)Q_1 + (A-1)Q_2) + 2\beta(1-\rho\delta)(A-1)Q_2). \end{aligned}$$

Finally, we use (30) and (26) to eliminate i_t and k_{t+1}^o from (69) and rearrange terms to obtain

$$i_{t-1} = \frac{W_0 - Z_1 \alpha_{02} + (W_2 - (Z_1 \alpha_{12} + Z_2)(1-\delta))k_t^o}{(Z_1 \alpha_{12} + Z_2)(1-\rho\delta) - W_1},$$

which is a linear (affine) function of k_t^o , consistently with our guess (30). We can now equate coefficients to obtain

$$\alpha_{02} = -\frac{Z_1 \alpha_{02} - W_0}{(Z_1 \alpha_{12}(1-\delta\rho) - W_1 + Z_2(1-\delta\rho))} \quad (70)$$

$$\alpha_{12} = -\frac{(Z_2(1-\delta) + Z_1 \alpha_{12}(1-\delta) - W_2)}{(Z_1 \alpha_{12}(1-\delta\rho) - W_1 + Z_2(1-\delta\rho))}. \quad (71)$$

To verify our other guess, (29), we return to the condition (67). Using the guess (30) to eliminate i_{t-1} yields,

$$\beta\tau_t = Q_0 + Q_1 \alpha_{02} + (Q_1 \alpha_{12} + Q_2)k_t^o,$$

which is also linear (affine), in line with our guess. Equating coefficients leads to

$$\beta\alpha_{11} = Q_1 \alpha_{12} + Q_2, \quad (72)$$

$$\beta\alpha_{01} = Q_0 + Q_1 \alpha_{02}. \quad (73)$$

The equations (70)-(73) defines a system of non-linear equations in the four unknown coefficients $\alpha_{01}, \alpha_{11}, \alpha_{02}, \alpha_{12}$. In general, we must resort to numerical analysis to solve such system.

10.3 Comparative statics of the Markov equilibrium

The comparative statics of λ are established as follows. We note that

$$\frac{d\lambda}{dA} = \frac{1}{2} \frac{1 - 2A(1 - \beta(1 - \rho)^2) + \sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)}}{\beta(1 - \rho)A^2\sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)}}$$

with

$$\left[\frac{d\lambda}{dA} \right]_{A=1} = (1 - \rho) > 0.$$

Furthermore, the derivative of the numerator is given by

$$\begin{aligned} & 2(1 - \beta(1 - \rho)^2) \frac{2(A - 1) - \left(\sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)} - 1 \right)}{\sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)}} \\ & > 2(1 - \beta(1 - \rho)^2) \frac{2(A - 1) - \left(\sqrt{1 + 4A(A - 1)} - 1 \right)}{\sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)}} \\ & = 0, \end{aligned}$$

so λ is increasing in A everywhere, positive and bounded from above by

$$\begin{aligned} \lim_{A \rightarrow \infty} \lambda &= \frac{1 - \sqrt{1 - \beta(1 - \rho)^2}}{(1 - \rho)\beta} \\ &< \frac{1 - (1 - \beta(1 - \rho)^2)}{(1 - \rho)\beta} \\ &= (1 - \rho). \end{aligned}$$

Furthermore,

$$\begin{aligned} \frac{d\lambda}{d\rho} &= - \left(\frac{(2A - 1)}{2} \frac{(2A - 1) - \sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)}}{\beta(1 - \rho)^2 A \sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)}} \right) \\ &< \left(\frac{(2A - 1)}{2} \frac{(2A - 1) - \sqrt{1 + 4A(A - 1)}}{\beta(1 - \rho)^2 A \sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)}} \right) = 0, \end{aligned}$$

with

$$\begin{aligned} \lim_{\rho \rightarrow 1} \lambda &= \lim_{\rho \rightarrow 1} \left(\frac{1}{\beta(1 - \rho)} \left(\frac{A - 1}{A} - \frac{\sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)} - 1}{2A} \right) \right) \\ &= \lim_{\rho \rightarrow 1} \left(\frac{4A(A - 1)(1 - \rho)}{\sqrt{1 + 4A(A - 1)(1 - \beta(1 - \rho)^2)}} \right) = 0, \end{aligned}$$

where we used l'Hôpital's rule in the last equation. Furthermore, τ_1 is given by

$$\begin{aligned}\tau_1 &= \alpha_{01} + \alpha_{11}k_1^o \\ &= \frac{\sqrt{1 + 4A(A-1)(1 - \beta(1 - \rho)^2)} - 1}{2A - 1 + \sqrt{1 + 4A(A-1)(1 - \beta(1 - \rho)^2)}} \\ &\quad + \frac{2\left(2A - 1 - \sqrt{1 + 4A(A-1)(1 - \beta(1 - \rho)^2)}\right)}{\beta^2(1 - \rho)^2\left(1 + \sqrt{1 + 4A(A-1)(1 - \beta(1 - \rho)^2)}\right)}k_1^o,\end{aligned}$$

where $\alpha_{01}, \alpha_{11} > 0$.

10.4 The stochastic case with government debt

In the case in which the government has no access to capital markets (section 6.2), the state-contingent tax plan in the first and second period prescribes given by the solution of the following linear system:

$$\begin{aligned}\tau_1 &= (1 + \beta(1 - \rho))\frac{A_l + p(A_h - A_l) - 1}{2A_1^e - 1} - \beta(1 - \rho)p\frac{A_l + A_h + p(A_h - A_l) - 1}{2A_1^e - 1}\tau_{h,2} \\ &\quad - \beta(1 - \rho)(1 - p)\frac{2A_l + p(A_h - A_l) - 1}{2A_1^e - 1}\tau_{l,2} \\ \tau_{l,2} &= \frac{(A_l - 1)\rho - (1 - \rho)(2A_l + p(A_h - A_l) - 1)\tau_1 - \beta(1 - \rho)^2 p(A_h + A_l - 1)\tau_{h,2}}{\left(1 - \beta(1 - \rho)^2 p\right)(2A_l - 1)} \\ \tau_{h,2} &= \frac{(A_h - 1)\rho - (1 - \rho)(A_h + A_l + p(A_h - A_l) - 1)\tau_1 - \beta(1 - \rho)^2(1 - p)(A_h + A_l - 1)\tau_{l,2}}{\left(1 - \beta(1 - \rho)^2(1 - p)\right)(2A_h - 1)}.\end{aligned}$$

10.5 Marginal distortionary costs and revenues balancing

Our results imply that taxes for periods $t > 1$ are independent of the amount of initial inelastic capital. To understand this result, it is helpful to consider the dual of the problem analyzed above, i.e., considering directly the optimality conditions for the sequence of taxes. To illustrate that it is immaterial for the argument whether we use the public good interpretation or the exogenous financing requirement interpretation we now use the latter.

The objective of the planner is to solve

$$\max_{\{\tau_t\}_0^\infty} U(\{\tau_t\}_0^\infty) + \Psi(R(\{\tau_t\}_0^\infty) - G)$$

where U denotes intertemporal private welfare after substituting the incentive compatibility constraint $i_t = \frac{\kappa - \tau_t}{2}$ and R and G are the present discounted values of tax revenues and expenditures of the government. Defining in current terms the marginal revenue of taxes and the marginal utility loss of taxes,

$$R_s \equiv \beta^{-s} \frac{\partial R}{\partial \tau_s}, M_s \equiv \beta^{-s} \frac{\partial U}{\partial \tau_s},$$

the first order condition for τ_s is given by

$$M_s + \Psi R_s = 0. \tag{74}$$

Let us then define the wedge between the marginal utility loss and marginal revenue of τ_s as the *intertemporal* marginal distortion measured in current terms, denoted $D_s \equiv -M_s - R_s$, and rewrite (74)

$$D_s = (\Psi - 1)R_s. \tag{75}$$

This is the standard condition that the marginal intertemporal distortion for each tax should be proportional to its marginal revenue and the factor of proportionality is the excess marginal value of public funds.

Now, we should note that since τ_t affects investments at all periods before t , the marginal *intertemporal* distortions D_t accumulate over time. In particular, if investments in periods before t are heavily distorted, the marginal intertemporal distortion of τ_t is high and vice versa. More specifically, recalling that $\frac{\kappa}{2} - i_t$ is the wedge between first best and actual investment levels at t , straightforward calculus yields that

$$\begin{aligned} D_s &= (1 - \delta) D_{s-1} + \delta (1 - \rho) \left(\frac{\kappa}{2} - i_{s-2} \right) + \frac{\kappa}{2} - i_{s-1} \\ D_1 &= \frac{\kappa}{2} - i_0. \end{aligned} \tag{76}$$

Similarly, since the tax base of τ_s is all the investments done before s that remains at s , also the marginal revenue R_s accumulates over time. Specifically,

$$\begin{aligned} R_1 &= k_1^o + 2i_0 - \frac{\kappa}{2} \\ R_s &= (1 - \delta) R_{s-1} + \delta (1 - \rho) \left(2i_{s-2} - \frac{\kappa}{2} \right) + 2i_{s-1} - \frac{\kappa}{2} \forall s > 1. \end{aligned} \tag{77}$$

Substituting from (76) and (77) into (75) yields the first-order condition for τ_s for all $s \geq 2$, i.e.,

$$\begin{aligned} &(1 - \delta) D_{s-1} + \delta (1 - \rho) \left(\frac{\kappa}{2} - i_{s-2} \right) + \frac{\kappa}{2} - i_{s-1} \\ &= (\Psi - 1) (1 - \delta) R_{s-1} + (\Psi - 1) \delta (1 - \rho) \left(2i_{s-2} - \frac{\kappa}{2} \right) + (\Psi - 1) \left(2i_{s-1} - \frac{\kappa}{2} \right). \end{aligned} \tag{78}$$

Now let us perform comparative statics on k_1^o . First, we note that the first-order condition for τ_1 , i.e., $D_1 = (\Psi - 1) R_s$ yields

$$\frac{\kappa}{2} - i_0 = (\Psi - 1) \left(k_1^o + 2i_0 - \frac{\kappa}{2} \right).$$

Suppose k_1^o goes up (from zero or any number consistent with an interior choice of τ_1). The marginal revenue of τ_1 (RHS) then increases, calling for an increase in the intertemporal distortion (LHS), and thus i_0 must fall and τ_1 should be increased.

Next, turn to the first-order condition for τ_2 , expressed as in (78) for $s = 2$. First, we note that the first term of the marginal revenue of τ_2 , (the RHS of (78) has increased since R_1 has gone up. However, this effect is exactly balanced by an increase in the first term of the intertemporal marginal cost of τ_2 (the RHS of (78) since $D_1 = (\Psi - 1) R_1$ by the first-order condition for τ_1 . Under geometric depreciation, the second terms on both sides of (78) vanish and what remains is simply to set $\frac{\kappa}{2} - i_{s-1} = (\Psi - 1) \left(2i_{s-1} - \frac{\kappa}{2} \right)$ by choosing $\tau_2 = \frac{\Psi - 1}{2\Psi - 1}$. The same is true for all $s \geq 2$, implying that under geometric depreciation, τ_s is independent of k_1^o given Ψ .

When $\rho \neq 1$, the second terms on both sides of (78) do not vanish. In particular, if $\rho < 1$, the fact that i_0 has decreased *increases* the intertemporal marginal distortion cost of τ_2 *more* than what is already captured in the increase in $(1 - \delta) R_1$ (the second term falls since i_0 has decreased). In addition, the marginal revenue of τ_2 increases *less* than what is captured by the increase in $(1 - \delta) R_{s-1}$, since the second term of the RHS has decreased. This is straightforward to understand: relative to the geometric case, a relatively smaller part of k_1^o remains in period 2 and the investments i_0 are relatively more sensitive to τ_2 . Thus, the increase in τ_1 has increased the marginal cost of τ_2 more and the marginal revenue less than under geometric depreciation (in which case they would have increased by the same amount). Therefore, i_1 must increase by a decrease in τ_2 , which reduces its intertemporal marginal distortionary cost (LHS) and increase its marginal revenue (RHS). The first-order condition for τ_3 is then also out of balance. The marginal distortionary cost is smaller than the marginal revenue and τ_3 should increase, implying oscillating dynamics in taxes and investments.

Finally, when $\rho > 1$, and increase in k_1^o again increases the marginal revenue of τ_1 and calls for it to increase and i_0 to fall. This, however, leads to a *smaller* increase in the marginal cost of τ_2 than what is captured by $(1 - \delta) D_1$ and a *higher* increase in the marginal revenue than under geometric depreciation. Therefore, also τ_2 should increase. Again, the first-order condition for τ_3 is then also out of balance. However, the marginal revenue of τ_3 is higher than its marginal distortionary cost, so τ_3 should also increase.