The Optimal Nonlinear Taxation of Capital in Models With Uninsurable Income Risk*

Michael Reiter, Universitat Pompeu Fabra, Barcelona

This version, August 2004

JEL Classification: H21, E62
Keywords: Optimal taxation, capital income taxation, bequest taxation, redistribution

Abstract

The paper studies the optimal taxation of capital or bequests in a model where households face uninsurable income shocks and a liquidity constraint. The government maximizes a Utilitarian welfare function, and can perfectly commit to future tax rates. The paper considers a fully nonlinear tax schedule and provides an analysis of the transition phase as well as the steady state.

A dynamic analogue to the well known "no distortion at the top" result from the static optimal nonlinear income taxation model is proven. Numerical examples show that the optimal tax rate can be large even in the long run if households are more risk averse than in the log-utility specification. The shape of the marginal tax function depends mainly on the upper tail of the income distribution.

Address of the author:

Michael Reiter
Department of Economics
Universitat Pompeu Fabra
Ramon Trias Fargas 25-27
E-08005 Barcelona
e-mail: michael.reiter@upf.edu

*I am grateful to Andres Erosa, Luisa Fuster, Albert Marcet, Leslie Reinhorn, Amedeu Spadaro and participants of the CEPR Conference on Dynamic Aspects of Taxation for helpful comments and discussions. Support from the Spanish Ministry of Education under the grant DGES PB98-1066 is gratefully acknowledged.

1 Introduction

Over the last years, the view that income on capital should not be taxed, because it distorts the saving decision and results in sizeable welfare losses, gains more and more ground among economists (Lucas, 1990, p.314 states the point forcefully by saying that eliminating capital taxation is "... the largest genuinely free lunch I have seen in 25 years in this business,..."). In contrast, actual economies continue to rely heavily on taxes on capital income and on bequests (cf. Krusell, Quadrini and Rios-Rull, 1996, Table 1). Recent efforts in the US to eliminate estate taxation are so far an exception. My interpretation of this fact is that the majority of politicians and the wider public think that some kind of capital income taxation is desirable on distributional grounds.

The famous results from the early literature on optimal dynamic taxation, showing that the optimal tax on capital is zero in the long run, seem to suggest that the public opinion on the distributional effects of capital taxation are misguided, or valid only in the short run. Judd (1985) shows that the result holds even in a heterogenous agents model, where the government cares only about workers, not capitalists. Chamley (1986), too, points out that the zero tax result carries over to the case of heterogenous agents.

A more recent literature has shown that certain types of market incompleteness may lead to different conclusions. Both Aiyagari (1995) and Chamley (2001) investigate a model where households face borrowing constraints and uninsurable idiosyncratic risk, and find reasons for a tax on capital income even in the long run.[1] Aiyagari's argument is based on the overaccumulation of capital by liquidity constrained households, but Chamley (2001) makes clear that what really matters is the possible role of capital taxation as a substitute for the missing insurance markets, if consumption is positively correlated with wealth. What is insurance ex ante is redistribution ex post, so this literature clearly attributes some role to the taxation of capital for distributional reasons. Calibrated models studying non-optimal tax reforms in incomplete market economies have come to similar conclusions. Both Garcia-Mila, Marcet and Ventura (1995) and Domeij and Heathcote (2000) find that abolishing capital taxes quickly makes a large part of the population worse off, despite an increase in macroeconomic activity.

The present paper analyzes a model in the Aiyagari/Chamley framework. It is now well understood that this framework calls for some (perhaps very small) positive taxation of capital in the long run. The present paper goes beyond this literature in several respects. It gives both a qualitative and quantitative analysis of a fully nonlinear capital tax struc-

[1] There are reasons other than distributional concerns why the optimal tax on capital might fail to be zero in steady state, for example the existence of untaxable production factors (Correia 1996), or the incomplete taxation of pure rents (Jones, Manuelli and Rossi 1997).

ture. Furthermore, it deals not just with steady state taxation, but considers the optimal time-varying tax structure under full commitment, including both a short-run and long-run analysis. In some numerical examples, the quantitative welfare effects of optimal capital taxation in this setting are explored. It is shown how the nonlinear tax structure depends on the characteristics of the income distribution, in particular on the shape of the upper tail.

In the recent literature, one can find different approaches to nonlinear optimal dynamic taxation. Costa and Werning (2001) and Kapicka (2002) study optimal nonlinear income taxation in setups that give rise to a stationary optimal tax structure. Conesa and Krueger (2002) determine the optimal nonlinear income tax in a rich model calibrated to the US economy. The tax structure is invariant over time, and comes from a three-parameter family. Saez (2002) deals with nonlinear capital taxation; the paper allows for time-varying tax rates, but uses more restrictive forms of nonlinear taxes. The model is deterministic, and the focus is on the redistribution of previously accumulated wealth, not on the insurance aspect of capital taxation that is a main ingredient of the present paper.

The strand of literature that is closest to this paper in the sense that it studies time-varying nonlinear tax schedules is the so-called "Mirrlees approach" to dynamic taxation (Golosov, Kocherlakota and Tsyvinski, 2003,Kocherlakota, 2004, Albanesi and Sleet, 2004). These papers solve for the first best allocation under an informational constraint (for example, that the government cannot observe individual abilities), and afterwards ask what is the tax structure that implements this optimal allocation. In contrast, the present paper follows the "Ramsey approach", looking for a second best under constraints on the tax system that do not necessarily stem from information constraints. The Mirrlees approach can more easily handle complex tax structures (the above mentioned papers deal with optimal nonlinear income and capital tax simultaneously, while I am looking at capital taxation only), since it is easier to solve for a constrained optimum rather than solving the first order conditions of a second best problem. On the other hand, that approach *forces* us to consider a tax structure that is general enough to implement the constrained optimum. This means, for example, that one cannot ask how much we would gain by going from the optimal linear to the optimal nonlinear tax, which is one of the issues addressed in the present paper.

The plan of the paper is as follows. Section 2 presents the model. Section 3 discusses the technical problems that have to be solved to find a solution to the optimal tax problem, and states the solution concept that will be used. Theoretical results on the optimal tax structure are presented in Section 4, while Section 5 provides some numerical illustrations. Section 6 concludes. The appendices contain the derivation of the analytical results, and a brief description of the numerical methods.

2 The model

2.1 Overview of the model

The model analyzed here is similar to Aiyagari (1995). I have simplified the model as much as possible, to make the optimal nonlinear tax problem tractable. In particular, I assume that labor supply is exogenous and that the economy works with a linear technology in labor and capital.

A key feature of the model is that households (dynasties) face each period an idiosyncratic labor income shock, which is the only source of uncertainty in the model. Asset markets are incomplete: the income shock cannot be insured, and households face a liquidity constraint in the form that their assets cannot become negative. In this environment, a Utilitarian government imposes taxes on capital, and redistributes the revenues lump sum. There is no public consumption, and capital taxes only serve for redistribution. In period 0, the government can commit to a sequence of time varying and nonlinear tax schedules on capital. The tax function is nonlinear as in Mirrlees (1971). The assumption of full commitment is not made for realism, but in order to give a normative analysis. We know that without government commitment, capital taxes can be very high in a political equilibrium (Klein and Ríos-Rull 2001).

Besides capital taxes, no other taxes appear in this model. Since labor supply is inelastic, labor income taxes would be be non-distortionary, and we can therefore interpret labor income in this model as after-tax income.

Let me now discuss some of the more important modeling choices.

2.1.1 Credit constraints vs. bequest constraints

The formal model of this paper can be interpreted either as a model of infinite lives with borrowing constraints, or a model of altruistic non-overlapping generations with a non-negativity constraint on bequests. The qualitative results of Section 4 are equally valid for both cases, but for the quantitative exploration in Section 5 we have to decide on an interpretation. For the present purpose, I find the latter one more attractive, for several reasons.

First, studying nonlinear tax schedules seems especially important for bequest taxes. In many countries, marginal taxes on bequests are more graduated than other taxes. In the US, for example, marginal estate tax rates in 2000 vary between 18 and 60 percent (for transfers between parents and children, more for other transfers), while the top marginal federal income tax rate was less than 40 percent, and the top rate on capital gains was about 20 percent. This may reflect the particular role of bequests for equity, since unequal bequests mean that people face different starting positions in life.

The second reason is theoretical. Borrowing constraints arise endogenously in an economy, for example through asymmetric information and limited enforceability of contracts. Modeling this constraint as exogenous is not innocuous and may distort the conclusions of the analysis (Krueger and Perri, 1999; cf. also Yotsuzuka, 1987 on the question of Ricardian Equivalence under borrowing constraints). In contrast, the non-negativity of bequests is a legal restriction that can be considered exogenous. Leaving bequests is a unilateral act, and while it is possible to make gifts, one cannot simply take money away from other people, not even from one's own children.

The third reason is technical. Interpreting the model period as a generation, the numerical examples can be computed based on a relatively small number of periods. The numerical solution technique that I employ in Section 5 would not be feasible, given my computer resources, if the model period were one year.

Of course it is true that the simple model analyzed here omits many of the specific characteristics of the bequest decision, for example that bequests have to be split among several children, or that parents already have a lot of information about the income realization of their children at the time when they leave the bequest (both aspects are analyzed in Cremer and Pestieau, 2001). But a similar critique could be made about the interpretation of the model as capital accumulation over life, since the model omits life cycle features, which have been shown to have important implications for the optimal taxation of capital (Erosa and Gervais 2002).

A critical feature of this analysis is that bequests are modeled as purely altruistic. Different assumptions on bequest motives (accidental bequests, joy of giving, strategic bequests) may have very different policy implications. Nevertheless, altruistic bequests have received so much attention in the literature that they deserve a separate analysis.

2.1.2 Linear production function

Like Chamley (2001), Saez (2002) and much of the public finance literature on optimal taxation, I assume a linear production function, so that the wage and the before tax interest rate are constant. This assumption simplifies the analysis and is not crucial for the qualitative results. Quantitatively, however, it may bias the results in the direction of a higher tax rate on capital, since we omit the negative effect on wages (and on the income distribution) from the reduction in the capital stock that is induced by the capital tax.

To guarantee a steady state in the absence of taxation, we have to make the following assumption on the before tax interest rate r and the discount factor β (Aiyagari 1995):

Assumption 1. *The interest rate is smaller than the rate of time preference:*

$$\beta(1+r) < 1 \tag{1}$$

2.1.3 Public debt

The model does not allow government debt, for three reasons. First, under Assumption 1, there is no steady state if the government can borrow money freely. In a steady state, the current value Lagrange multiplier ν of the budget constraint is constant. Next periods revenues are therefore valued today at $\beta\nu$. Since the government borrows at the before tax interest rate r, Equ. (1) implies that the government always wants to borrow more money. A steady state is only possible if the government is borrowing constrained, just as the individuals.

Second, the focus of the paper is on the distributional role of capital taxation, not on the role of government debt in a world where people are liquidity constrained. If we allow public debt, the two effects get mixed up (see also the introduction of Chamley (2001) on this issue).

The third reason is empirical relevance. Especially from a long-run point of view, where government debt redistributes between generations, the recent political opposition to unfunded pension systems and public debt in general indicates that there are limits to what the government can do. It is convenient to set this limit to zero.

2.2 Households

I now turn to a formal description of the model. For the reasons explained above, I adopt the interpretation of the model as a sequence of non-overlapping generations, where capital is accumulated through bequests. The economy is then populated by a continuum of dynasties with constant mass 1. Each generation lives for only one period, but is altruistically linked to future generations. A member of generation t splits its resources into consumption C_t and bequests B_t so as to maximize (in this section we write all variables without household indices)

$$\mathrm{E}_t \sum_{s=t}^{\infty} \beta^{s-t} U(C_s) \tag{2a}$$

subject to a budget constraint and a non-negativity constraint on bequests:

$$B_t = (1+r)\left(B_{t-1} + z_t - C_t - T_t\left(B_t\right)\right) \tag{2b}$$

$$B_t \geq 0 \tag{2c}$$

The household has inherited the amount B_{t-1} from the last generation, and is endowed with one unit of labor, which it supplies inelastically. While there is no aggregate uncertainty in this model, individual real labor productivity (or labor income) z_t is stochastic. The household has to pay a tax $T_t(B_t)$ on the bequests it leaves. For convenience, taxes are written here as a function of after-tax bequests B_t, not of before-tax savings, which are given by $S_t = B_{t-1} + z_t - C_t$. In the optimum, it will turn out that $T_t(0)$ is negative, which can then be interpreted as a lump sum transfer. Implicit in (2b) is the assumption that labor income fluctuations cannot be insured.

In the following, we will abbreviate the available private resources of a household of generation t (its capital) by

$$k_t \equiv B_{t-1} + z_t \tag{3}$$

Capital k_t is the only individual state variable. The optimal bequest and consumption function of the household with capital k are denoted by $B_t[k]$ and $C_t[k]$, respectively. They are related by the budget constraint (2b). For notational clarity, function arguments in brackets always refer to the household with capital k.

We make the following assumptions on labor productivity z_t:

Assumption 2. *i) The distribution of z_t is identical and independent over time and across households;*

ii) it has bounded support $(\underline{z}, \overline{z})$ with $\underline{z} > 0$;

iii) it has a density $\pi(z)$ which is continuously differentiable on $\Re$.

iv) The density $\pi(z)$ is weakly decreasing in a neighborhood of $(\overline{z})$.

Note that Part iv) of the assumption is quite natural, given that $\pi(\overline{z}) = 0$ by continuity. It is made to exclude certain types of oscillatory behavior at the upper end of the support, which is used in the proof of Proposition 2.

We make standard assumptions on the utility function:

Assumption 3. *i) The utility function $U(C)$ is strictly increasing, strictly concave and twice continuously differentiable.*

ii) $\lim_{c\to 0} U'(c) = \infty$

For technical convenience, we assume that the initial capital distribution is bounded (later assumptions will guarantee that it stays bounded):

Assumption 4. *i) The initial cross-sectional distribution of capital is given and has bounded support, with a supremum value $\bar{K}_0$.*

ii) The distribution is continuous with a density denoted by $\phi_0[k]$.

From our assumptions on the income distribution, it follows immediately that the cross-sectional distributions for periods $t = 1, 2, \ldots$ are continuous. We denote their density by $\phi_t[k]$.

2.3 The problem of the government

The government maximizes a Utilitarian welfare function, that means, an unweighted sum of individual utilities. The only fiscal instrument available to the government is a tax on capital (bequests) $T_t(B)$. We assume $T_t(B)$ to be continuous and piecewise smooth, in the sense of being twice continuously differentiable everywhere except possibly at a countable number of values of b. We can represent the tax function then by a lump sum subsidy, $-T_t(0)$, and a marginal tax function $T'_t(B)$. Tax rates can change over time, and in period 0 the government can fully commit to a fiscal policy $\Theta = T_t(0), T_1(0), \ldots; T'_0(B), T'_1(B), \ldots$.

The government faces the period-by-period budget constraint

$$0 = \int_{\Re} \phi_t[k] \, T_t(B_t[k]) \, \mathbf{d}k \tag{4}$$

Two aspects of this formulation deserve notice. First, the revenues of the tax on bequests are redistributed lump sum within the generation that pays the tax; the government does not redistribute between generations. This is important, since Ricardian Equivalence does not hold in this model, and it might contribute to the strong negative effect of taxation on capital formation that we will observe later. Second, the government cannot levy a tax on initial capital (which would be a lump sum tax). The first tax it can impose is on the bequests of the first generation, which is already distortionary. It is therefore not necessary to place binding upper limits on the tax in the first period, unlike in many other models of capital taxation (cf. for example Chamley, 1986).

To make sure that the value function of the household is bounded (cf. Lemma 2), it is convenient to impose some very loose upper and lower bounds on taxes:

Condition 1. *There exist constants $\underline{\tau}$ and $\bar{\tau}$ such that for all t,*

$$T_t(0) \leq \bar{\tau} < \underline{z} \tag{5a}$$

$$T_t(B_t[k]) \geq -\underline{\tau} \int_{\Re} k \phi_t[k] \, \mathbf{d}k \tag{5b}$$

These constraints should not be binding (as will be confirmed in the numerical simulations), but I do not find a way to proof this. The upper bound guarantees a minimum consumption level for all households. This constraint is innocuous, since we expect $T_t(0)$ to be negative. The lower bound on taxes (upper bound on subsidies) prevents the government from creating an unbounded cross-sectional wealth distribution through unbounded subsidies to the richest people. Since we have a continuum of households, the government could theoretically do this, although such a policy seems to make no sense for a Utilitarian government under concave utility.

Definition 1. By $\bar{K}_t$ we denote the supremum of the support of the cross-sectional distribution $\phi_t[k]$.

By $\bar{\bar{K}}_t$ we denote the maximum amount of capital that a household can accumulate at t, i.e., the k_t that is achieved by starting from $\bar{K}_0$, always drawing the highest possible income shock $\overline{z}$, and consuming zero in every period.

Clearly, both $\bar{K}_t$ and $\bar{\bar{K}}_t$ are contingent on tax policy.

Based on Condition 1, we can prove (cf. Appendix A) the following

Lemma 1. *a) For each t, the cross-sectional distribution has bounded support.*

b) For any $\hat{\beta} < \frac{1}{1+r}$,

$$\lim_{t\to\infty} \hat{\beta}^t \bar{\bar{K}}_t = 0 \tag{6}$$

We are now ready to state the problem of the government:

Program P1: for $t = 0, \ldots, \infty$, choose transfers $T_t(0)$, functions of marginal tax rates $T'_t(B)$, cross-sectional distributions $\phi_t[k]$ and bequest functions $B_t[k]$ so as to maximize

$$\sum_{t=0}^{\infty} \beta^t \int_{\Re} \phi_t[k]\, U\left(k - \frac{B_t[k]}{(1+r)} - T_t(B_t[k])\right) \mathbf{d}k \tag{7}$$

subject to the following constraints:

i) the budget constraint (4) and Condition 1.

ii) the cross-sectional distributions of wealth holdings satisfy the dynamic equation

$$\phi_{t+1}[k] = \int_{\Re} \phi_t[j]\, \pi(k - B_t[j])\, \mathbf{d}j, \qquad \forall t \tag{8}$$

with given initial distribution $\phi_0[k]$.

iii) The marginal tax functions are piecewise smooth in the sense defined above.

iv) The bequest functions $B_t[k]$ represent the utility maximizing choices of households, given the households' initial level of capital, their history of shocks and the sequence of policy parameters. (More precisely, the function $B_t[k]$ can be any selection of the optimal bequest correspondence of the households.)

3 Solving the optimal tax problem

The aim of this section is to treat the technical issues in the solution of the problem of the government, and to describe the solution concept. The proofs of all lemmas can be found in Appendix A.

3.1 Solution of the household problem

Let us define Θ_t as the announced policy from time period t onwards. Then the solution to the household (sub)problem starting at time t can be characterized by the usual value function $v_t(k; \Theta_t)$, which is a function of the household's capital stock at time t and is conditional on Θ_t. For notational convenience, the dependence on Θ_t will usually be suppressed, so we write $v_t(k)$. We think of $v_t(k)$ as defined on $k \in (\underline{z}, \bar{\bar{K}}_t)$, which contains the support of $\phi_t[k]$.

Lemma 2. *i) For all t and all $k \in (\underline{z}, \bar{\bar{K}}_t)$, the value function $v_t(k)$ is bounded and strictly increasing in k.*

ii)

$$\lim_{t \to \infty} \beta^t v(\bar{\bar{K}}_t) = 0 \tag{9}$$

iii) For each t, the value function is differentiable almost everywhere and can be represented as

$$v_t(k) = v_t(\underline{z}) + \int_{\underline{z}}^{k} v_t'(j) \, \mathbf{d}j \tag{10}$$

The value function satisfies the Bellman equation

$$v_t(k; \Theta_t) = \max_b \{U_t(k, b; \Theta_t) + \beta \, \mathrm{E}_t \, v_{t+1}(b + z; \Theta_{t+1})\} \tag{11}$$

The optimal choices of b that solve the rhs of (11) are collected in the bequest correspondence $B_t^*[k]$. Using the value function we can now establish the following lemma:

Lemma 3. *For any sequence of tax policies that allows a solution of the household problem, the optimal bequest correspondence $B_t^*[k]$ is non-decreasing in k for all t, that means,*

$$k_0 < k_1, \quad b_0 \in B_t^*[k_0], \quad b_1 \in B_t^*[k_1] \qquad \Rightarrow \qquad b_0 \leq b_1 \tag{12}$$

Note that this monotonicity implies that $B_t^*[k]$ is single-valued almost everywhere.

3.2 Rewriting the government problem

In order to make the government problem **P1** tractable, it will be necessary to work with the first order conditions of the household problem, as is commonly done in optimal taxation problems. With nonlinear tax functions, we face the problem that the budget set of the household is not concave if marginal tax rates are decreasing in some regions, which will be the case in most of the applications below. Then the first order conditions do not guarantee that the household is at the optimal point. This is a serious problem. We know from Mirrlees (1986) that in this case not even the derivation of the necessary first order conditions of the tax problem is valid.

To solve this problem, we can first think of the government as choosing a sequence of household value functions and bequest functions which have the properties established in Lemma 2, and together satisfy the Bellman equation (11). We know that this is sufficient for a household optimum. In a second step, we restrict the government policy in a way that the household problem is "not too non-concave", such that the first order conditions of the household are sufficient for an optimum. Exploiting a single-crossing property of the household indifference curves, it is possible to show that this restriction is innocuous in the sense that it does not affect the real choices available to the government. Following this line of argument, we can show (cf. Lemma 4) that problem **P1** can be replaced by the following, more tractable problem:

Program P2: For $t = 0, \ldots, \infty$, choose piecewise smooth tax functions $T(b) : (0, \infty) \to \Re$, u.h.c. bequest correspondences $B_t^* [k] : (\underline{z}, \bar{\bar{K}}_t) \to (0, \infty)$ bequest functions $B_t [k]$ with $B_t [k] \in B_t^* [k]$, household value functions $v_t(k) : (\underline{z}, \bar{\bar{K}}_t) \to \Re$ with $\lim_{t\to\infty} \beta^t \bar{\bar{K}}_t = 0$ and cross-sectional distribution functions $\phi_t [k]$ so as to maximize (7) subject to

i)–iii) of Program P1

iv) the household first order conditions

$$U'\left(k - \frac{b}{1+r} - T(b)\right) \geq \beta \frac{1+r}{1+(1+r)T'_{t+}(b)} \int_{\Re} v'_{t+1}(b+z)\pi(z)\mathbf{d}z \tag{13a}$$

$$U'\left(k - \frac{b}{1+r} - T(b)\right) \leq \beta \frac{1+r}{1+(1+r)T'_{t-}(b)} \int_{\Re} v'_{t+1}(b+z)\pi(z)\mathbf{d}z \tag{13b}$$

hold for all $b \in B_t^* [k]$. Furthermore,

$$v'_t(k) = U'\left(k - \frac{B_t [k]}{1+r} - T(B_t [k])\right), \qquad \text{a.e.} \tag{14}$$

$$v_t(\underline{z}) = U\left(\underline{z} - \frac{B_t [\underline{z}]}{1+r} - T(B_t [\underline{z}])\right) + \beta \int_{\Re} v_{t+1}(B_t [\underline{z}] + z)\pi(z)\mathbf{d}z \tag{15}$$

Equs. (13) take into account that the tax function is not necessarily differentiable, but has a left and right derivative, $T'_{t+}(b)$ and $T'_{t-}(b)$, respectively. The envelope condition (14) is standard. Equ. (15) is the HJB equation at the minimal capital level $\underline{z}$, and serves to make sure that the government really chooses the value function, not just its derivative. The expression $v_t(k)$ should always be understood as an abbreviation for the rhs of (10).

The following lemma is proven in Appendix A.1.

Lemma 4. *To any solution of program* **P1***, there is a solution of program* **P2** *(and vice versa) that is equivalent in the sense that it induces the same bequest behavior for almost all households, raises the same tax revenue and yields the same value of the government objective function.*

The term “almost all” means that the set of households for which this is not true is of measure zero w.r.t. to the initial distribution of capital.

3.3 A Lagrangian approach to the government problem

Both the theoretical analysis of Section 4 and the numerical results in Section 5 will focus on the following class of solutions to **P2**:

Definition 2. A *smooth interior solution* of Program *P2* is a solution where, for all t, the household bequest correspondence is single-valued on the support of $\phi_t\,[k]$, which can then be written as a function $B_t\,[k]$, and where this bequest function is continuously differentiable with $B_t'\,[k] > 0$ for all k with $B_t\,[k] > 0$.

In particular, these solutions exclude two types of “anomalous” behavior:

1. tax functions with kinks; this would lead to a bequest function where a range of households with different capital stocks are bunched at the same (positive) level of bequests;
2. discontinuous bequest functions.

One could allow for these cases in the derivation of the optimal policy, but it would further complicate the already lengthy derivations, and it turns out that the additional generality is not needed. For the numerical analysis, the restriction to smooth interior solutions is innocuous, since they can approximate the more general solutions arbitrarily well. Kinks or discontinuities should be detected in the process of the numerical solution, since they would generate bequest functions with a derivative that tends either to zero or to infinity. In all the numerical experiments undertaken so far, this has never been the case. This is also what one would expect a priori. Bunching should not be optimal (except at a zero level of bequests, where the constraints (2c) kicks in), because the government should be able to increase welfare by making some of the bunched households with higher capital stock consume less (bequest more), and some with lower capital stock, which have a higher marginal utility of consumption, consume more. In the case of discontinuities, households which differ in their capital stock only by an ϵ, differ in their consumption by a discrete amount. Here again, averaging the consumption level between the two types of households should increase aggregate welfare. Unfortunately, I cannot prove this formally.[2]

The analysis will be based on the necessary conditions for a smooth interior solution, which will be derived using the following Lagrangian

$$
\begin{aligned}
\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Bigg\{ \int_{\Re} \phi_t[k] \Bigg[& U(C_t[k]) + \xi_t[k] B_t[k] + \zeta_t[k] \left(-v_t'(k) + U'(C_t[k])\right) \\
& + \mu_t[k] \left(-U'(C_t[k]) + \beta \bar{R}_t[k] \int_{\Re} v_{t+1}(B_t[k] + z)\, \pi(z) \mathbf{d}z \right) \\
& + \lambda_t[k] \left(-\phi_t[k] + \int_{\Re} \phi_{t-1}[j]\, \pi(k - B_{t-1}[j])\, \mathbf{d}j \right) \Bigg] \mathbf{d}k \\
& + \nu_t \left(T_t(0) + \int_{\Re} \phi_t[k] \int_0^k T_t'(B_t[j])\, B_t'[j]\, \mathbf{d}j\, \mathbf{d}k \right) \\
& + \chi_t \left(-v_t(\underline{z}) + U(C_t[\underline{z}]) + \beta \int_{\Re} v_{t+1}(B_t[\underline{z}] + z) \pi(z) \mathbf{d}z \right) \Bigg\}
\end{aligned} \tag{16}
$$

In (16),

- The variables $\xi_t[k]$, $\zeta_t[k]$, $\mu_t[k]$, $\lambda_t[k]$, ν_t and χ_t are Lagrange multipliers.
- The after tax interest factor $\bar{R}_t[k]$ as is an abbreviation for
$$\bar{R}_t[k] = \frac{1+r}{1+(1+r)T_t'(B_t[k])} \tag{17}$$
- Total taxes have been rewritten in terms of marginal taxes. Since the taxes are a differentiable function of B, and B is differentiable in k by the definition of a smooth interior solution, the taxes to be paid by a household with capital k are given by
$$T_t(B_t[k]) = \int_0^{B_t[k]} T_t'(b)\, \mathbf{d}b = \int_0^k T_t'(B_t[j])\, B_t'[j]\, \mathbf{d}j \tag{18}$$
- The expression $C_t[k]$ is an abbreviation for
$$C_t[k] \equiv k - \frac{B_t[k]}{(1+r)} - T_t(B_t[k]) \tag{19}$$
$$= k - \frac{B_t[k]}{(1+r)} - \int_0^k T_t'(B_t[j])\, B_t'[j]\, \mathbf{d}j \tag{20}$$
- In a smooth interior solution, the household Euler equation is
$$U'(C_t[k]) \geq \beta \bar{R}_t[k] \int_{\Re} U'(C_{t+1}[B_t[k] + z])\, \pi(z) \mathbf{d}z \tag{21}$$
with equality if $B_t[k] > 0$. Note that the Lagrange multiplier $\mu_t[k]$ is equal to zero for those k where $B_t[k] = 0$.

[2]The main difficulty lies in the fact that any change in taxes at time t leads to behavioral responses in earlier periods that have external effects due to their impact on tax revenues.

We are looking for stationary points of the Lagrangian:

Definition 3. A **stationary point** of the Lagrangian (16), is defined as sequences of functions $T_t'(b)$, $B_t[k]$, $\phi_t[k]$, $\lambda_t[k]$, $\mu_t[k]$, and $\xi_t[k]$, and real numbers ν_t and $T_t(0)$ such that, for each t,

1. The derivative of $\mathcal{L}$ w.r.t. $T_t(0)$ is zero.

2. The derivative of $\mathcal{L}$ w.r.t. all feasible variations $\delta T_t'(b)$, $\delta B_t[k]$ and $\delta\phi_t[k]$ is equal to zero.

3. The Kuhn-Tucker conditions

$$0 = \mu_t[k]\left(-U_t'[k] + \beta\bar{R}_t[t+1\int_{\Re} U_{t+1}'[B_t[k]+z]\pi(z)\mathbf{d}z\right) \tag{22a}$$

$$0 = \xi_t[k]B_t[k], \qquad B_t[k], \xi_t[k] \geq 0, \tag{22b}$$

   hold together with Equs. (4), (8), (14), (15) and (21).

The following lemma justifies the Lagrangian approach. A sketch of a proof is given in Appendix A.2.

Lemma 5. *A smooth interior solution of Program P2 is a stationary point of the Lagrangian (16).*

4 Theoretical results

This section presents some analytical results on the structure of optimal nonlinear capital taxation, which can be interpreted as extensions of results in the theory of optimal nonlinear income taxation. These results are independent of which interpretation of the model (non-overlapping generations or infinite lives) we adopt.

4.1 The optimality conditions and their interpretation

Proposition 1. *A smooth interior solution of Program P2 satisfies*

$$\int_{\Re}\phi_t[k]\,W_t[k]\,\mathbf{d}k = \nu_t \tag{23}$$

$$\nu_t B_t'[k]\int_k^{\infty}\phi_t[j]\,\mathbf{d}j = B_t'[k]\int_k^{\infty}\phi_t[j]\,W_t[j]\,\mathbf{d}j + \phi_t[k]\,\mu_t[k]\bar{R}_t[k]U_t'[k] \tag{24}$$

for $t = 0, 1, \ldots,$ *where*

$$W_t[k] \equiv U_t'[k] + \lambda_t^*[k] + \frac{U_t''[k]}{\phi_t[k]}\frac{\partial C_t[k]}{\partial k}\int_{\Re}\phi_{t-1}[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k-B_{t-1}[j])\,\mathbf{d}j \tag{25}$$

and

$$\phi_t\,[k]\,\mu_t[k] = -\phi_t\,[k]\,\frac{\lambda_t^*[k]}{U_t''[k]} + \frac{B_t'\,[k]}{\bar{R}_t[k]}\int_{\Re}\phi_{t-1}\,[j]\,\bar{R}_{t-1}[j]\mu_{t-1}[j]\pi(k - B_{t-1}\,[j])\,\mathbf{d}j \tag{26}$$

$$\lambda_t^*[k] = \nu_t T_t'\left(B_t\,[k]\right)B_t'\,[k] + \beta B_t'\,[k]\int_{\Re}\lambda_{t+1}^*[j]\pi\left(j - B_t\,[k]\right)\,\mathbf{d}j \tag{27}$$

with initial value

$$\mu_{-1}[k] = 0 \tag{28}$$

where

$$\lambda_t^*[k] \equiv \lambda_t'[k] - U_t'[k] \tag{29}$$

Equs. (24) and (26)–(27) hold at all points where $B_t\,[k]$ is differentiable (all points except $\max\{k : B_t\,[k] = 0\}$).

Proof. See Appendix B.1.2. □

To understand the optimality conditions in Proposition 1, let us start with a discussion of the Lagrange multipliers λ and μ. The multiplier $\lambda_t[k]$ of the dynamic equation for the distribution of capital (8) gives the shadow value of having one more household with capital k. Its derivative $\lambda_t'[k]$ gives the value of having a household with capital $k+1$ rather than k, which can be interpreted as the marginal social value of one more unit of capital given to a household with capital k. The variable $\lambda_t^*[k]$ in (29) is then the marginal social value of one unit of capital in excess of its private marginal utility. This variable satisfies the integral equation (27), which has a straightforward interpretation. $\lambda_t^*[k]$ is equal to the current marginal tax contribution of the additional capital, valued with the shadow price of public revenues ν_t, plus the marginal contribution of the descendents that obtain additional bequests of $B_t'\,[k]$. This contribution is the product of $B_t'\,[k]$ and the expected marginal contribution of an additional unit of capital. The variable $\lambda_t^*[k]$ is positive if marginal tax rates are positive.

To understand the multiplier $\mu_t[k]$ of the household Euler equation (21), note that in the optimum, the government would like households to save more, since this generates more tax revenues (assuming that marginal tax rates are positive). That means, the government would like to increase $U_t'[k]$ relative to $U_{t+1}'[B_t\,[k]+z]$. It is prevented from doing so by the constraint $-U_t'[k] + \beta\bar{R}_t[k]\int_{\Re}U'[B_t\,[k]+z]\pi(z)\mathbf{d}z = 0$, which is therefore binding for the government (not for the individual!) in the direction $\geq$. The Lagrange multiplier $\mu_t[k]$ will therefore be positive. The multiplier satisfies the integral equation (26), which has the following interpretation. Some algebra shows that a relaxation of the constraint $-U_t'[k] + \beta(1+\bar{r}_t)\int_{\Re}U'[B_t\,[k]+z]\pi(z)\mathbf{d}z \geq 0$ by one unit allows the government to increase $B_t\,[k]$ by $-\frac{B_t'[k]}{U_t''[k]}$ units. The same effect on bequests would be brought about by an increase in capital of $\frac{-1}{U_t''[k]}$ units, and the value of the increase in revenues is therefore equal to $-\frac{\lambda_t^*[k]}{U_t''[k]}$; this explains the first term on the rhs of (26). Furthermore, this increase in bequests decreases consumption in t by $\frac{-B_t'[k]}{R_t[k]U_t''[k]}$ units, which changes the Euler equation of time $t-1$ by $\frac{B_t'[k]}{R_t[k]}$ units. This change affects all those households in period $t-1$ that leave a bequest such that they will end up with capital k in period t with positive probability. This is captured by the second term on the rhs of (26).

An interpretation of (28) is given in Marcet and Marimon (1998). In period t, government policy is constrained by the commitments made in period 0. This constraint works through the household Euler equation, and its effect is captured by the Lagrange multiplier μ_t. In period 0, the government can choose all parameters optimally, without being constrained by past promises, and the multiplier is therefore equal to 0.

Understanding the Lagrange multipliers, we can now interpret the expression $W_t[k]$ in (25) as the *net social marginal valuation of income*, a concept often used in the theory of optimal income taxation. It is the sum of three terms. First, the private marginal utility of income, which is equal to $U_t'[k]$ by the envelope theorem for the household. Second, the social value that it creates through current and future tax payments, equal to $\lambda_t^*[k]$. Third, the effect that it generates on the decisions of the household in earlier periods through the Euler equation of $t-1$. This is captured by the last term in (25), since the additional unit of income changes the Euler equation of $t-1$ by $U_t''[k]\frac{\partial C_t[k]}{\partial k}$ units.

Equ. (23) is the first order condition for an increase in the lump sum subsidy. It says that, in the optimum, the marginal value ν of public funds is equal to the average net social marginal valuation of income. This is completely analogous to results from the income taxation literature.

Equ. (24) is the first order condition related to an increase of the marginal tax rate $T_t'(B_t[k])$ that applies to household k. This increases by $B'[k]$ the tax burden of all households that bequeath more then $B[k]$ (cf. the representation of taxes at the rhs of Equ. (18)). Therefore $B'[k]\int_k^\infty \phi[j]\,\mathbf{d}j$ measures the resulting increase in revenue, holding household behavior fixed. Multiplying this by ν_t we obtain the lhs of (24), which measures the marginal value of the increased revenues. On the rhs of (24), the first term measures the reduction in social welfare from the reduced net income of households, while the second term measures the marginal excess burden that comes from the additional distortion of the household Euler equation.

Equ. (24) is still rather opaque; to obtain a better insight into the determinants of the optimal marginal tax rate, we can rewrite it in the following way (see Appendix B.2 for the derivation):

$$\frac{T_t'(B_t[k])}{\frac{1}{1+r}+T_t'(B_t[k])} = \left(1-\frac{Ave_t^k(W)}{Ave_t^0(W)}\right)\frac{1-\Phi_B[B_t[k]]}{B_t[k]\,\phi_B(B_t[k])}\frac{1}{\eta_R^{B^s}[t,k]} - \frac{\beta\bar{R}_t[k]}{\nu_t}\hat{\lambda}_{t+1}[k] + \frac{U_t''[k]}{\nu_t\phi_t[k]}\hat{\mu}_{t-1}[k] \quad (30)$$

Here ϕ_B and Φ_B are the cross-sectional density and distribution function of bequests, respectively; $\eta_R^{B^s}[t,k]$ is defined as the compensated elasticity of bequests w.r.t. the interest rate factor $\bar{R}_t[k]$;

$$Ave_t^k(W) \equiv \int_k^\infty \phi[j]\,W_t[j]\,\mathbf{d}j \;/\; \int_k^\infty \phi[j]\,\mathbf{d}j \quad (31)$$

is the average marginal valuation of income for households above capital level k; and finally

$$\hat{\lambda}_{t+1}[k] \equiv \int_{\Re} \lambda_{t+1}^*[j]\pi(j-B_t[k])\,\mathbf{d}j \quad (32a)$$

$$\hat{\mu}_{t-1}[k] \equiv \int_{\Re} \phi_{t-1}[j]\,\bar{R}_{t-1}[j]\mu_{t-1}[j]\pi(k-B_{t-1}[j])\,\mathbf{d}j \quad (32b)$$

The first line of the rhs of (30) is very similar to formulas derived in the optimal income tax literature (cf. Atkinson, 1995, Equ. (3.12)). It says that the marginal tax rate of a household with capital k is the higher

1. the lower is the average social marginal value of income $Ave_t^k(W)$ for people with capital above k, compared to the average $Ave_t^0(W)$.

   We can expect social marginal utility to be decreasing in k, and this term reflects the desirable "distribution effect" of increasing the marginal bequest tax and handing the revenues back as lump sum transfers.

2. the higher is the fraction of people with capital above k (who pay more tax if we increase the marginal tax rate at k)

3. the lower is the amount of bequests $B_t[k]\,\dot{\phi}_b(B_t[k])$ that is affected by the distortion from the increase in the marginal tax

4. the lower is the compensated elasticity of bequests w.r.t. the after tax interest rate.

The first item expresses a redistributive concern, the last three items are about efficiency. In addition to these effects, which already appear in static models, there are two dynamic determinants of the marginal tax function, which are captured in the second line of (30):

1. The term in $\hat{\lambda}_{t+1}[k]$ says that the optimal marginal tax of a household today is the lower the higher is the expected marginal contribution of this household to future tax revenues. This tells you to leave more money to people who will pay back more of it in the future.

2. The term in $\hat{\mu}_{t-1}[k]$ says that the optimal tax at k in t is lower if the Euler equation of those households in $t-1$ that are dynamically linked to k in t is more distorted (means that $\mu_{t-1}[j]$ is higher). From this effect, the optimal tax rate tends to decrease over time, since $\hat{\mu}$ starts at 0 and builds up over time.

4.2 No distortion at the top

Based on Proposition 1, we can now derive an analogue to the well known "no-distortion-at-the-top" result from the theory of nonlinear income taxation (see for example Myles, 1995, p.151).

Proposition 2. *In a smooth interior solution with $B_t\left[\bar{K}_t\right] > 0$, we have $\mu_t[\bar{K}_t] = 0$.*

The interesting aspect of this result is that, in a dynamic context, it is not the marginal tax rate that is set to zero at the top[3], but the Lagrange multiplier related to the household Euler equation, which measures the distortion. The marginal tax rate is probably negative. Consider, for example, the following

Corollary 1. *If $\hat{\lambda}_1[\bar{K}_0] > 0$ then $T_0'\left(\bar{K}_0\right) < 0$.*

To see this, combine (27) and (26) to get

$$\mu_t[k] = -\frac{\nu_t T_t'\left(B_t\left[k\right]\right) B_t'\left[k\right]}{U_t''[k]} - \frac{\beta B_t'\left[k\right]}{U_t''[k]}\hat{\lambda}_{t+1}[k] + \frac{B_t'\left[k\right]}{\phi_t\left[k\right]\bar{R}_t[k]}\hat{\mu}_{t-1}[k] \tag{33}$$

and observe that $\hat{\mu}_{-1}[k] = 0$ (recall the definitions of $\hat{\lambda}$ and $\hat{\mu}$ in (32)).

The corollary says that the top household in period 0 faces a negative marginal tax if its expected future marginal tax contribution, as measured by $\hat{\lambda}_1[\bar{K}_0]$, is positive. This is likely to be the case, since the distribution is mixing every period, and the same household will not be at the top of the wealth distribution in future periods.

5 Some numerical illustrations

5.1 Parameter values and functional forms

The utility function is of the CRRA form

$$U(C) = \frac{C^{1-\gamma} - 1}{1-\gamma} \tag{34}$$

with γ taking the values 1 or 2.

[3] To be precise, what one can show is that $\lim_{k\to\bar{K}_t}\mu_t[k] = 0$, while Equ. (26) does not pin down $\mu_t[\bar{K}_t]$, because $\phi_t\left[\bar{K}_t\right] = 0$. The distinction is irrelevant, since there is no mass at $\bar{K}_t$.

For the lifetime income process, I consider two different distributions. The first is the lognormal distribution, truncated at the lower end at the point "mean minus 4 standard deviations", where income is very close to zero. To this distribution I add a fixed income of 1, so that the minimum income is approximately 1, which is about 12 percent of median income in the baseline specification with a Gini coefficient of 0.3. The lognormal distribution is the one most intensively studied in the optimal taxation literature. However, recent work by Diamond (1998) and Saez (2001) has made clear that the choice of distribution drives to a large extent the results of optimal income tax exercises. The fact that the lognormal distribution has very thin tails is mainly responsible for the result that optimal marginal tax rates are usually found to decline for higher incomes. The upper tail of the American income distribution can be better described by a Pareto distribution, with exponential parameter of about 2 (Saez 2001). The second distribution I consider is therefore a mixed distribution, where the lower 9 deciles follow a lognormal distribution, while the upper decile is Paretian. To obtain a density function that is twice differentiable (as is required in the calculations above), the distribution was smoothed in the region around the switch-point.

For both distributions, I investigated different levels of inequality, with Gini coefficients between 0.3 and 0.4. With the mixed distribution described above, a Gini coefficient of 0.3 implies a variance of log income of about 0.23, which is close to the most recent estimates for log lifetime income in Haider (2001, Table 5), obtained from PSID data. It is well known that the type of models that we analyze do not generate enough wealth inequality. To partially compensate for that, I am also considering a Gini coefficient of 0.4. The variance of log income then increases to 0.47.[4]

The before tax interest rate was set to 1.0236^{30}, to be understood as the return over 30 years with an annual rate of 2.36 percent, which is the average real return on 1 year US treasury bonds over 1959-1998. The discount factor β was set to 0.429, which conforms to a value of 0.972 annually. With this parameter choice, the fraction of households leaving zero bequests (in the long run under nonlinear taxation) varies between 10 and 50 percent for the different parameter constellations considered. In the data, this number is about 30 percent. (Menchik and David 1983, Table 3.1)).

In the numerical application, I have to approximate the nonlinear tax schedule by a flexible functional form. The results below are based on tax functions that have 11 free parameters for every period. More details are given in Appendix C.

[4]Diaz-Gimenez, Quadrini and Rios-Rull (1997, Table 1) find a Gini coefficient of 0.63 for recent US annual earnings. The coefficient for lifetime earnings should be considerably smaller.

5.2 Results

5.2.1 Average tax rates and welfare gains

Table 1 provides a number of summary statistics on optimal tax rates and their effects on aggregate variables. For each parameter constellation, the first line gives the results for the linear tax, the second line for the nonlinear tax.

I consider specification 1) as the benchmark. The "M" in the column titled "ϕ" indicates that the income distribution is of the "mixed type", the lower part lognormal, the upper part Paretian. $\gamma = 1$ means we have log-utility, the Gini of labor income is set to 0.3, and the income distribution was truncated at the upper end such that the maximal income is 73.8 times the median income. In the initial period ($t = 0$), the linear tax on bequests is 23.03 percent, while the nonlinear tax is on average 22.40 percent. Both regimes give rise to lump sum subsidies of about 7.3 percent of median income. Marginal taxes are progressive in the sense that the correlation $\rho_{T',B}$ between bequests and the marginal tax rate is positive. Nine generations later ($t = 9$), the last period that we simulate, tax rates are much lower, and the subsidy has decreased to about 2 percent of median income. This reflects the higher efficiency costs of taxes that are announced a long time in advance, since households have more time to evade it by reducing savings (cf. the effect of $\hat{\mu}$ in Formula (30)).

In the next column, Δv reports the total welfare gain for the Utilitarian government. It is expressed in the table in terms of an equivalent proportional, permanent change in consumption of all households. In the benchmark case, the welfare gain is equivalent to about 2 percent of consumption. This is a significant, but not a huge gain. It may be surprising the the welfare gain of the nonlinear tax is only slightly higher than that of the linear tax. We will look at this in more detail later.

Since the bequest tax is redistributive, it is clear that some households will gain and others loose. The column "Frac Gain" in the table reports the percentage of households that gains from having the optimal tax rather than no tax, in the sense of having a higher value function. Notice that, for any parameter constellation, all tax regimes (no tax, linear tax, nonlinear tax) start from the same initial distribution, which corresponds to the steady state distribution under the no tax regime. In the initial period, we see that the fraction of households that gain is much higher under the nonlinear than under the linear tax, despite the fact that the average welfare gains are so close. This makes clear that if we ask for the possible political support of a bequest tax, the availability of nonlinear taxes may be much more important than the Utilitarian welfare criterion suggests.

The column $t = 1$ reports the same statistic for the following period. Notice that now we have to compare welfare across different wealth distributions that arise endogenously in the

					t=0			t=9				Frac Gain		Frac $B_0 = 0$		
	ϕ	γ	Gini z	$\frac{Y^{max}}{Y^{med}}$	$\frac{T}{B}$	$\frac{M}{Y^{med}}$	$\rho_{T',B}$	$\frac{T}{B}$	$\frac{M}{Y^{med}}$	$\rho_{T',B}$	Δv	$t = 0$	$t = 1$	no tax	tax	$\Delta \ln k$
1)	M	1	0.3	73.8	23.03	7.34	0.00	7.98	2.19	0.00	1.96	38.1	2.8e-08	33.0	61.9	-0.24
					22.40	7.26	0.58	6.90	1.92	0.75	2.03	54.2	5.4e-06	33.0	45.8	-0.24
2)	M	1	0.3	17.7	17.41	5.72	0.00	6.58	1.84	0.00	1.33	47.5	2.9e-06	30.7	52.5	-0.21
					17.38	5.73	0.46	6.08	1.70	0.65	1.36	55.0	2.1e-06	30.7	45.0	-0.21
3)	M	2	0.3	53.9	54.15	42.11	0.00	21.55	6.92	0.00	10.98	79.9	2.0e-09	9.1	20.1	-0.60
					56.07	43.18	0.60	18.22	6.44	0.20	11.31	65.0	1.4e-06	9.1	9.8	-0.57
4)	M	2	0.3	17.7	49.46	37.47	0.00	17.19	6.25	0.00	8.97	83.6	5.3e-09	8.2	16.4	-0.53
					48.17	36.91	-0.05	17.21	6.25	-0.26	8.98	85.2	1.2e-08	8.3	14.8	-0.53
5)	M	1	0.4	23.6	20.79	12.89	0.00	8.24	3.91	0.00	2.69	68.0	1.2e-05	17.3	32.0	-0.32
					20.86	12.97	-0.80	8.39	3.94	-0.35	2.71	65.7	2.3e-05	17.3	34.3	-0.32
6)	L	1	0.3	11.3	10.89	3.56	0.00	4.59	1.31	0.00	0.64	63.2	4.9e-06	24.9	36.8	-0.15
					11.21	3.68	-0.90	5.14	1.46	-0.82	0.65	58.8	9.4e-06	24.9	41.2	-0.15

first line of each parameter constellation: linear tax; second line: nonlinear tax

ϕ: type of income distribution; "L":lognormal; "M": mixed lognormal and Paretian

γ: risk aversion parameter; Gini z: Gini coefficient of income distribution, Y^{max}/Y^{med}: maximum over median income

T/B: total tax over total bequests; M/Y^{med}: subsidy in percent of median income

$\rho_{T',B}$: correlation coefficient between bequests and marginal tax rate

Δv: welfare gain of taxation

Frac Gain: percentage of households that gain in tax vs. no-tax regime; cf. text for details

Frac $B_0 = 0$: percentage of households in $t = 0$ leaving zero bequests, in tax and no-tax regime

$\Delta \ln k$: reduction in log of average capital, tax compared to no-tax regime

Table 1: Results for numerical examples

different regimes. Therefore, we do not compare households with the same level of capital k in both regimes, but households that occupy the same position in the wealth distribution (for example, we compare the household in the fifth wealth percentile under the no-tax and under the nonlinear tax regime).[5] Recall that we do not compare instantaneous utility, but the value function of the household as of time $t = 1$. With this criterion, from the second period onwards practically nobody gains! We can understand this result if we see that the high tax rates in the initial period strongly reduce the total amount of bequests. The columns "Frac $B_0 = 0$" compare the fraction of households that leave zero bequests, under the no-tax and the tax regimes, in $t = 0$. The introduction of the tax raises this number from 33 to 62 and 46 percent in the linear and nonlinear case, respectively. From a distributional perspective, the fact that now there are many more households who inherit nothing, more than compensates the lump sum subsidy that is financed by the tax. Second, this extreme result is also a consequence of the fact that in this model the lower part of the income distribution is lognormal. If there were more people bunched at zero market income and zero bequests, only depending on subsidies, these would benefit from the introduction of the tax also in later periods. In any case, it seems that only the very poorest households in the economy still benefit from the tax in the second period. Finally, the column "$\Delta \ln k$" reports the difference in the log of average wealth between the tax and the no-tax regime, at $t = 9$. We see that the decrease in the capital stock accumulates to more than 20 percent in the long run.

The parameter constellation 2) differs from 1) in the parameter $\frac{Y^{max}}{Y^{med}}$, the ratio of maximal to median income. While log-linear and Pareto distributions are unbounded, for numerical reasons they are truncated at a certain point. When $\frac{Y^{max}}{Y^{med}}$ changes, the standard deviation of income is adjusted such that the same Gini coefficient is maintained. The result is a considerable reduction in average tax rates, and the advantage of nonlinear over linear taxation is even further reduced.

This last conclusion is borne out even more clearly by the comparison of cases 3) and 4), where the risk aversion parameter γ is increased to 2. The gain of nonlinear taxation is non-negligible if the upper tail of the distribution is long, but it is again very low if the tail is short. Tax rates are generally much higher, even in the long run. As a consequence of the high tax rates, the capital stock shrinks by almost 50 percent in the long run.

In Case 5), the degree of inequality is increased to a Gini coefficient of 0.4. As one would expect, tax rates and the welfare gain go somewhat up compared to Case 2), but tax rates

[5] Notice that this is not equivalent to following the dynasties across different regimes. For example, a household that received zero inheritance and got a rather bad income shock may be at point 0.01 in the income distribution in the no-tax regime, while under the linear tax the same household may end up at position 0.02, since there are now many more households that got no inheritance.

go down compared to Case 1) with the longer tail. The final Case 6 considers a lognormal distribution, which is a distribution with a very thin tail. Both tax rates and welfare gain drop strongly.

In sum, the optimal average tax rate is increasing in the level of inequality, as measured by the Gini coefficient of labor income, and in the level of risk aversion (γ). Most importantly, taxes are higher if the tail of the income distribution is longer (keeping the Gini coefficient constant). While the total welfare gain can be substantial, it is a temporary phenomenon. The gain accrues almost exclusively to the first generation. These results apply both to the linear and the nonlinear tax.

5.2.2 The shape of the tax function

From the graphs of the optimal nonlinear tax rates in Fig. 1, we see that the marginal tax rate drops to low or negative levels for very high bequests, which is broadly consistent with the no-distortion-at-the-top results of Section 4.2. That tax rates are not always negative at the top is probably due to numerical problems: in the numerical computations we are using a finite parameterization of the tax function. Particularly at the upper end, there are so few people out there that it is very difficult to pin down the marginal tax rates precisely.

For low and middle level of bequests, the marginal tax rate is either increasing or U-shaped in cases where the income distribution has a Pareto upper tail, while it is decreasing with a lognormal distribution. This is very similar to the results in the income taxation literature (Saez 2001), and can be understood from Equ. (30). It is driven by the term $\frac{1-\Phi_B[B_t[k]]}{B_t[k]\phi_B(B_t[k])}$, which is constant if $B_t[k]$ follows a Paretian distribution, while it decreases rapidly in the lognormal case. A thin tail of the income distribution induces a thin tail of the bequest distribution, and this causes a falling optimal marginal rate.

To see how many people are affected by these tax rates, Fig. 2 displays the same functions as Fig. 1, but now there are not the bequests on the x-axis, but the cumulative distribution of households. The value at $x = 0.9$ means the marginal tax rate faced by the household that is at the 90-th percentile of the wealth distribution. We see that in the Paretian case, only a small, sometimes negligible, number of households are in the range of falling tax rates.

Fig. 3 should shed light on two questions: Why is the welfare gain of nonlinear taxation only minimally higher than that of linear taxation? Have we really exhausted the possible gains from nonlinear taxation by our parametric approximation, which contains 11 parameters in every period? The figure displays the residual of the first order condition (24) of the government problem, in $t = 0$. On the x-axis we now have capital k, expressed as a fraction of the upper bound $\bar{K}_0$. The range is truncated above 0.7, because the residual is indistinguishable from 0 for higher values of k. The reason is that the cross-sectional density

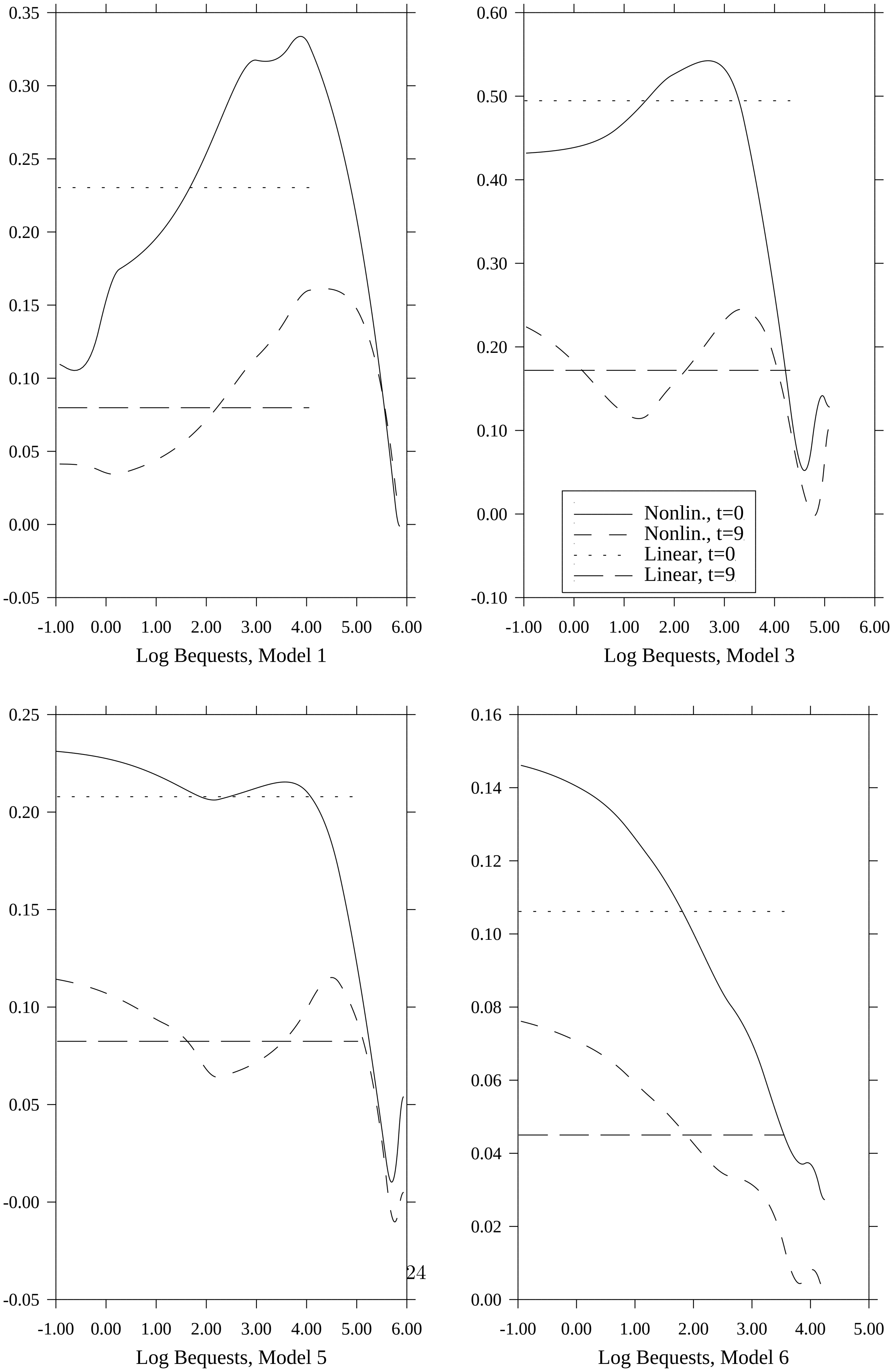

Figure 1: Optimal marginal tax rates

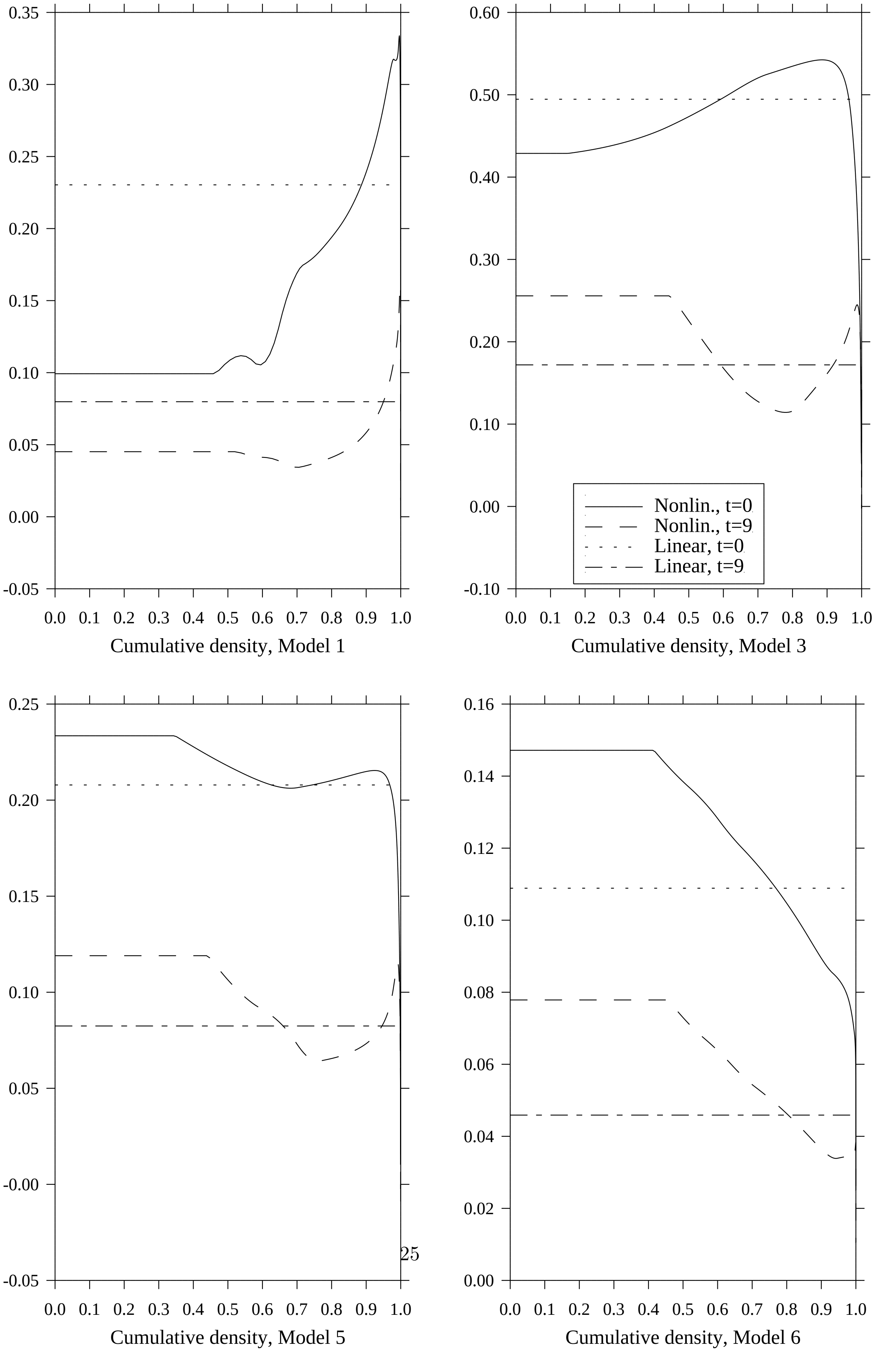

Figure 2: Optimal marginal tax rates

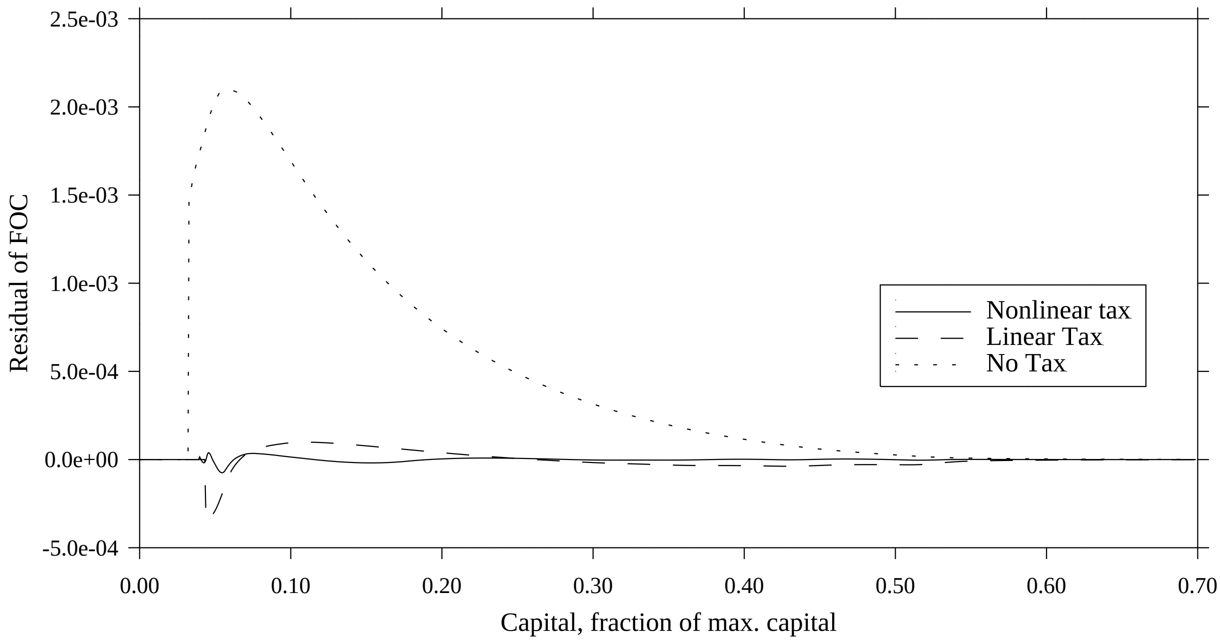

Figure 3: First order condition, benchmark model, $t = 0$

$\phi_0 [k]$ enters the first order condition, and $\phi_0 [k]$ is extremely small for very high levels of capital. By definition, this residual is identically zero for the fully nonlinear optimal tax (if it is computed exactly). For the optimal linear tax, it is straightforward to show[6] that the integral of this function over k must be zero. What does this picture tell us about welfare gains? To a quadratic approximation, the value loss from having chosen a suboptimal policy is a quadratic function of the first order residuals.[7] In other words, the welfare gain that can be realized by going from a suboptimal solution (no tax, linear tax, imprecisely calculated nonlinear tax) to the fully optimal tax system is a quadratic function of the residuals displayed in Fig. 3. Since the residuals of the optimal linear tax are almost one order of magnitude smaller than the residuals with zero tax, the welfare gain left after linear taxation is almost two orders of magnitude smaller than the gain starting from no tax, which is in line with the numbers in Table 1 Regarding the accuracy of the computed optimal nonlinear tax, we see that it does not give a zero residual everywhere, but it seems to exhaust the biggest part of the gain that is left by the linear tax.

[6] We first consider (73) where we take $\delta T_t' [j]$ as constant over j (the case of an optimal linear tax). Then we integrate the first order condition (24) over k, and we see that the two equations are equivalent, after a simple change of variables.

[7] Minimizing $f(x) \equiv \frac{1}{2} x' Q x$ over the vector x, for positive definite matrix Q, the value loss $f(x) - f(0)$ is proportional to $R(x)' Q^{-1} R(x)$, a quadratic function of the first order residual $R(x) \equiv Qx$.

6 Conclusions

This paper has analyzed the role of capital (or bequest) taxation for the purpose of redistribution. The theoretical analysis has established strong analogies between the optimal nonlinear taxation of capital and the well known static theory of optimal nonlinear income taxation, including a non-distortion-at-the-top result.

The quantitative analysis has used a calibration of the model with a time period of 30 years. This was interpreted as a model of non-overlapping generations, where savings come in the form of bequests. It has been found that optimal bequest tax rates are high in the long run only if the risk aversion coefficient is higher than 1. The welfare gain may then be substantial, but in all the examples considered, the welfare gain goes almost exclusively to members of the first generation. From the second generation on, the decrease in inheritances more than compensates the welfare gains from redistributional taxes, for all but a tiny number of very poor households. The degree of income inequality has an impact on average tax rates and even more so on the shape of the optimal tax function, but what counts is the upper tail of the distribution, not so much overall inequality measures such as the Gini coefficient. As in optimal income taxation, marginal tax rates are decreasing if income is lognormally distributed, while they are increasing or U-shaped if the upper tail is Paretian.

In all model specifications, the welfare gain from using the optimal nonlinear tax versus the optimal linear tax is much smaller than the gain from the linear tax versus no tax. This conclusion may have to be modified in models that give a more accurate description of the very upper tail of the income and wealth distribution (in the simulations, I truncated the income process at maximally 75 times median income, for numerical reasons). In general, the model of this paper is very stylized, and future work should test the numerical conclusions by considering richer models; to make the analysis tractable, one would have to use more restricted forms of nonlinear taxation, such as fewer brackets and less variation over time. The aim of the present analysis was to study the mechanism of optimal nonlinear dynamic taxation, to show what optimal tax functions look like and what determines their characteristics. The results should also be helpful in future work to find the right simplifications to analyze richer models.

Appendix

A Proofs for Section 3

Proof of Lemma 1. a) Plugging Equ. (5b) into the budget constraint (2b), the claim follows from Assumptions 2ii) and 4ii).

To show b), define $m_{k,t} \equiv \int_{\Re} k\phi_t[k]\,\mathbf{d}k$ and $m_z \equiv \int_{\Re} z\pi(z)\,\mathbf{d}z$. The aggregate resource constraint (obtained from the government budget constraint and the sum of household budget constraints) then implies $m_{k,t} \leq (1+r)m_{k,t-1} + m_z$. Defining $\tilde{m}_{k,t} \equiv (1+r)^{-t}m_{k,t}$, we get $\tilde{m}_{k,t} \leq \tilde{m}_{k,t-1} + (1+r)^{-t}m_z$, therefore $\tilde{m}_{k,t} \leq \tilde{m}_{k,0} + \sum_{i=1}^{t}(1+r)^{-i}m_z$. Straightforward algebra gives $m_{k,t} \leq (1+r)^t(m_{k,0} + m_z/r)$.

The household budget constraint together with the last inequality then implies

$$\bar{K}_{t+1} \leq (1+r)\bar{K}_t + \overline{z} + \underline{\tau}[(1+r)^t(m_{k,0} + m_z/r)]$$

A similar argument as above shows that $\bar{K}_t \leq (1+r)^t\left(\bar{K}_0 + t(\underline{\tau}(m_{k,0} + m_z/r) + \overline{z})\right)$ The claim then follows immediately.

□

Proof of Lemma 2. Strict monotonicity follows from standard arguments. That $v_t(0) > -\infty$ follows directly from Equ. (5a)

To see that $v_t(\bar{K}_t) < \infty$, take any $\hat{\beta}$ with $\beta \leq \hat{\beta} \leq 1/(1+r)$. Lemma 1 shows that $\lim_{t\to\infty}\hat{\beta}^t\bar{K}_t = 0$. Together with Equ. (5b) this implies that $\lim_{t\to\infty}\hat{\beta}^t\bar{c}_t = 0$, where $\bar{c}_t$ is the maximal possible level of consumption in t. Because of the concavity of $U(.)$, this implies

$$\lim_{t\to\infty}\hat{\beta}^t U(\bar{c}_t) \leq \lim_{t\to\infty}\hat{\beta}^t\left[U(\underline{z}) + U'(\underline{z})(\bar{c}_t - \underline{z})\right] = 0 \tag{35}$$

From (35), for any $\epsilon > 0$ there exists t_ϵ s.t. $\hat{\beta}^t U(\bar{c}_t) \leq \epsilon$ for all $t \geq t_\epsilon$. Then, for any t,

$$\begin{aligned} v(\bar{K}_t) &\leq \sum_{i=0}^{t_\epsilon-1}\beta^i U(\bar{c}_{t+i}) + \sum_{i=t_\epsilon}^{\infty}\beta^i U(\bar{c}_{t+i}) \leq \sum_{i=0}^{t_\epsilon-1}\beta^i U(\bar{c}_{t+i}) + \hat{\beta}^{-t}\sum_{i=t_\epsilon}^{\infty}\left(\frac{\beta}{\hat{\beta}}\right)^i\hat{\beta}^{t+i}U(\bar{c}_{t+i}) \\ &\leq \sum_{i=0}^{t_\epsilon-1}\beta^i U(\bar{c}_{t+i}) + \hat{\beta}^{-t}\sum_{i=t_\epsilon}^{\infty}\left(\frac{\beta}{\hat{\beta}}\right)^i\epsilon \leq \sum_{i=0}^{t_\epsilon-1}\beta^i U(\bar{c}_{t+i}) + \hat{\beta}^{-t}\frac{\epsilon}{1-\beta/\hat{\beta}} \leq \infty \end{aligned}$$

ii) For any $\epsilon > 0$, define t_ϵ as above. Then, for any $t \geq t_\epsilon$,

$$\beta^t v\left(\bar{K}_t\right) \leq \beta^t\sum_{i=0}^{\infty}\beta^i U(\bar{c}_{t+i}) = \sum_{i=t}^{\infty}\beta^i U(\bar{c}_i) = \sum_{i=t}^{\infty}\left(\frac{\beta}{\hat{\beta}}\right)^i\hat{\beta}^i U(\bar{c}_i) \leq \sum_{i=t}^{\infty}\left(\frac{\beta}{\hat{\beta}}\right)^i\epsilon \leq \frac{\epsilon}{1-\beta/\hat{\beta}}$$

Since such a t_ϵ can be found for any ϵ, the claim follows.

iii) From the boundedness of $v_{t+1}(k)$ and Assumption 3ii), optimal consumption is bounded away from zero at each t. Denote the infimum of consumption at t by $\underline{c}_t > 0$. Then we have, for all k and all $\epsilon \leq \underline{c}_t/2$ that $v_t(k) - v_t(k-\epsilon) \leq U(C_t[k]) - U(C_t[k] - \epsilon) \leq U(\underline{c}_t) - U(\underline{c}_t - \epsilon)$. This implies that $v_t(k)$ is Lipschitz continuous with constant $(U(\underline{c}_t) - U(\underline{c}_t - \epsilon))/\epsilon$. The claim then follows from Foran (1991, Theorem (8.17)). □

Proof of Lemma 3. Any $b_k \in B^*_t[k]$ satisfies

$$b_k = \underset{b}{\operatorname{argmax}} \left\{ U\left(k - \frac{b}{(1+r)} - T_t(b)\right) + \beta \int_{\Re} v_{t+1}(b+z)\pi(z)\mathbf{d}z \right\} \tag{36}$$

Now assume that $k_1 > k_0$, and $b_1 < b_0$, where $b_0 \in B^*_t[k_0]$ and $b_1 \in B^*_t[k_1]$. Define $\Delta v \equiv \beta \int_{\Re} v_{t+1}(b_0+z) - v_{t+1}(b_1+z)\pi(z)\mathbf{d}z$. This implies

$$\begin{aligned} U\left(k_1 - \frac{b_1}{(1+r)} - T_t(b_1)\right) - U\left(k_1 - \frac{b_0}{(1+r)} - T_t(b_0)\right) &\geq \Delta v \\ U\left(k_0 - \frac{b_1}{(1+r)} - T_t(b_1)\right) - U\left(k_0 - \frac{b_0}{(1+r)} - T_t(b_0)\right) &\leq \Delta v \end{aligned}$$

This contradicts the concavity of U. □

A.1 Admissible tax functions

To proof Lemma 4, we first show that we can restrict the government's choice of tax functions to a class of functions that we call "envelope tax functions", because they are essentially the envelope of the households indifference curves in (B,T)-space. Then we show that for these functions, the conditions Equs. (13)–(15) are sufficient for a solution of the household problem, which is the core part of Lemma 4.

The indifference curves in (B,T)-space are defined by the indirect utility function $V_t(B,T;k,\Theta_{t+1})$:

$$V_t(B,T;k,\Theta_{t+1}) \equiv U(k - B/(1+r) - T) + \beta\,\mathrm{E}_t\, v_{t+1}(B+z;\Theta_{t+1}) \tag{37}$$

Given next period's value function, it expresses the value that a household with capital k reaches in period t as a function of bequests B and taxes paid T. Everything is conditional on future policy Θ_{t+1}, which is suppressed as an argument in the following. The slope $I_t(B,T;k)$ of this indifference curve is given as

$$I_t(B,T;k) = \frac{dT}{dB} = \frac{\beta\,\mathrm{E}_t\, v'_{t+1}(B+z)}{U'(k - B/(1+r) - T)} - \frac{1}{1+r} \tag{38}$$

which can be positive or negative.[8] The following derivations will make essential use of the fact that the indifference curves fulfill a single crossing condition. The indifference curve in

[8]Note that $v'_{t+1}(B+z)$ may not exist at a countable number of points, but the expectation is still well defined, because it is an integral w.r.t. to a continuous distribution.

a given point is the steeper the higher is k, because

$$\frac{\partial I_t\,(B,T;k)}{\partial k} = -\frac{U''(k-B/(1+r)-T)}{[U'(k-B/(1+r)-T)]^2}\beta\,\mathrm{E}_t\,v'_{t+1}\,(B+z) > 0 \tag{39}$$

using the strict monotonicity of the value function (Lemma 2i).

We are now ready to define envelope tax functions, which basically eliminate the "irrelevant parts" of the tax function.

Definition 4. Take as given a series of piecewise smooth tax functions $T_t^0\,(b) : (0,\infty) \to \Re$. Denote by $B_t^0\,[k] : (\underline{z}, \bar{\bar{K}}_t) \to (0,\infty)$ the optimal bequest correspondence under $T_t^0\,(k)$, and by $v_t^0(k) : (\underline{z}, \bar{\bar{K}}_t) \to \Re$ the corresponding household value function. Define R_t^B as the range of $B_t^0\,[k]$.

For every t, the **envelope tax function** $T_t^{env}\,(b)$ relating to $T_t^0\,(b)$ is defined as follows:

1. If $b \in R_t^B$, then $T_t^{env}\,(b) = T_t^0\,(b)$.

2. For $b \leq b_t^{inf}$, where $b_t^{inf} \equiv \inf_{\tilde{b}}\left\{\tilde{b} \in R_t^B\right\}$, define $T_t^{env}\,(b) \equiv T_t^0\left(b_t^{inf}\right) - \frac{1}{1+r}(b - b_t^{inf})$. This construction means that bequests are fully subsidized up to b_t^{inf}, so that a lower level of b is not optimal for any household, nor are the first order conditions of any household satisfied at those levels.

3. If $b \geq b_t^{sup}$, where $b_t^{sup} \equiv \sup_{\tilde{b}}\left\{\tilde{b} \in R_t^B\right\}$, define $T_t^{env}\,(b) \equiv T_t^0\,(b_t^{sup}) + \kappa(b - b_t^{sup})$, where the constant κ is chosen big enough so that bequests above $\bar{K}_t$ are prohibitively expensive, that means, those levels are neither optimal nor do they satisfy the first order conditions for any household. Since all the relevant variables are bounded, such a κ can always be found.

4. If $b_t^{inf} < b < b_t^{sup}$ and $b \notin R_t^B$, define $b^0 \equiv \sup_{\tilde{b}\in R_t^B}\{\tilde{b} < b\}$ and $b^1 \equiv \inf_{\tilde{b}\in R_t^B}\{\tilde{b} > b\}$. From the continuity of the utility and the value function, it is clear that there is exactly one k_0 such that a household with level k_0 is indifferent between b^0 and b^1. Define $T_t^{env}\,(b)$ by the level of taxes such that household k_0 is indifferent between b and b^0, that means,
$U\left(k - \frac{b}{1+r} - T_t^{env}\,(b)\right) + \int_{\Re} v_{t+1}(b+z)\pi(z)\mathbf{d}z =$
$U\left(k - \frac{b^0}{1+r} - T_t^{env}\left(b^0\right)\right) + \int_{\Re} v_{t+1}(b^0+z)\pi(z)\mathbf{d}z.$
What is important here is that the single-crossing condition (39) guarantees that the switch from $T_t^0\,(b)$ to $T_t^{env}\,(b)$ does not affect the bequests of any household other than k_0. For any $b \in (b^0, b^1)$, households with $k > k_0$ prefer b^1 over b, while those with $k < k_0$ prefer b^0 over b.

Furthermore, this construction guarantees that the bequest correspondence under the envelope tax function is u.h.c., because at all jump points, the household with capital k_0 is indifferent between all bequest levels in (b_0, b_1).

This construction is illustrated in Figure 4. The solid line gives the original tax func-

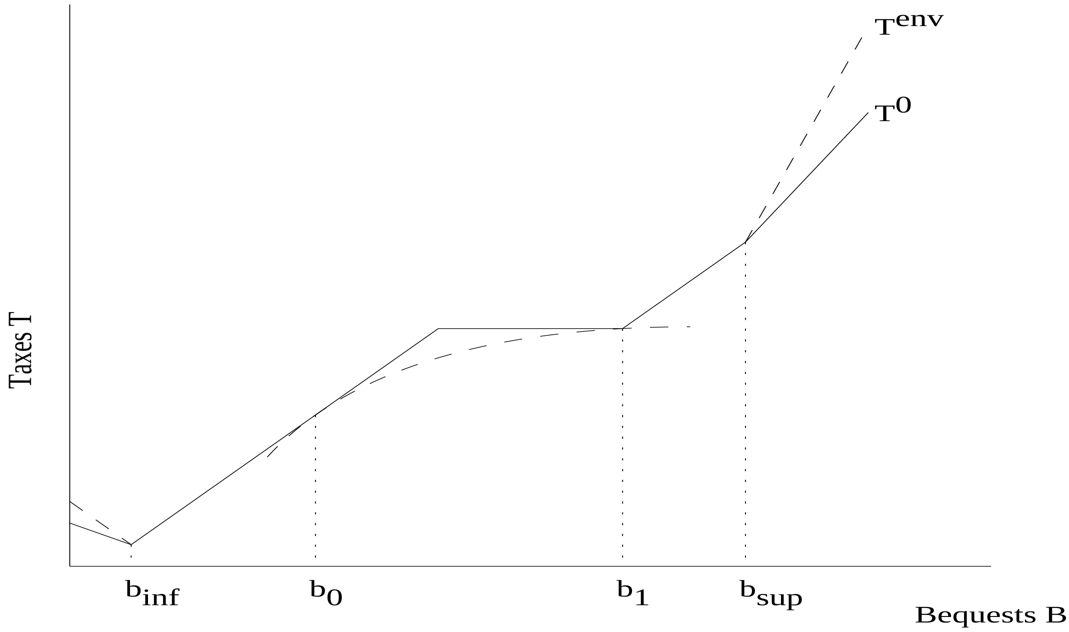

Figure 4: Admissible tax functions

tion $T_t^0(b)$. The envelope tax function $T_t^{env}(b)$ coincides with with $T_t^0(b)$, except for three intervals. For $b \in (b_0, b_1)$, the new tax function is given by the indifference curve (dashed line) of the household that is indifferent between b_0 and b_1. For $b < b_t^{inf}$, it is given by the dashed line that is steeper than the original line. Note that the graph covers the case that the household with the lowest capital still leaves a positive bequest (which is very unlikely to happen in the optimal solution, but is possible with government subsidies). For $b > b_t^{sup}$, the new tax function is illustrated by the steep dashed line.

Lemma 6. *To any piecewise smooth tax policy $T_t^0(b)$, the corresponding envelope tax policy $T_t^{env}(b)$ is u.h.c. and is equivalent to $T_t^0(b)$ in the sense that it induces the same bequest behavior for almost all households and raises the same tax revenue.*

Proof. The explanations given in Definition 4 have shown that the bequest function is u.h.c., and that the value function of households is unchanged if we replace $T_t^0(b)$ by $T_t^{env}(b)$. The change can only affect the bequest correspondence of those households where a discontinuity appears under $T_t^0(b)$. Since the bequest correspondence is monotonic, this can happen at only a countable number of capital values. Since the cross-sectional distribution functions of k are continuous w.r.t. Lebesgue measure, only a zero measure set of households are affected. For that reason, there is no effect on tax revenues. $\square$

Lemma 7. *Any sequences of value functions $v_t(\underline{z})$, $v_t'(k)$, weakly increasing and upper hemi-continuous (u.h.c.) bequest correspondences $B_t^0[k]$ and envelope tax functions $T_t(B_t[b])$ satisfy Equs. (13)–(15) if and only if they satisfy*

$$B_t^0[k] = \underset{b}{\operatorname{argmax}} \left\{ U\left(k - \frac{b}{1+r} - T(b)\right) + \beta \int_{\Re} v_{t+1}(b+z)\pi(z)\mathbf{d}z \right\} \tag{40}$$

$$v_t(k) = \max_b \left\{ U\left(k - \frac{b}{1+r} - T(b)\right) + \beta \int_{\Re} v_{t+1}(b+z)\pi(z)\mathbf{d}z \right\} \tag{41}$$

Proof. Deriving Equs. (13)–(15) from Equs. (40) and (41) is standard.

Equ. (41) is derived from Equs. (13)–(15), as follows:

$$\begin{aligned}
v_t(k) &= v_t(\underline{z}) + \int_{\underline{z}}^k v_t'(j)\,\mathbf{d}j \\
&= U(C_t[\underline{z}])) + \beta \int_{\Re} v_{t+1}(B_t[\underline{z}] + z)\pi(z)\mathbf{d}z + \int_{\underline{z}}^k \Bigg\{ U'(C_t[j])) \\
&- \left[\left(\frac{1}{1+r} + T'(B_t[j])\right) U'(C_t[j])) + \beta \int_{\Re} v_{t+1}'(B_t[j] + z)\pi(z)\mathbf{d}z\right] \frac{dB_t[j]}{dj} \Bigg\} \mathbf{d}j \\
&= U(C_t[\underline{z}]) + \int_{\underline{z}}^k U'(C_t[j])) \left(1 - \left(\frac{1}{1+r} + T'(B_t[j])\right) \frac{dB_t[j]}{dj}\right) \mathbf{d}j \\
&+ \beta \int_{\Re} \left[v_{t+1}(B_t[\underline{z}] + z) + \int_{\underline{z}}^k v_{t+1}'(B_t[j] + z) \frac{dB_t[j]}{dj} \mathbf{d}j \right] \pi(z)\mathbf{d}z \\
&= U(C_t[k]) + \beta \int_{\Re} v_{t+1}(B_t[k] + z)\pi(z)\mathbf{d}z
\end{aligned} \tag{42}$$

In (42), the term $C_t[k]$ should always be understood as an abbreviation for $k - \frac{B_t[k]}{(1+r)} - T_t(B_t[k]))$. The second line of (42) uses (14) and (15). The third line is equal to zero, because the Euler equation (13) holds with equality for almost all values of k. The latter holds because $B_t[k]$ is differentiable w.r.t. k almost everywhere (because it is monotonic, c.f. Foran, 1991, Theorem (8.2)), and because $T_{t+}'(b) = T_{t-}'(b)$ for almost all b (by the definition of a piecewise smooth tax function). The forth and fifth line are just a slight reordering, and the last line follows because the terms under the integral $\int_{\underline{z}}^k$ are the derivatives of the utility function and the value function w.r.t. k, respectively. The last line is equivalent Equ. (41), because $B_t[k]$ is defined as the maximizing b.

It remains to derive Equ. (40) from Equs. (13)–(15). To that end, define

$$w(k,b) \equiv U\left(k - \frac{b}{1+r} - T(b)\right) + \beta \int_{\Re} v_{t+1}(b+z)\pi(z)\mathbf{d}z \tag{43}$$

We then have to show that $w\left(k, b^0\right) \geq w(k,b)$ for each $b^0 \in B_t^0[k]$ and any b. Since we already know that any b not in the range R_t^B of $B_t^0[k]$ is not optimal for any household under the policy $T_t(B_t[b])$ (by construction of an envelope tax policy, cf. the explanations in Definition 4), we can restrict ourselves to the $b \in R_t^B$. (We can further restrict attention to those b that leave a positive level of consumption.) Then, for every k, $b^0 \in B_t^0[k]$ and $b \in R_t^B$,

$$\begin{aligned} w\left(k,b^0\right) - w(k,b) &= \int_b^{b^0} \frac{\partial w\left(k,\tilde{b}\right)}{\partial \tilde{b}} \mathbf{d}\tilde{b} && (44)\\ &\geq \int_b^{b^0} \frac{\partial\left[U\left(\mathcal{K}\left(\tilde{b}\right) - \frac{\tilde{b}}{1+r} - T\left(\tilde{b}\right)\right) + \beta\int_{\Re} v_{t+1}(\tilde{b}+z)\pi(z)\mathbf{d}z\right]}{\partial \tilde{b}} \mathbf{d}\tilde{b} && (45)\\ &= 0 && (46)\end{aligned}$$

where $\mathcal{K}\left(\tilde{b}\right) \equiv \sup\left\{k : \tilde{b} \in B_t^0[k]\right\}$. For $\tilde{b} \leq b^0$, we have $\mathcal{K}\left(\tilde{b}\right) \leq k$ and the inequality (45) follows from the concavity of $U(c)$. In the case $b \geq b^0$, the terms in the integral in (45) are bigger than their counterparts in (44), but are multiplied by -1 in the integration. (46) follows again from (13), since the Euler equation holds with equality for almost all k.

The critical step in the derivation is that the whole range of integration of (44) is in R_t^B, which is a closed interval, because $B_t^0[k]$ is u.h.c. □

Proof of Lemma 4. For any solution of Program **P1**, we know from Lemma 6 that there exists an equivalent envelope tax function. From Lemma 2ii) we know that the corresponding household value function $v_t(k)$ satisfies $\lim_{t\to\infty} \beta^t \bar{\bar{K}}_t = 0$. From Lemma 7 we know that it satisfies the first order conditions Equs. (13)–(15). This shows that there exists an equivalent solution of Program **P2**.

Now take any solution of Program **P2**. From Lemma 7 we know that the value function satisfies the HJB equations (40) and (41). From the condition $\lim_{t\to\infty} \beta^t \bar{\bar{K}}_t = 0$ we know that the HJB equations are sufficient for the solution of the household problem (Stokey and Lucas 1989, Theorem 9.2). Any selection of the bequest correspondence then gives us a solution of Program **P1**. □

A.2 Stationary points of the Lagrangian

The Lagrangian approach of Section 3.3 is very intuitive, and the results we obtain seem to confirm its validity. Nevertheless, it is preferable to have a rigorous justification for the approach, that means, a proof of Lemma 5. For this, we have to show that the problem can be brought into the framework of Zeidler (1985, Ch. 48). Since a detailed proof would fill several pages, I ignore a number of technical issues and confine myself in the following to a sketch of a proof.

Outline of a proof of Lemma 5. Take a candidate solution $B_t[k], T_t(0), T_t'(b), \phi_t[k], v_t(\underline{z}), v_t'(k)$ to Program P2 such that $B_t[k]$ is continuously differentiable with $B_t'[k] > 0$ and (6) is satisfied.

Denote by $C^1[a, b; c]$ the Banach space of once continuously differentiable functions on the closed interval (a, b), with the norm $\|f\| \equiv \max_{a \le x \le b} f(x) + \frac{1}{c} \max_{a \le x \le b} f'(x)$. To derive necessary first order conditions, we consider variations

- $\delta B_t[k] \in C^1\left(k_t^*, \bar{k}; \min_{k \in \left(k_t^*, \bar{k}\right)} B_t'[k]\right)$; where k_t^* denotes the k such that $B_t[k] = 0$ for $k < k_t^*$ and $B_t[k] > 0$ for $k > k_t^*$. Notice that $\min_{k \in \left(k_t^*, \bar{k}\right)} B_t'[k] > 0$ is well defined, since a continuous function on a compact interval has a minimum.
- $\delta T_t(0) \in \Re$
- $\delta T_t'(b) \in C^1\left(0, B_t\left[\bar{K}_t\right], 1\right)$,
- $\delta \phi_t[k] \in C^1\left(\underline{z}, \bar{K}_t, 1\right)$
- $v_t(\underline{z}) \in \Re$
- $v_t'(k)) \in C^1\left(\underline{z}, \bar{K}_t, 1\right)$

Each of the spaces in which the variations live is a Banach space. Denote the product of the above spaces by $\mathcal{X}_t$, which is again a Banach space under the sum norm, which we denote by $|X|_t$.

We then define our space of variations $\mathcal{X}$ as

$$\mathcal{X} \equiv \left\{ X \in \prod_{t=0,\ldots} \mathcal{X}_t : \|X\| \equiv \sum_{i=0}^{\infty} |X_t|_t < \infty \right\} \tag{47}$$

This is again a Banach space under the norm $\|X\|$. With this, we satisfy the main requirements of Zeidler (1985, Theorem 48B), which shows the existence of Lagrange multipliers for our problem. [I have ignored here the questions of whether the Frechet derivative of the constraints has a closed range, and whether a solution is necessarily a "regular point" where the Lagrange multiplier λ_0 of the objective function is strictly positive. The numerical solutions strongly suggest that the problem is well-behaved in this sense.] The theorem gives us the Lagrange multiplier as a linear functional; its integral representation is in general of the form of a Stieltjes integral like $\int \phi(k)\, \mathbf{d}\Lambda(k)$. Writing it as such, and considering all possible variations of $\phi_t[k]$ it becomes clear that the function $\Lambda(k)$ cannot have discontinuities, and we can write the integral as $\int \phi(k)\lambda\,\mathbf{d}k$ as required in the lemma. (Loosely speaking, this is shown by the fact that the Lagrange multipliers $\lambda_t[k]$ and $\mu_t[k]$ are finite in the optimal solution, cf. remarks after Equ. (59), and Lemma 8i.) $\square$

One should note that in all the calculations we consider only variations in bequests that leave the critical point k_t^* unchanged. For $k < k_t^*$, we set the bequests to 0. In this way, the inequality constraints are effectively removed from the variational problem. It turns out that one need not vary k_t^* to derive the necessary first order conditions that we use.

B Proofs for Section 4

B.1 Proof of Proposition 1

For Proposition 1, we have to characterize a stationary point of the Lagrangian (16). In order to differentiate the Lagrangian w.r.t. parameters at time t, it is useful to collect all terms in (16) that refer to a fixed t, divide them by β^t and apply some transformations:

$$\begin{aligned}
\mathcal{L}/\beta^t = \ldots + \int_{\Re} &\phi_t[k]\,U(C_t[k]) + \phi_t[k]\,\xi_t[k]B_t[k] + \phi_t[k]\,\zeta_t[k]\left(-v_t'(k) + U'(C_t[k])\right) \\
&+ \phi_t[k]\,\mu_t[k]\left(-U'(C_t[k]) + \beta\bar{R}_t[k]\int_{\Re} v_{t+1}'(j)\,\pi(j - B_t[k])\,\mathbf{d}j\right)\mathbf{d}k \\
&+ \nu_t\left(T_t(0) + T_t^{tot}\right) - \int_{\Re}\phi_t[k]\,\lambda_t[k]\,\mathbf{d}k + \beta\int_{\Re}\phi_t[k]\int_{\Re}\lambda_{t+1}[j]\pi\left(j - B_t[k]\right)\mathbf{d}j\,\mathbf{d}k \\
&+ \int_{\Re} v_t'(k)\int_{\Re}\phi_{t-1}[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}[j])\,\mathbf{d}j\,\mathbf{d}k \\
&+ \chi_t\left(-v_t(\underline{z}) + U(C_t[\underline{z}])) + \beta\int_{\Re} v_{t+1}(B_t[\underline{z}] + z)\pi(z)\mathbf{d}z\right) \\
&+ \chi_{t-1}\left(v_t(\underline{z}) + \int_{\Re}\int_{\underline{z}}^{B_{t-1}[\underline{z}]+z} v_t'(j)\,\mathbf{d}j\pi(z)\mathbf{d}z\right) + \ldots
\end{aligned} \tag{48}$$

In (48), total taxes T_t^{tot} are an abbreviation for

$$T_t^{tot} = \int_{\Re}\phi_t[k]\int_0^k T_t'(B_t[j])\,B_t'[j]\,\mathbf{d}j\,\mathbf{d}k \tag{49}$$

For the term in the second line of (48) we have used

$$\int_{\Re} U'(C_{t+1}[B_t[k] + z])\,\pi(z)\mathbf{d}z = \int_{\Re} U'(C_{t+1}[j])\,\pi(j - B_t[k])\,\mathbf{d}j \tag{50}$$

In the third line we have used

$$\begin{aligned}
\int_{\Re}\lambda_{t+1}[k]\int_{\Re}\phi_t[j]\,\pi(k - B_t[j])\,\mathbf{d}j\,\mathbf{d}k &= \int_{\Re}\int_{\Re}\phi_t[j]\,\lambda_{t+1}[k]\pi(k - B_t[j])\,\mathbf{d}k\,\mathbf{d}j \\
&= \int_{\Re}\phi_t[k]\int_{\Re}\lambda_{t+1}[j]\pi(j - B_t[k])\,\mathbf{d}j\,\mathbf{d}k
\end{aligned}$$

The forth line has used

$$\int_{\Re} \phi_{t-1}[k]\,\mu_{t-1}[k]\bar{R}_{t-1}[k] \int_{\Re} U'\left(C_t\left[B_{t-1}[k]+z\right]\right)\pi(z)\mathbf{d}z\,\mathbf{d}k$$
$$= \int_{\Re} U'\left(C_t[k]\right) \int_{\Re} \phi_{t-1}[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k-B_{t-1}[j])\,\mathbf{d}j\,\mathbf{d}k \quad (51)$$

which is an application of the transformation

$$\int_{\Re} f[k] \int_{\Re} g[B_{t-1}[k]+z]\pi(z)\,\mathbf{d}z\,\mathbf{d}k = \int_{\Re} f[k] \int_{\Re} g[w]\pi\left(w-B_{t-1}[k]\right)\,\mathbf{d}w\,\mathbf{d}k$$
$$= \int_{\Re} g[w] \int_{\Re} f[k]\pi\left(w-B_{t-1}[k]\right)\,\mathbf{d}k\,\mathbf{d}w = \int_{\Re} g[k] \int_{\Re} f[j]\pi\left(k-B_{t-1}[j]\right)\,\mathbf{d}j\,\mathbf{d}k \quad (52)$$

which holds for any continuous functions f and g, where the first equality is a variable substitution, the second applies Fubini's theorem (which holds because we work with bounded functions, cf. Lemma 8), and the third equality is simply a renaming of variables. The last line has used (10).

For the following derivations, it is convenient to list formulas for some frequently used derivatives, which are obtained from (19) and (17):

$$\frac{\partial C_t[k]}{\partial k} = 1 - B_t'[k]\left(T_t'\left(B_t[k]\right) + \frac{1}{1+r}\right) = 1 - B_t'[k]\,\bar{R}_t^{-1}[k] \quad (53)$$
$$\frac{\partial C_t[k]}{\partial B_t[k]} = -T_t'\left(B_t[k]\right) - \frac{1}{1+r} = -\bar{R}_t^{-1}[k] \quad (54)$$
$$\frac{\partial B_t[k]}{\partial C_t[k]} = -\frac{1+r}{1+(1+r)T_t'\left(B_t[k]\right)} = -\bar{R}_t[k] \quad (55)$$
$$\frac{\partial \bar{R}_t[k]}{\partial B_t[k]} = -\frac{(1+r)^2 T_t''\left(B_t[k]\right)}{\left(1+(1+r)T_t'\left(B_t[k]\right)\right)^2} = -\bar{R}_t^2[k]T_t''\left(B_t[k]\right) \quad (56)$$

B.1.1 Deriving equations for the Lagrange multipliers

Let us first handle the constraint (41). Since (41) is the only place in the Lagrangian that contains the variable $v_t(\underline{z})$, the latter can be chosen such that the constraint is met, and the Lagrange multiplier should be zero. To see this formally, take the derivative w.r.t. $v_t(\underline{z})$ in (16) (or in (48)) to obtain $\chi_t = \chi_{t-1}$, so that χ is constant over time. To see that it is really 0, take the derivative w.r.t. $B_t[\underline{z}]$ to obtain (note that, apart from (41), $B_t[\underline{z}]$ enters the Lagrangian only through integral terms that make no contribution to the derivative)

$$\chi_t\left(-\bar{R}_t[\underline{z}]U_t'[\underline{z}] + \beta \int_{\Re} v_{t+1}'(B_t[\underline{z}]+z)\pi(z)\mathbf{d}z\right) = 0 \quad (57)$$

The term in parentheses in (57) will be strictly negative whenever the non-negativity constraint (2c) is binding. So (57) can be met for all t only if $\chi_t = 0$. In the following, we can therefore ignore any terms in χ_t.

We now focus on the optimal choice of $\phi_t[k]$. A necessary condition for $\phi_t[k]$ to be optimal is that no feasible variation $\epsilon \cdot \delta\phi_t[k]$, where ϵ is understood to be a small real number, improves the objective function. Inserting $\phi_t[k] + \epsilon \cdot \delta\phi_t[k]$ into (48), the derivative w.r.t. ϵ at $\epsilon = 0$ must be zero. Using the Kuhn-Tucker conditions (22), this gives

$$\int_{\Re} \delta\phi_t[k] \left(U(C_t[k]) + \nu_t T_t(B_t[k]) - \lambda_t[k] + \int_{\Re} \beta\lambda_{t+1}[j]\pi(j - B_t[k])\, \mathbf{d}j \right) \mathbf{d}k = 0 \tag{58}$$

The standard argument from the calculus of variations (see, e.g., Kamien and Schwartz, 1991, p. 16) shows that this can be fulfilled *for all feasible variations* only if the term in parentheses is zero, which gives

$$\lambda_t[k] = U(C_t[k]) + \nu_t T_t(B_t[k]) + \beta \int_{\Re} \lambda_{t+1}[j]\pi(j - B_t[k])\, \mathbf{d}j \tag{59}$$

Since $\beta < 1$ and $\pi(.)$ is a probability density, we can iterate the rhs of (59) forward and obtain a unique bounded solution whenever $\sup_k |U(C_t[k]) + \nu_t T_t(B_t[k])|$ grows in t at a rate smaller than $1/\beta$. Furthermore, since π is differentiable, it is clear that the rhs of (59) is differentiable w.r.t. k at all points where $B[k]$ is differentiable, therefore $\lambda_t[k]$ is differentiable w.r.t. k at these points. From the definition of a smooth interior solution, this is true at all points where $B_t[k] > 0$. Differentiating (59) and using (53) as well as

$$\int_{\Re} \lambda_{t+1}[j]\pi'(j - B_t[k])\, \mathbf{d}j = -\int_{\Re} \lambda'_{t+1}[j]\pi(j - B_t[k])\, \mathbf{d}j \tag{60}$$

we then obtain at the points of differentiability

$$\lambda'_t[k] = U'_t[k]\left(1 - B'_t[k]\,\bar{R}_t^{-1}[k]\right) + \nu_t T'_t(B_t[k])\, B'_t[k] + B'_t[k] \int_{\Re} \beta\lambda'_{t+1}[j]\pi(j - B_t[k])\, \mathbf{d}j \tag{61}$$

Since

$$B'_t[k]\, U'_t[k]\bar{R}_t^{-1}[k] = \beta B'_t[k] \int_{\Re} U'_{t+1}[j]\pi(j - B_t[k])\, \mathbf{d}j$$

we can transform (61) to

$$\lambda'_t[k] - U'_t[k] = \nu_t T'_t(B_t[k])\, B'_t[k] + \beta B'_t[k] \int_{\Re} \left(\lambda'_{t+1}[j] - U'_{t+1}[j]\right)\pi(j - B_t[k])\, \mathbf{d}j \tag{62}$$

Using the definition (29), Equ. (62) can be written compactly as (27).

Next we consider policy variations $\epsilon \cdot \delta v'_t(k)$. Inserting this into into (48) and differentiating w.r.t. ϵ at $\epsilon = 0$ we get

$$\int_{\Re} \delta v'_t(k) \left\{ -\phi_t[k]\,\zeta_t[k] + \int_{\Re} \phi_{t-1}[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}[j])\, \mathbf{d}j \right\} \mathbf{d}k = 0 \tag{63}$$

Again, this holds for all feasible variations only if

$$\phi_t[k]\,\zeta_t[k] = \int_{\Re} \phi_{t-1}[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}[j])\, \mathbf{d}j = 0 \tag{64}$$

We now turn to the variation $\delta B_t\,[k]$. Applying (53) and (54), and using (14) and (64), we obtain

$$\begin{aligned}\int_{\Re} \delta B_t\,[k] \Bigg\{ &-\phi_t\,[k]\,\bar{R}_t^{-1}[k]U_t'[k] + \phi_t\,[k]\,\mu_t[k]\bar{R}_t^{-1}[k]U_t''[k] \\ &- \beta\phi_t\,[k]\,\mu_t[k]\bar{R}_t[k] \int_{\Re} U_{t+1}'[j]\pi'(j - B_t\,[k])\,\mathbf{d}j \\ &- \beta\phi_t\,[k]\,\mu_t[k]\bar{R}_t^2[k]T_t''\,(B_t\,[k]) \int_{\Re} U_{t+1}'[j]\pi(j - B_t\,[k])\,\mathbf{d}j + \phi_t\,[k]\,\xi_t[k] + \nu_t\phi_t\,[k]\,T_t'\,(B_t\,[k]) \\ &- \bar{R}_t^{-1}[k]U_t''[k] \int_{\Re} \phi_{t-1}\,[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}\,[j])\,\mathbf{d}j \\ &- \phi_t\,[k] \int_{\Re} \beta\lambda_{t+1}[j]\pi'\,(j - B_t\,[k])\,\mathbf{d}j \Bigg\}\,\mathbf{d}k = 0 \end{aligned} \tag{65}$$

By the same argument as above we get

$$\begin{aligned} &- \phi_t\,[k]\,\bar{R}_t^{-1}[k]U_t'[k] + \phi_t\,[k]\,\mu_t[k]\bar{R}_t^{-1}[k]U_t''[k] \\ &- \beta\phi_t\,[k]\,\mu_t[k]\bar{R}_t[k] \int_{\Re} U_{t+1}'[j]\pi'(j - B_t\,[k])\,\mathbf{d}j \\ &- \beta\phi_t\,[k]\,\mu_t[k]\bar{R}_t^2[k]T_t''\,(B_t\,[k]) \int_{\Re} U_{t+1}'[j]\pi(j - B_t\,[k])\,\mathbf{d}j + \phi_t\,[k]\,\xi_t[k] + \nu_t\phi_t\,[k]\,T_t'\,(B_t\,[k]) \\ &- \bar{R}_t^{-1}[k]U_t''[k] \int_{\Re} \phi_{t-1}\,[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}\,[j])\,\mathbf{d}j \\ &- \phi_t\,[k] \int_{\Re} \beta\lambda_{t+1}[j]\pi'\,(j - B_t\,[k])\,\mathbf{d}j = 0 \end{aligned} \tag{66}$$

To obtain an expression for $\mu[k]$, we look at those k where $B\,[k] > 0$. Then $\xi[k] = 0$. To eliminate the integrals, we differentiate the household first order constraint w.r.t. k, which is binding whenever $\mu[k]$ is nonzero. This implies (use (54)–(56))

$$\begin{aligned}\phi_t\,[k]\,\mu_t[k] \Bigg\{ &U_t''[k]\left(1 - B_t'\,[k]\,\bar{R}_t^{-1}[k]\right) + \beta\bar{R}_t^2[k]T_t''[k]B_t'\,[k] \int_{\Re} U_{t+1}'[j]\pi(j - B_t\,[k])\,\mathbf{d}j \\ &+ \beta B_t'\,[k]\,\bar{R}_t[k] \int_{\Re} U_{t+1}'[j]\pi'(j - B_t\,[k])\,\mathbf{d}j \Bigg\} = 0 \end{aligned} \tag{67}$$

Now we multiply (66) by $B'\,[k]$ and subtract (67), which gives

$$\begin{aligned} &\phi_t\,[k]\,\mu_t[k]U_t''[k] - \phi_t\,[k]\,\bar{R}_t^{-1}[k]U_t'[k]B_t'\,[k] + \nu_t\phi_t\,[k]\,T_t'\,(B_t\,[k])\,B_t'\,[k] \\ &- B_t'\,[k]\,U_t''[k]\bar{R}_t^{-1}[k] \int_{\Re} \phi_{t-1}\,[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}\,[j])\,\mathbf{d}j \\ &- \phi_t\,[k]\,B_t'\,[k] \int_{\Re} \beta\lambda_{t+1}[j]\pi'\,(j - B_t\,[k])\,\mathbf{d}j = 0 \end{aligned} \tag{68}$$

Multiplying (61) by $\phi\,[k]$ and subtracting from (68), using (60), we get

$$\phi_t\,[k]\,\mu_t[k]U_t''[k] = B_t'\,[k]\,U_t''[k]\bar{R}_t^{-1}[k]\int_{\Re}\phi_{t-1}\,[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}\,[j])\,\mathbf{d}j \\ - \phi_t\,[k]\,(\lambda_t'[k] - U_t'[k]) \quad (69)$$

Slight reordering gives Equ. (26).

B.1.2 First order conditions of the government problem

Differentiating the Lagrangian (16) w.r.t. $T_t(0)$ and using (64), we obtain

$$\int_{\Re}\phi_t\,[k]\,\{U_t'[k] - \mu_t[k]U_t''[k]\} + U_t''[k]\int_{\Re}\phi_{t-1}\,[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}\,[j])\,\mathbf{d}j\,\mathbf{d}k = \nu_t \quad (70)$$

Inserting (26) into (25) and using (53), we see that $W_t[k]$ can be written as

$$W_t[k] \equiv U_t'[k] - \mu_t[k]U_t''[k] + \frac{U_t''[k]}{\phi_t\,[k]}\int_{\Re}\phi_{t-1}\,[j]\,\mu_{t-1}[j]\bar{R}_{t-1}[j]\pi(k - B_{t-1}\,[j])\,\mathbf{d}j \quad (71)$$

Using (71), (70) simplifies to (23).

Next we are going to consider variations of the form $T_t'\,[k] + \epsilon\delta T_t'\,[k]$. We then need the following derivatives, which are derived from (20), (49) and (17):

$$\left.\frac{\partial C_t[k]}{\partial\epsilon}\right|_{\epsilon=0} = -\int_0^k \delta T_t'\,[j]\,B_t'\,[j]\,\mathbf{d}j$$
$$\left.\frac{\partial T_t^{tot}}{\partial\epsilon}\right|_{\epsilon=0} = \int_{\Re}\phi_t\,[k]\int_0^k \delta T_t'\,[j]\,B_t'\,[j]\,\mathbf{d}j\,\mathbf{d}k$$
$$\left.\frac{\partial\bar{R}_k[t]}{\partial\epsilon}\right|_{\epsilon=0} = \frac{-(1+r)^2\delta T_t'\,[k]}{(1+(1+r)T_t'\,[k])^2}$$

Differentiating the Lagrangian w.r.t. ϵ and using (64) then gives

$$\int_{\Re}\phi_t\,[k]\left\{(-U_t'[k] + \mu_t[k]U_t''[k])\int_0^k \delta T_t'\,[j]\,B_t'\,[j]\,\mathbf{d}j - \mu_t[k]\frac{(1+r)\delta T_t'\,[k]}{1+(1+r)T_t'\,[k]}U_t'[k]\right\}\mathbf{d}k \\ - \int_{\Re}U_t''[k]\int_{\Re}\phi_{t-1}\,[l]\,\mu_{t-1}[l]\bar{R}_{t-1}[l]\pi(k - B_{t-1}\,[l])\,\mathbf{d}l\int_0^k \delta T_t'\,[j]\,B_t'\,[j]\,\mathbf{d}j\,\mathbf{d}k \\ + \nu_t\int_{\Re}\phi_t\,[k]\int_0^k \delta T_t'\,[j]\,B_t'\,[j]\,\mathbf{d}j\,\mathbf{d}k = 0 \quad (72)$$

For the fourth term, we used again (21). Using (71), (72) simplifies to

$$\int_{\Re}\phi_t\,[k]\left\{(\nu_t - W_t[k])\int_0^k \delta T_t'\,[j]\,B_t'\,[j]\,\mathbf{d}j - \delta T_t'\,[k]\,\mu_t[k]\bar{R}_k[t]U_t'[k]\right\}\mathbf{d}k = 0 \quad (73)$$

Applying the transformation

$$\int_{\Re}f[k]\int_0^k g[j]\,\mathbf{d}j\,\mathbf{d}k = \int_{\Re}g[j]\int_j^{\infty}f[k]\,\mathbf{d}k\,\mathbf{d}j = \int_{\Re}g[k]\int_k^{\infty}f[j]\,\mathbf{d}j\,\mathbf{d}k \quad (74)$$

to (72), we obtain

$$\int_{\Re} \delta T_t'[k] \left\{ B_t'[k] \int_k^{\infty} (\phi_t[j] (\nu_t - W_t[j])) \,\mathbf{d}j - \phi_t[k] \mu_t[k] \bar{R}_t[k] U_t'[k] \right\} \mathbf{d}k = 0 \tag{75}$$

Again, this holds for all feasible $\delta T'[k]$ only if (24) holds for all k.

This completes the proof of Proposition 1.

B.2 Derivation of Formula (30)

Using (31), the definition of the cross-sectional distribution $\Phi_K(k) \equiv \int_0^k \phi[j]\,\mathbf{d}j$ and the fact that $\nu_t = Ave_t^0(W)$ (cf. (23)), we can rewrite (24) as

$$Ave_t^0(W) = Ave_t^k(W) + \frac{\phi_t[k] \mu_t[k] \bar{R}_t[k] U_t'[k]}{B_t'[k] (1 - \Phi_K(k))} \tag{76}$$

Equ. (30) can be obtained by rearranging (76), using the definition (17), Equ. (33), the identity $\phi_B(\mathcal{B}_t(k)) = \phi_t[k] / B_t'[k]$ which links the cross-sectional densities $\phi_t[k]$ of capital and $\phi_B(\mathcal{B}_t(k))$ of bequests, and finally using

$$\frac{U_t'[k]}{U_t''[k]} = -\eta_R^{B^s}[t,k] \frac{B_t[k]}{B_t'[k]} \tag{77}$$

Equ (77) is derived as follows. Differentiating the Euler equation (cf. Equ. (67)) and solving for $B_t'[k]$, we get

$$B_t'[k] = \frac{U_t''[k]}{U_t''[k] \bar{R}_t^{-1}[k] - \beta \bar{R}_t[k] \int_{\Re} U_t'[j] \pi'(j - B_t[k]) \,\mathbf{d}j - \bar{R}_t[k] T_t''[k] U_t'[k]} \tag{78}$$

Next we develop an expression for the substitution effect of interest rate changes on bequests. For this, consider a household with capital k which faces the after tax discount factor $\bar{R}_t[k]$. We will derive the change in bequests as a reaction to a change in $\bar{R}_t(B_t[k])$, keeping $\frac{d\bar{R}_t(B_t[k])}{dB}$ constant (that means we change the marginal tax rate $T'[k]$ while keeping $T''[k]$ fixed). Denote the household's chosen consumption-bequest bundle by $(\bar{C}, \bar{B})$. To isolate the substitution effect, the household obtains a Slutsky compensation, so that its budget constraint reads $C = \bar{C} - \bar{R}_t[k]^{-1}(B - \bar{B})$. Inserting this into the household Euler equation (21) (with equality) and taking total derivatives at point $(\bar{C}, \bar{B})$ we obtain

$$\begin{aligned} - U_t''[k] \bar{R}_t[k]^{-1} dB^s[k] &= -\beta \bar{R}_t[k] \int_{\Re} U_t'[j] \pi'(j - B_t[k]) \,\mathbf{d}j dB^s[k] \\ &+ \beta \frac{d\bar{R}_t(B_t[k])}{dB} \int_{\Re} U_t'[B_t[k] + z] \pi(z) \mathbf{d}z dB^s[k] + \beta \int_{\Re} U_t'[B_t[k] + z] \pi(z) \mathbf{d}z d\bar{R}[k] \end{aligned} \tag{79}$$

Using (56), and the Euler equation (21) with equality, we obtain from (79)

$$\frac{dB^s[k]}{d\bar{R}[k]} = -\frac{\bar{R}_t[k]^{-1} U_t'[k]}{U_t''[k] \bar{R}_t[k]^{-1} - \beta \bar{R}_t[k] \int_{\Re} U_t'[j] \pi'(j - B_t[k]) \,\mathbf{d}j - \bar{R}_t[k] T_t''[k] U_t'[k]} \tag{80}$$

Defining $\eta_R^{B^s}[t,k] = \frac{dB^s[k]}{d\bar{R}[k]} \frac{\bar{R}_t[k]}{B_t[k]}$ and combining (78) and (80), we then obtain (77).

B.3 Proof of Proposition 2

Lemma 8. *In a smooth interior solution, for all t,*

i) the functions $C_t[k]$, $U_t'[k]$, $U_t''[k]$, $\bar{R}_t[k]$, $T'(B_t[k])$, $B_t'[k]$, $\lambda_t^*[k]$, $\lambda_t^*[k]/B_t'[k]$, $\mu_t[k]$ and $\mu_t[k]/B_t'[k]$ are bounded on the support of $\phi_t[k]$.

ii) there exists a constant $M_t > 0$ such that $U_t'[k] \geq M_t$, $|U_t''[k]| \geq M_t$ and $\bar{R}_t[k] \geq M_t$.

Proof. The existence of a lower bound $\underline{c}_t > 0$ and an upper bound $\bar{c}_t$ on consumption has been established in the proof of Lemma 2. The boundedness of $U_t'[k]$ and $U_t''[k]$ then follows from the fact that they are continuous on the compact interval $(\underline{c}_t, \bar{c}_t)$.

In a smooth interior solution, the household Euler equation holds with equality for all k such that $B_t[k] > 0$, which implies that $\bar{R}_t[k] \in \left(\beta^{-1}\frac{U'(\bar{c}_t)}{U'(\underline{c}_t)}, \beta^{-1}\frac{U'(\underline{c}_t)}{U'(\bar{c}_t)}\right)$. This shows the boundedness of $\bar{R}_t[k]$ and that of $T'(B_t[k])$, since $T'(B_t[k]) = \bar{R}_t^{-1}[k] - 1/(1+r)$, cf. Equ. (17).

To show the boundedness of $\lambda_t^*[k]$, consider Equ. (27). Because of the properties of $\pi(z)$ and the continuity of $B_t[k]$, the integral on the rhs of (27) is continuous in k, and therefore bounded on the compact support of $\phi_t[k]$. This implies the boundedness of $\lambda_t^*[k]$, because all other functions on the rhs of (27) have been shown to be bounded. Similarly, we obtain the boundedness of $\lambda_t^*[k]/B_t'[k]$ by dividing the rhs of (27) by $B_t'[k]$.

For $\mu_t[k]$, we consider only the k where $\phi_t[k] > 0$ ($\mu_t[k]$ is not pinned down by the first order conditions at points with zero density). Then divide (26) by $\phi_t[k]$ and use (8) to obtain

$$\mu_t[k] = -\frac{\lambda_t^*[k]}{U_t''[k]} + \frac{B_t'[k]}{\bar{R}_t[k]} \frac{\int_{\Re} \phi_{t-1}[j]\, \bar{R}_{t-1}[j]\mu_{t-1}[j]\pi(k - B_{t-1}[j])\, \mathbf{d}j}{\int_{\Re} \phi_{t-1}[j]\, \pi(k - B_{t-1}[j])\, \mathbf{d}j} \tag{81}$$

We know that $\mu_{-1}[k]$ is bounded because it is identically zero. Having established the boundedness of $\mu_{t-1}[k]$, (81) shows the boundedness of $\mu_t[k]$: the fraction is bounded since $\phi_t[k]$ is nonnegative, and $\bar{R}_{t-1}[j]\mu_{t-1}[j]$ is bounded. The other functions on the rhs of (81) are bounded, and $U_t''[k]$ and $\bar{R}_t[k]$ are bounded away from zero.

Similarly, we obtain the boundedness of $\mu_t[k]/B_t'[k]$ dividing the rhs of (81) by $B_t'[k]$ and using Equ. (27).

Part ii) of the lemma then follows from the boundedness of the relevant functions and the fact that $U_t'[k]$ and $\bar{R}_t[k]$ are strictly positive and $U_t''[k]$ is strictly negative (cf. Assumption 3i). □

Proof of Proposition 2. 1) We first show that $\phi_t[k]$ is non-increasing in a left neighborhood of $\bar{K}_t$. Since $\bar{K}_{t-1}$ is the upper bound of the support in $t-1$, we can write (8) as $\phi_t[k] = \int_{-\infty}^{\bar{K}_{t-1}} \phi_{t-1}[j]\, \pi(k - B_{t-1}[j])\, \mathbf{d}j$. Differentiating w.r.t. k we obtain

$$\phi_t'[k] = \int_{-\infty}^{\bar{K}_{t-1}} \phi_{t-1}[j]\, \pi'(k - B_{t-1}[j])\, \mathbf{d}j \leq 0 \tag{82}$$

Since $B_{t-1}[j] \leq \bar{K}_t - \overline{z}$ for all j in the support of $\phi_t[k]$, we see that $k - B_{t-1}[j] \geq \overline{z} - \epsilon$ for any $\epsilon > 0$ and $k \in (\bar{K}_t - \epsilon, \bar{K}_t)$. From Assumption 2iv), there is an $\epsilon > 0$ s.t. $\pi'(z) \leq 0$ for $z \in (\overline{z} - \epsilon, \overline{z})$. The claim then follows from (82).

2) Next we show that

$$\lim_{k \to \bar{K}_t} \frac{1}{\phi_t[k]} \int_k^{\bar{K}_t} \phi_t[j] f(k,j) \, \mathbf{d}j = 0 \tag{83}$$

for any bounded function $f(k,j)$. Take any M such that $|f(k,j)| \leq M$ and define ϵ as above, then

$$\frac{1}{\phi_t[k]} \left| \int_k^{\bar{K}_t} \phi_t[j] f(k,j) \, \mathbf{d}j \right| \leq \frac{1}{\phi_t[k]} \int_k^{\bar{K}_t} \phi_t[k] \, |f(k,j)| \, \mathbf{d}j \leq M(\bar{K}_t - k) \tag{84}$$

for all $k \geq \bar{K}_t - \epsilon$, and the claim follows.

3) Applying (69) to (71), we obtain the following representation for $W_t[k]$:

$$W_t[k] = U_t'[k] - \mu_t[k] U_t''[k] + \frac{\bar{R}_t[k]}{B_t'[k]} \left(\mu_t[k] U_t''[k] + \lambda_t^*[k] \right) \tag{85}$$

From Lemma 8, we see then that $W_t[k]$ is bounded on the support of $\phi_t[k]$.

4) Divide (24) by $\phi_t[k]$ and make the upper bound of integration $\bar{K}_t$ explicit. This gives

$$B_t'[k] \frac{1}{\phi_t[k]} \int_k^{\bar{K}_t} \phi_t[j] \left(\nu_t - W_t[j] \right) \mathbf{d}j = \mu_t[k] \bar{R}_t[k] U_t'[k] \tag{86}$$

The boundedness of $W_t[j]$ was shown in 3), and that of $B_t'[k]$ in Lemma 8. From 2) it then follows that the lhs of (86) goes to 0 as k goes to $\bar{K}_t$. On the rhs of (86), we see from Lemma 8ii) that $\bar{R}_t[k] U_t'[k]$ is bounded away from 0. Then the rhs of (86) can only go to zero if $\lim_{k \to \bar{K}_t} \mu_t[k] = 0$, which is what had to be shown (cf. Footnote 3).

□

C Computation

Both household bequests, as a function of k, and the marginal tax schedule are approximated by flexible functional forms. A natural choice would be Chebyshev polynomials. However, since the nonlinear tax function gives rise to somewhat irregular bequest functions, after some experimentation I found that the following ad-hoc construction worked better. For the bequest function, I use the parameters i) k_t^*, the level of capital where the non-negativity constraint starts to bind; ii) $B(k_i)$ at $(n_b - 1)$ different node points, where the nodes are endogenously determined through k_t^* and the upper bound of the distribution. iii) the derivative $B'(k_i)$ at k_t^* and the node points. This makes $2n_b$ parameters. To these data I then a apply a piecewise quadratic interpolation, using a construction as in Schumaker's quadratic splines (cf. Judd, 1998, Section 6.11), but with the interior free node always at the mean of adjacent nodes. This construction turned out to be very good in giving a stable bequest function and guaranteeing that $B'(k_i) > 0$, which is essential for the calculation. I usually worked with 15 node points (including k_t^*), which means 30 free parameters in each period.

For the marginal tax function, I took a cubic spline, usually on 8 node points. Fixing the first derivative at both end points, this gives 10 parameters per period. The subsidy gives one additional parameter per period.

The distribution function and the Lagrange multipliers where defined on a grid of 400 points. All the integrals (in (8), (26), (27) etc. are computed by a simple midpoint rule on this grid. It was then not necessary to interpolate these functions.

An approximate solution of the model is found through the following nonlinear root finding problem. Given a set of parameters of the bequest functions, the tax functions, and the shadow value of tax revenues ν_t,

1. calculate the cross-sectional distributions by solving (8), starting from the given $\phi_0\,[k]$.

2. calculate the $\lambda_t^*[k]$ backwards using (27). Initialize $\lambda_T^*[k] = 0$. Then apply (27) ten times, and use the result as value of $\lambda_T^*[k]$. Then compute $\lambda_{T-1}^*[k]$ by (27) and so on.

3. calculate $\mu_t[k]$; for this, I was not using (26), but the following equivalent formulation:

$$\mu_t^*[k] = -\phi_t\,[k]\,\bar{R}_t[k]\frac{\lambda_t^*[k]}{U_t''[k]} + B_t'\,[k]\int_{\Re} \mu_{t-1}^*[j]\pi(k - B_{t-1}\,[j])\,\mathbf{d}j \tag{87}$$

   where we have defined

$$\mu_t^*[k] = \phi_t\,[k]\,\mu_t[k]\bar{R}_t[k] \tag{88}$$

   We start with the initialization $\mu_{-1}^*[k] = 0$.

4. Calculate the following residuals:

   - the residuals of the Euler equation $U'\left(C_t\,[k]\right) = \beta\bar{R}_t[k]\int_{\Re} U'\left(C_{t+1}\,[B_t\,[k] + z]\right)\pi(z)\mathbf{d}z$ and of Equ. (78) at the node points. To have as many residuals as parameters, we get a bequest function in $T+1$ by $B_{T+1}\,[k] = B_T\,[k]$, which is justified since we should be approximately in a steady state after 10 periods (means 300 years).
   - the residuals of the government first order conditions (23) and (75), where $\delta T_t'\,[k]$ is the derivative of $T_t'\,[k]$ w.r.t. any of the parameters of the representation of the tax function. (It is more natural and probably more accurate to use directly (75), rather than the residual of (24) at some node points).
   - The residual of the government budget constraint (4).

We replace the infinite horizon model by 10 periods (which conforms to 300 years). With the above approximation, we have a problem of 420 nonlinear equations in 420 unknowns. We have to find parameter values that make these residuals equal to zero. This was a difficult task; computing the set of residuals once is already very costly, due to the integral equations that determine the distribution and the Lagrange multipliers. The task was solved by Broyden's algorithm (Press, Flannery, Teukolsky and Vetterling 1986, Section 9.7), in combination with the use of continuation methods (Judd 1998, Section 5.8).

To check accuracy, I sometimes increased the number of parameters in the approximations. For the bequest function, I was using up to 50 parameters per period, and I increased the number of the nodes in the quadrature grid to 800. The numerical results then change slightly, but the conclusions that we reach in the text were never changed. It should be clear, however, that we are not talking about a high-precision solution here. So far I did not increase the number of tax parameters beyond 11. The accuracy of this approximation is discussed in the text.

References

Aiyagari, S. R.: 1995, Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting, *Journal of Political Economy* **103**(6), 1158–75.

Albanesi, S. and Sleet, C.: 2004, Dynamic optimal taxation with private information, mimeo.

Atkinson, A. B.: 1995, *Public Economics in Action: The Basic Income/Flat Tax Proposal*, Clarendon Press, Oxford.

Chamley, C.: 1986, Optimal taxation of capital income in general equilibrium with infinite lives, *Econometrica* **54**(3), 607–22.

Chamley, C.: 2001, Capital income taxation, wealth distribution and borrowing constraints, *Journal of Public Economics* **79**(1), 55–69.

Conesa, J. C. and Krueger, D.: 2002, On the optimal progressivity of the income tax code, mimeo.

Correia, I. H.: 1996, Should capital income be taxed in the steady state?, *Journal of Public Economics* **60**(1), 147–51.

Costa, C. d. and Werning, I.: 2001, On the optimality of the Friedman rule with heterogeneous agents and non-linear income taxation, University of Chicago.

Cremer, H. and Pestieau, P.: 2001, Non-linear taxation of bequests, equal sharing rules and the tradeoff between intra- and inter-family inequalities, *Journal of Public Economics* **79**(1), 35–53.

Diamond, P. A.: 1998, Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates, *American Economic Review* **88**(1), 83–95.

Diaz-Gimenez, J., Quadrini, V. and Rios-Rull, J.-V.: 1997, Dimensions of inequality: Facts on the u.s. distributions of earnings,income, and wealth, *Federal Reserve Bank of Minneapolis Quarterly Review* **21**(2), 3–21.

Domeij, D. and Heathcote, J.: 2000, Factor taxation with heterogeneous agents, Stockholm School of Economics.

Erosa, A. and Gervais, M.: 2002, Optimal taxation in life-cycle economies, *Journal of Economic Theory* **105**(2), 338–69.

Foran, J.: 1991, *Fundamentals of Real Analysis*, Marcel Dekker, New York.

Garcia-Mila, T., Marcet, A. and Ventura, E.: 1995, Supply side interventions and redistribution, Universitat Pompeu Fabra, Economics Working Paper 115.

Golosov, M., Kocherlakota, N. and Tsyvinski, A.: 2003, Optimal indirect and capital taxation, *Review of Economic Studies* **70**, 569–587.

Haider, S. J.: 2001, Earnings instability and earnings inequality of males in the United States: 1967-1991, *Journal of Labor Economics* **19**(4), 799–836.

Jones, L. E., Manuelli, R. E. and Rossi, P. E.: 1997, On the optimal taxation of capital income, *Journal of Economic Theory* **73**(1), 93–117.

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

Judd, K. L.: 1998, *Numerical Methods in Economics.*, MIT Press, Cambridge and London.

Kamien, M. I. and Schwartz, N. L.: 1991, *Dynamic Optimization*, 2nd edn, North Holland, New York.

Kapicka, M.: 2002, Optimal income taxation and human capital accumulation, University of Chicago.

Klein, P. and Ríos-Rull, J.-V.: 2001, Time-consistent optimal fiscal policy, mimeo.

Kocherlakota, N. R.: 2004, Zero expected wealth taxes: A mirrlees approach to dynamic optimal taxation, Federal Reserve Bank of Minneapolis Sta .Report.

Krueger, D. and Perri, F.: 1999, Risk sharing: Private insurance markets or redistributive taxes?, Federal Reserve Bank of Minneapolis.

Krusell, P., Quadrini, V. and Rios-Rull, J.-V.: 1996, Are consumption taxes really better than income taxes?, *Journal of Monetary Economics* **37**(3), 475–503.

Lucas, Robert E., J.: 1990, Supply-side economics: An analytical review, *Oxford Economic Papers, N. S.* **42**(2), 293–316.

Marcet, A. and Marimon, R.: 1998, Recursive contracts, UPF.

Menchik, P. L. and David, M.: 1983, Income distribution, lifetime savings, and bequests, *American Economic Review* **73**(4), 672–90.

Mirrlees, J. A.: 1971, An exploration in the theory of optimum income taxation, *Review of Economic Studies* **38**, 175–208.

Mirrlees, J. A.: 1986, The theory of optimal taxation, *in* K. J. Arrow and M. D. Intriligator (eds), *Handbook of Mathematical Economics*, North-Holland, Amsterdam.

Myles, G. D.: 1995, *Public Economics*, Cambridge University Press.

Press, W., Flannery, B., Teukolsky, S. and Vetterling, W.: 1986, *Numerical Recipes*, Cambridge University Press.

Saez, E.: 2001, Using elasticities to derive optimal income tax rates, *Review of Economic Studies*.

Saez, E.: 2002, Optimal progressive capital income taxes in the infinite horizon model, NBER Working Paper 9046.

Stokey, N. L. and Lucas, R. E.: 1989, *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge, MA.

Yotsuzuka, T.: 1987, Ricardian equivalence in the presence of capital market imperfections, *Journal of Monetary Economics* **20**(2), 411–36.

Zeidler, E.: 1985, *Nonlinear Functional Analysis and its Applications III*, Springer, New York.