Optimal Estate Taxation in the Steady State

Wojciech Kopczuk*

June 21, 2001

JEL classification: H21, H23, H24, E6.

Abstract

The optimal income taxation problem is extended to analyze estate taxation. In the model, people have preferences over bequest, consumption and leisure, with a potential external aspect of bequests. The optimization is subject to the steady state constraint on the distribution of bequests, under a simplifying assumption of perfect correlation of abilities within the family line. The optimal taxes have a modular structure, with components reflecting redistribution, Pigouvian correction of bequest externality, and linear contributions due to pecuniary externalities acting through factor prices. Without a bequest externality, estate taxation is governed by commodity tax rules and in a special case it is not necessary. When the external effect is present, inheritance becomes a (partially) exogenous source of inequality and it should be treated analogously to the regular income implying that estate taxation is useful for redistribution. This effect is separate from the Pigouvian correction. I also demonstrate that the Atkinson-Stiglitz commodity taxation logic applies in the steady state even if it implies saving distortions. This result is related to the "no capital taxation" propositions known from the literature.

*Address: Department of Economics, University of British Columbia, #997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail: wojtek@umich.edu.

1 Introduction

The purpose of this paper is to consider the role of estate taxation in the optimal tax system, accounting for its redistributive role and the distortions it introduces. The key question is whether the estate tax is a part of the optimal tax mix, in particular whether it is necessary when the income tax may be optimally designed. To preview the main results: it might be, and for a number of reasons, including redistribution and correction of externalities. In particular, I demonstrate that the no-capital income taxation results of, e.g., Chamley (1986) and Judd (1985), cannot be used to justify no individual capital income distortions in a realistic public finance context that accounts for informational constraints and, additionally, there are even circumstances when aggregate distortions ought to be present as well.

Bequests are in an important aspect different than other goods: donor's consumption (of a gift) becomes donee's income. Consequently, the decision to bequeath determines how much inequality the next generation wil face. At the same time, bequests represent saving and thus their taxation can have long-run consequences. Capital accumulation and inequality are among the most important economic issues, so that it comes as no surprise that tax treatment of bequests is a controversial topic.

People favoring estate taxation often stress perils of inheritance: concentration of wealth and economic power, the Carnegie effect etc. Behind each of these, there is presumably some social cost that the donors do not account for. I will refer to this cost as an "externality". If bequests are a form of contracts between parents and children, they are efficient from the social point of view only as long as neither player has market power. Yet, the models of such contracts usually assume some form of strategic behavior, and, consequently, the resulting allocation is inefficient. These are negative external effects, but there are also arguments for a positive externality from bequests. For example, if people enjoy giving, a gift yields a double dividend: it is enjoyed by both the donor and the donee. In principle, the tax policy can be used to correct any such imperfection. Understanding the consequences of potential externalities would help to identify tax policies that are certainly not optimal even if one shares the view that such an externality exists and to identify the ones that could be optimal only if such an externality is present. Therefore, theory can systematize the debate about estate taxation by limiting the set of options under consideration and revealing assumptions that proponents of certain policies implicitly adhere to.

The mix of redistributive, dynamic and external effects makes the problem of determining the optimal tax treatment of bequests complicated. Many questions ought to be addressed in this context. What is the appropriate objective of the government in the context involving multiple generations? How should it deal with inequality of wealth caused by inheritances? Does redistribution necessarily require estate taxation or is income taxation enough? Should we worry about distorting wealth accumulation and, if so, how does it interact with redistributional objectives? Are Pigouvian taxes feasible and if not, how should the externalities be addressed? Do negative externalities necessarily imply positive marginal estate tax rates? What are the incentive problems involved and how do they interact with externalities — do they impede or facilitate redistribution? Can correcting externalities become more important than reducing inequality? Finally, what does it all tell us about the actual tax system?

These questions are addressed in the model developed in this paper. There are two major simplifying assumptions maintained throughout. First, skill levels in a given family line are assumed to be perfectly correlated so that, in the steady state, children look exactly like their parents. Second, only the long-run is considered and the government is assumed to maximize the welfare of a "representative" generation. In particular, it implies that only the question of the optimal redistribution within a generation is considered and that intergenerational redistribution is ignored. The model is still rich enough to account for major considerations. It builds on the optimal taxation model of Mirrlees (1971), therefore it allows for analyzing the inherent efficiency-equity tradeoff. Wealth inequality is endogenous and in a sense extreme: wealth is directly related to the skill level. The steady state approach makes it possible to analyze capital distortions, questions of dynamic efficiency and the role played by factor prices. My specification of the utility function has as its special cases models known from the literature, in particular the altruistic model and the bequests-in-utility model, as well as it allows for the presence of externalities. Finally, the information structure and associated incentive problems are taken seriously.

The main point of reference for the results in the paper is the optimal indirect taxation theorem of Atkinson and Stiglitz (1976). In a static economy, when skills are the only source of inequality, under mild assumptions (the utility function being separable between leisure and other commodities), labor income taxation is sufficient for redistribution and relative prices of other goods should not be distorted. As Kaplow (2001) stresses, if bequests are like other goods, then there should be no estate taxation. The framework of Kaplow (2001) is, however, completely static and it does not feature inequality other than that of ability levels. In particular, there is no inequality of wealth. Observe, however, that in the steady state, as long as social and private returns to bequeathing are equal, bequests are like other goods even though they generate wealth inequality. When people have identical preferences, as I assume in this paper, wealth inequality arises only as a consequence of inequalities in the skill levels. Under the strong assumption of perfect intergenerational correlation of skills, inequalities over many generations reduce to inequality of individuals' *own* skill levels. Without any externality present, for the steady-state individual, bequests that one intends to leave yield the same benefit as the inheritance that one receives. Effectively then, there is no difference between bequest and other commodities and therefore intergenerational transfers are not an exogenous source of inequality. Hence, their tax treatment should follow the rules analogous to those that govern tax treatment of other commodities.

When private and social benefits to bequeathing are different, bequests are no longer like other commodities. The Pigouvian taxes would do the job of correcting such an imperfection, however, they are unlikely to be feasible. Many types of potential externalities from bequests are person specific because they arise as consequences of individual actions, not the aggregated action of the whole society. For example, when one worries about the concentration of economic power, a huge bequest received by a single person has vastly different consequences than the same amount divided between many individuals. It implies that the strength of an external effect usually varies by the skill level. In principle then, the first-best solution would require the ability to vary the Pigouvian taxes by the unobservable skill level. This problem is analogous to the optimal income taxation problem and, similarly, the first-best correction is often informationally infeasible. Therefore, in the second-best optimum, differences between private and social benefits from bequeathing have to persist. By itself, it implies that in the steady state bequests retain their external aspect. In this way, they are similar to labor income which is affected in part by the, externally given, skill level; in fact this is what distinguishes labor income from other commodities and is responsible for its special role in the Atkinson and Stiglitz (1976) solution.

I characterize the optimal income and estate taxation in this context. The optimal tax formulae reflect the symmetry in the role of labor income and bequests: each of the optimal taxes plays a redistributive role, but this happens only to the extent that there is an external aspect of the corresponding source of inequality. This is always so with labor income, but it remains true for bequests only if some sort of externality from bequeathing is present. If the separability assumption identified by Atkinson and Stiglitz (1976) does not hold, each of these taxes additionally supports the other one in addressing inequality from its relevant source. This effect is analogous to commodity taxation. Finally, the taxes reflect a direct, Pigouvian-like, correction of externalities: both from bequeathing and pecuniary ones. All these effects add up, reflecting the, so called, “principle of targeting”.[1] Imperfections should be targeted directly, even if other tax distortions need to be present and as soon as imperfections are internalized the remaining problem becomes a standard second-best issue, because the imperfection no longer has to be dealt with.

The plan of this essay is as follows. In Section 2 I set up the model. In Section 3 the first-best optimal policy is characterized and I deal with the dynamic issues. The optimal taxation problem under imperfect information is the topic of Section 4. In this section, I derive the solutions for the optimal tax rates and I demonstrate some results about the direction of redistribution. Section 5 deals with the redistributive role of factor prices and other aggregate considerations. I discuss the relationship of my results to the no-capital income taxation theorem in Section 5.1. The last section concludes.

2 Framework

I assume that each generation consists of a continuum of individuals with a certain (exogenously given) distribution of skills. The skill level of α implies that each unit of labor that an individual supplies results in α efficiency units of labor. The wage rate per efficiency unit of labor is given by w. An individual can choose her amount of leisure, L. The total endowment of leisure is normalized to one, so that labor income is given by $\alpha w(1-L)$. In addition to labor income, each individual receives an (after-tax) inheritance of X from her parent. Skill levels and inheritances are the only source of differences between individuals. In particular, everybody shares the same preferences.

[1] The result was proven in a special case of linear commodity taxation by Sandmo (1975) and Dixit (1985). See Kopczuk (2000a) for a general statement.

2.1 Dynamics

The generations follow each other, and they do not overlap. Each individual has a single child who is the sole recipient of the inheritance. Furthermore, I assume an extreme form of correlation between skills of subsequent generations: all members of the same family line share the same skill level α.[2] As the result, a bequest is just a function of α and, in the steady state, the bequest that an individual receives is directly related to the bequest that he leaves. I also assume that there is no population growth, so that in the steady state it must be

$$B = X\,, \tag{1}$$

where B is the bequest left by the individual.

2.2 Individual and Social Preferences

Each individual lives a single period during which he consumes a part of the lifetime income (C) and appropriates the remainder to a bequest (B), so that his utility level is maximized. The budget constraint of the individual is

$$C + \frac{B}{1+r} + T = \alpha w(1-L) + X\,, \tag{2}$$

where w is wage rate per unit of effective labor, r is the interest rate, T are taxes paid and X is the bequest received.

[2] Although this assumption is certainly not true literally, there is some evidence that that the intergenerational correlation of incomes may be relatively high, of the order of 0.4 (Solon, 1992; Zimmerman, 1992). Allowing for the non-perfect correlation of incomes would add an additional dimension of differences between individuals: each generation would be characterized by a two-dimensional (bequest/skill) distribution, with one of the dimensions (bequests) endogenous. Apart from reducing the problem to just one dimension, it seems that the key insight that is foregone due to this assumption has to do with the informational content of bequests left by altruistic parents who might have some private knowledge about ability of their kids. This is of interest, but it seems unlikely to drive the decisions of the very rich because labor income of their children is likely to be minor compared to the size of inheritance.

The utility function is given by

$$u(C, B, L, X)\,. \tag{3}$$

The impact of X on the utility is not recognized by the individual, so that it represents an external aspect of bequests. The government is interested in maximizing the welfare of a given generation and accounts for the dependence on X in its objective function. It is useful to rewrite the utility function by substituting from the budget constraint for consumption,

$$u(X + \alpha w(1 - L) - \frac{B}{1 + r} - T, B, L, X)\,. \tag{4}$$

Note that X appears in this expression twice and each of these components represents the external effect, which is therefore given by $u_C + u_X$. The exposition will be simplified by introducing the following notation,

$$v(C, B, L, X) \equiv u(X + C, B, L, X)\,. \tag{5}$$

Here, C is consumption *minus* inheritance and the external effect is simply v_X.

In order to guarantee that the government has redistributive intentions, I assume a variant of concavity of the utility function:

Assumption 1 *Define $v^*(C, B, L) \equiv v(C, B, L, B)$. I assume that v^* is jointly concave in its arguments.*

Note that this is a weaker statement than $v(\cdot)$ or $u(\cdot)$ being concave in its arguments, but because the government recognizes the equivalence of B and X it implies that a reduction in inequality is valuable.

This specification has a number of special cases that may be interpreted as different types of bequest motives known from the literature:

- $u_X = 0$, $u_B > 0$. This is simply the joy-of-giving bequest motive of e.g., Yaari (1965) and Hurd (1989). An individual does not take into account the effect that a bequest has on the donee and thus the inheritance received is an additional source of utility not taken into account in the optimization process.

- $u_B = \mu u_C$ and $u_X + u_C = 0$ (i.e., $v_X = 0$). It corresponds to the Barro (1974) altruism with discounting at the rate of μ: the parent values consumption of the child, but discounts it at the rate of μ (reflected by μu_C). There is no external effect present. More explicitly, this utility function is given by $v(C, B, L, X) = h(C + \mu B, L)$. It needs to be stressed that this is only a reduced form of the utility function, valid only in the steady state with perfect correlation of abilities in the family. In what follows, only the case of $\mu = 1$ will correspond to the standard altruistic model. This is because, by assuming that only the steady state matters, I implicitly assume that the social discount rate is equal to one.

- $v_X < 0$. This is the case of a negative externality such as, e.g., due to concentration of wealth.

This specification ignores accidental bequests (Yaari, 1965; Davies, 1981; Kopczuk, 2000b) and strategic interactions present if bequests are contracts (e.g., Bernheim, Shleifer and Summers, 1985; Kotlikoff and Spivak, 1981). It also does not allow for the wealth-in-utility approach of Carroll (2000).

2.3 Production

The production side of the economy is characterized by the aggregate constant-returns-to-scale production function $F(K, N)$, where K is the aggregate capital stock and N is the aggregate labor supply. The only source of capital stock are estates. The amount of productive capital supplied by each individual is then, $\frac{B}{1+r}$. Augmented by the return on these assets, the actual bequest is B, as expected. The amount of productive labor provided by each individual is given by $\alpha(1 - L)$. The factor markets are competitive, so that $F_K = r + \delta$, $F_N = w$, where $0 < \delta < 1$ is a depreciation rate.

Assume also that there is a given amount of revenue, R, to be collected every period. Moreover, assume that the government budget in every period is balanced.

3 First-Best Optimum

The first task is to characterize the allocation that maximizes the steady state welfare level in a single-consumer economy. This is important, because it allows us to get some insight into the nature of forces at work. I treat B (which in a single person economy directly determines the capital stock) as a free choice variable. While the optimal amount of bequests may exceed initial resources, it does not prevent it from being attained in the steady state.

The government's objective is to maximize (4), subject to the revenue and production constraints. Ignoring incentives for a moment, we can set $T(B) = R$. The production function is assumed to be constant returns to scale, which implies that $w(1-L)+(r+\delta)K = F(K, L)$. Using the steady state relationship $X = B$, the Lagrangian for the government's problem may be expressed as,

$$\mathcal{L} = u\left(F\left(\frac{B}{1+r}, 1-L\right) - \frac{\delta B}{1+r} - R, B, L, B\right) + \pi(F_K - \delta - r) \, .$$

Solving for π from the first-order condition for r and substituting the result into the first order conditions for B leads to,

$$r = -(1 + r + F_{KK}K)\left(\frac{u_B + u_X}{u_C}\right) \tag{6}$$

This condition is best understood in terms of the golden-rule. When $u_B + u_X = 0$, from the social point of view, bequests serve only as a source of capital and they do not have any independent effect on utility. Consequently, they should be selected to maximize consumption: the golden-rule solution implies $r = 0$, exactly as in formula 6. When

$u_B + u_X \neq 0$, bequests directly affect the utility level. Consequently, maximization of consumption is no longer the right objective: at the standard golden-rule allocation, a first-order change in the capital stock has only a second order effect on consumption (by the envelope theorem), but it has the first-order effect on bequests.[3] In fact, $\frac{dB}{dK} = 1+r+KF_{KK}$, so that formula (6) shows that the stronger bequests respond to a change in capital stock (i.e., the weaker is the response of the interest rate), the more consumption should deviate from the golden rule level. I will maintain the assumption that $\frac{dB}{dK} > 0$.[4] Then, if $u_B + u_X > 0$ (higher bequests have a positive effect on utility), the capital stock should be increased above the golden rule level in order to reap the benefits of generations passing a higher stock of wealth from one to another.

The first-order condition for B and L may be expressed to highlight the optimal distortions that the government needs to introduce:

$$\frac{u_B}{u_C} - \frac{1}{1+r} = -\frac{u_X + u_C}{u_C} + \frac{rKF_{KK}}{(1+r)(1+r+KF_{KK})} \tag{7}$$

$$\frac{u_L}{u_C} - w = -\frac{rKF_{KN}}{1+r+KF_{KK}}. \tag{8}$$

The optimal distortion on the consumption/bequest margin, equation (7), has two components. First, it serves to correct the externality from bequeathing with the optimal Pigouvian rate of $\frac{u_X+u_C}{u_C}$ (equivalently: $\frac{v_X}{v_C}$). Second, it serves to facilitate the adjustment toward the golden rule level of capital stock. Such a correction may be necessary because the individuals need not internalize the effect that their saving decisions have on the long-run capital stock and, therefore, on the long-run consumption level. The second term accounts for this effect. In particular, it disappears if the golden rule is indeed the

[3] Note that $B = K(1+r)$, so that an increase in the capital stock affects bequests directly and through the effect on the interest rate. Higher capital stock reduces interest rate and requires more resources to be used in order to achieve a given increase in bequests.

[4] This assumption is satisfied by e.g., the Cobb-Douglas production function. A more general sufficient condition for it is $-\frac{KF_{KK}}{F_K} < 1$.

optimum, $r = 0$.

The optimal wage rate (and the corresponding labor supply) is characterized by equation (8). Similarly as in the case of bequests, it is optimal to introduce some distortion at this margin because a change in labor supply affects the interest rate ($F_{KN} > 0$ by concavity and constant returns to scale). This optimal solution may be easily implemented by imposing marginal tax rates that introduce distortions of equations (7-8) and the lump-sum tax that guarantees the balanced government's budget

I finish this section by highlighting the structure of the optimum in the altruistic case with $\mu = 1$. In that case, $u_B = u_C = -u_X$, so that by equation 6 the optimum is characterized by $r = 0$. From equations (7-8), it follows that there should be no distortions. This is intuitive, because it is well known that the optimum under perfect altruism is efficient. It also corresponds directly to the result of Chamley (1986) that in such a model the capital accumulation decisions should not be distorted.

In the next section, I extend this analysis by allowing for heterogeneous individuals with the government interested in redistribution. The crucial complication is that the government is subject to informational constraints, i.e., that it is unable to directly observe individual types.

4 Redistribution with Incentive Constraints

When the government has redistributional objectives, the first-best solution would be to distribute the tax burden unequally across individuals (individualized T's), while preserving the incentive structure given by equations (7-8). However, as originally emphasized by Mirrlees (1971), individualized taxes are not possible in practice, because they require the ability to observe individual types. A feasible solution is to set up a single tax function, and let individuals optimize subject to the constraints it introduces. For the purpose of selecting such a function, it is useful to apply the revelation principle: policy instruments are assumed to depend on individuals' types with people "choosing" their type. The optimization is subject to the self-selection constraint that everybody selects his own type.

4.1 Government's Problem

The problem will be solved using the original Mirrlees' approach treating the utility level v as a state variable. In this approach, taxes are not explicitly present in the problem, but they are defined implicitly by the paths of utility and consumers' allocations. In order to set up the problem, one needs to calculate $\dot{v}$. By the envelope theorem,

$$\dot{v} = w(1-L)(1-T_L)v_C + v_X\dot{X}\,. \tag{9}$$

In order to eliminate the tax terms from the problem, observe that utility maximization implies that $v_L = v_C(1-T_L)\alpha w$. Consequently,

$$\dot{v} = \frac{1-L}{\alpha}v_L + v_X\dot{X}\,. \tag{10}$$

This is only the first-order condition for individual optimization. In the appendix, the second-order condition is derived as well. That condition is not explicitly imposed in what follows so that the optimal policy found is a correct solution only to the extent that it is satisfied. The optimal tax formulae that are derived should be understood as a characterization of the optimal policy in the intervals where the second-order constraint is not binding.

Imposing the steady state condition $B = X$, the optimal tax problem may be expressed formally as

$$\max \int \omega(v(\alpha))\, g(\alpha) \tag{11}$$

$$\dot{v} \;=\; \frac{1-L}{\alpha}v_L + xv_X\,, \tag{12}$$

$$\dot{B} = x\,, \tag{13}$$

$$\dot{R} = Tg(\alpha)\,, \tag{14}$$

$$\dot{K} = \frac{B}{1+r}g(\alpha)\,, \tag{15}$$

$$\dot{N} = \alpha(1-L)g(\alpha)\,. \tag{16}$$

Apart from these conditions, optimization is subject to terminal conditions determining the optimal factor prices. These conditions do not play a direct role in determination of the marginal tax rates (other than via determining the equilibrium interest and wage rates), so that they are suppressed for now. I analyze them in Section 5. The control variables in this problem are x and L, and the state variables are v, B, R and K. The total tax liability, $T(B, L, \alpha, w, u)$, is defined implicitly by $v(\alpha) = v(\alpha w(1-L) - \frac{B}{1+r} - T, B, L, B)$. By the implicit function theorem this equation yields partial derivatives of $T(\cdot)$: $T_v = -\frac{1}{v_C}$, $T_L = \frac{v_L}{v_C} - \alpha w$ and $T_B = \frac{v_B + v_X}{v_C} - \frac{1}{1+r}$. In order to apply the optimal control approach, I set up the Hamiltonian as follows,

$$\mathcal{H} = \omega(v)g(\alpha) + \lambda\dot{v} + \beta x + \rho\dot{R} + \kappa\frac{B}{1+r}g(\alpha) + \eta\alpha(1-L)g(\alpha)\,. \tag{17}$$

The first order conditions with respect to the control variables are,

$$0 = \beta + v_X\lambda\,, \tag{18}$$

$$0 = \rho\frac{d\dot{R}}{dL} + \lambda\frac{d\dot{v}}{dL} - \eta\alpha g(\alpha)\,. \tag{19}$$

The first-order conditions with respect to R and K imply simply that $\dot{\rho} = \dot{\kappa} = 0$. Respective conditions for u and B are

$$-\dot{\lambda} = \omega'(v)g(\alpha) + \rho\frac{d\dot{R}}{dv} + \lambda\frac{d\dot{v}}{dv}\,, \tag{20}$$

$$-\dot{\beta} = \rho\frac{d\dot{R}}{dB} + \lambda\frac{d\dot{v}}{dB} + \kappa\frac{g(\alpha)}{1+r}\,. \tag{21}$$

4.2 Optimal Marginal Tax Rates

These equations may be manipulated to derive an expression for the optimal marginal income and estate tax rates.

Proposition 1 *The optimal marginal income tax rate is given by*

$$\frac{v_L}{v_C} - \alpha w = \frac{\lambda v_C}{\rho \alpha g(\alpha)} \frac{\partial}{\partial L} \left\{ (1-L) \frac{v_L}{v_C} \right\} - \frac{\lambda v_C (1-L) x}{\rho \alpha g(\alpha)} \frac{\partial}{\partial X} \left\{ \frac{v_L}{v_C} \right\} + \frac{\eta}{\rho}. \tag{22}$$

Proof: Directly from formula 19. □

The first term of this formula is identical to the optimal solution derived by Atkinson and Stiglitz (1976). Its sign depends on the sign of λ. Intuitively, it is the direction of redistribution that will be discussed in more details in what follows. It also depends on the sign of $\frac{\partial}{\partial L}\left\{(1-L)\frac{v_L}{v_C}\right\}$, which is traditionally assumed to be negative, reflecting the necessary single-crossing property.[5] The last term reflects some aggregate effects and is discussed in Section 5. The second term is of particular interest. Its form is reminiscent of the formula for the optimal commodity tax rates derived by Atkinson and Stiglitz. It is just the optimal tax formula for the labor tax when bequests are the source of inequality. This simply reflects that when inequality of bequests has to be dealt with, labor income has an aspect of being a commodity. Consequently, its taxation should be (partially) determined by the commodity taxation considerations. This term is equal to zero when the externality from bequests is separable from other commodities.

Next proposition shows that similar structure of the optimal tax is apparent also in the optimal estate tax formula.

[5]It is this assumption that guarantees a local optimum of individual problem is a global one. Although ignored here, the second-order condition of individual optimization ought to be imposed to guarantee that the proposed tax policy corresponds to the maxima of the individual problems not the minima.

Proposition 2 *The optimal marginal estate tax rate is given by*

$$\frac{v_B}{v_C} - \frac{1}{1+r} = -\frac{v_C\lambda(1-L)}{\rho g(\alpha)\alpha}\frac{\partial}{\partial L}\left\{\frac{v_B}{v_C}\right\} - \frac{\kappa}{(1+r)\rho} - \frac{\omega'(v)v_X}{\rho} - \frac{\lambda v_C x}{\rho g(\alpha)}\frac{\partial}{\partial X}\left\{\frac{v_B}{v_C}\right\}. \quad (23)$$

Proof: In the appendix. □

The formula for the optimal estate tax differs from the formula for the optimal income tax in two qualtitative respects. First, it is about the distortion on the bequest-consumption margin ($\frac{v_B}{v_C}$) instead of the leisure-consumption margin ($\frac{v_L}{v_C}$). Second, it depends on the bequest externality (the third term).

In the Atkinson and Stiglitz (1976) world, the marginal commodity tax rate would be given by the first term. When the utility function is weakly separable between leisure and the composite of consumption and bequests, $\frac{v_B}{v_C}$ is not a function of leisure so that $\frac{\partial}{\partial L}\left\{\frac{v_B}{v_C}\right\}$ is zero. This just says that when the utility function is weakly separable, observing consumption of different commodities, in addition to just observing labor income, does not provide any new information about the skill level. This is because, in the separable case, consumption and bequests are a function of total income, $X + \alpha w(1-L)$ and not the skill level α directly. Consequently, the pattern of commodity consumption cannot provide any information about the skill level other than what is already known from observing the total taxable income and inheritance.

The following terms in this formula are due to the presence of imperfections. The second term reflects the aggregate effect of the capital stock not internalized by individuals (this is discussed in Section 5). The third term is just a Pigouvian correction for the externality from bequeathing.

The last term in formula (23) is particularly important. Note first that it is zero when externalities are not present. But, it also bears a close resemblance to the standard formula for the optimal income tax rate, i.e. the first term in Proposition 1, with the marginal impact of higher wage rate, $(1-L)v_L$, replaced by the marginal impact of higher inheritance, xv_B and derivatives taken with respect to a change in inheritance instead of a change in labor income.[6]

The last term is proportional to the effect of higher inheritance on the marginal rate of substitution between bequests and consumption. In other words, it indicates how consumption patterns of those with higher inheritance differ from those of the poor. The presence of this term and its analogy to the standard labor income tax formula simply reflects that, in the presence of externalities, intergenerational transfers are an independent source of inequality. When full social benefits or costs of bequests are not internalized by individuals, they contain some information about the skill level that ought to be exploited for redistributional reasons. As the result, estates should be treated analogously to the way that labor income is treated. Loosely speaking, the inheritance becomes a type of "income" not a commodity. As such, it should be a direct target of the tax policy, instead of being only used as a supplementary source of information about the skill level.

The sign of the various terms in (23) depends on the sign of λ. This is the multiplier on the self-selection constraint, and its sign indicates which of the incentive constraints is binding. Just as is the case in the optimal income taxation problem, λ is negative whenever the incentive constraints of the rich bind, in other words, when the tax system is redistributive. I will demonstrate in Section 6 that $\lambda < 0$ under reasonable conditions.

In general, it is not possible to sign formula (23), although its components can be signed. The task is easier when the special case of additive separability is considered. Then, the first term is equal to zero. The second term is uniform across individuals and is not directly related to the presence of externality. When $\lambda < 0$, the last term is positive. When a negative externality is present, the contribution of the last two terms is therefore unambiguously positive. Negative externalities imply that estate tax rates should be above

[6] The first term in Proposition 1 may be alternatively expressed as $-\frac{\lambda v_C}{\rho g(\alpha^*)} \frac{\partial}{\partial \alpha^*(1-L)} \left\{ \frac{(1-L)v_L}{v_C} \right\}$, where α^* is used to highlight that α is fixed.

the level[7] required to achieve the long-run efficiency. When externalities are positive, the estate tax must balance redistribution and correction of the externality, with ambiguous sign in general.

The results of this section may be summarized as follows. Under reasonable conditions, the optimal steady state policy should be redistributive (i.e., $\lambda < 0$, it is demonstrated in Section 6), although it need not automatically guarantee that the marginal income and estate tax rates are positive. The optimal taxes have a modular structure that reflects different features of the model. First, there is a redistributive component. Different tax bases can be classified as either reflecting the exogenous source of inequality (being "income") or not (being a "commodity"). "Income" always has to be taxed for redistributive reasons, while "commodities" should be taxed only if they provide an independent source of information about the source of inequality (which they do only if a suitable separability assumption does not hold). In the model, there may be two exogenous sources of inequality: skill levels and inheritances, although the latter one is such a source only if there is an external aspect of bequests. With two sources of inequality, both labor income and bequests play simultaneous roles of "income" and "commodity" and contribute two terms to the optimal tax formulae. The second type of component reflects the Pigouvian correction of the bequest externality, if present. By the logic of the principle of targeting, it affects only the structure of estate taxation. Finally, the optimal tax rules reflect the effect that taxes play in determination of the aggregate capital stock and labor supply. This is discussed in the next section.

[7] Given by $-\frac{\kappa}{(1+r)\rho}$. The rationale for this statement is explained in Section 5.

5 Factor Prices

Until now, the factor prices and the multipliers corresponding to the capital and labor supplies (κ and η) were taken as given. These are, however, endogenous variables.

Recall that the multiplier κ affects the optimal estate tax rate (derived in Proposition 2), and the multiplier η affects the optimal income tax rate. They reflect aggregate externalities acting through the aggregate supply of factors of production.[8]

Define $U_r \equiv \int \lambda \frac{d\dot{v}}{dr}\, d\alpha$ and $U_w \equiv \int \lambda \frac{d\dot{v}}{dw}\, d\alpha$. In the appendix, it is demonstrated that κ and η are given by,

$$\frac{\kappa}{\rho(1+r)} = \frac{U_r F_{KK} + U_w F_{KN}}{\rho(1+r+KF_{KK})} - \frac{rKF_{KK}}{(1+r)(1+r+KF_{KK})}, \tag{24}$$

$$\frac{\eta}{\rho} = \frac{(1+r)(U_r F_{KN} + U_w F_{NN})}{\rho(1+r+KF_{KK})} - \frac{rKF_{KN}}{1+r+KF_{KK}}. \tag{25}$$

These formulae should be related to both the first-best solution (equations 7-8) and the optimal marginal tax rates given by Propositions 1 and 2. The aggregate contributions to the optimal taxes in redistributive context consist of two types of terms: the redistributive components (first terms of both of the above formulae) and the first-best adjustments that pushes the economy toward the "modified golden-rule" level. The latter one was discussed in Section 3. This effect is not, however, specific to the redistributive problem and is marginal to the topic of this paper. Furthermore, it may be kept separate from other effects.

Here, I concentrate on the question whether the aggregate distortion is nessary for redistributive reasons, i.e., whether the first terms in formulae 7 and 8 are zero.

[8] There would not be such an externality in a perfectly competitive, complete information setting. However, when the economy does not have to be dynamically efficient and the skill level remains private information, prices do not accurately reflect all social benefits and costs. Consequently, there may be an external effect acting through the factor prices that can and ought to be exploited by the policy. Greenwald and Stiglitz (1986) discuss this point in more details.

Proposition 3 *When $v(\cdot)$ is additively separable,* $\frac{\kappa}{\rho(1+r)} = \frac{rKF_{KK}}{(1+r)(1+r+KF_{KK})}$ *and* $\frac{\eta}{\rho} = -\frac{rKF_{KN}}{1+r+KF_{KK}}$.

Proof: Straightforward calculations yield $\frac{d\dot{v}}{dr} = \frac{B}{(1+r)^2}\left(\frac{1-L}{\alpha}v_{CL} + v_{XC}x\right)$ and $\frac{d\dot{v}}{dw} = \alpha(1-L)\left(\frac{1-L}{\alpha}v_{CL} + v_{XC}x\right)$. When the utility function is additively separable $v_{LC} = v_{XC} = 0$, so that $U_r = U_w = 0$. □

In the additively separable case, the aggregate distortions should be set at the first-best level as in section 3. In other words, affecting relative prices plays no role in the optimal redistributive policy. Note that the production side of the economy does not affect the underlying degree of inequality because differences between individuals are determined by the exogenous distribution of α. Thus, the policy could be useful only to the extent that it relaxes incentive constraints. However, in the separable case distorting the interest rate may not play this role because of the familiar Atkinson-Stiglitz logic: distorting relative prices of non-labor goods does not provide any information about the underlying skill level. In this case, also distorting the wage rate does not play a redistributive role because it affects all types proportionally and thus does not help in distinguishing between them. From the proof of Proposition 3 it is clear, however, that when the utility function is not separable the redistributive components of formulae 24 and 25 do not disappear. Therefore, the optimal policy may then still be needed to affect the long-run mix of the production factors in order to relax the incentive constraints. This is discussed in the next section.

It is interesting to point out the relationship of this discussion and that of Naito (1999). In that model, there are two types of skills with different wage rates. By affecting relative supplies of factors of production and therefore distorting the optimal prices, different types are affected differently. Thus, distorting the relative prices is useful even in the separable case. In the context of this model, it would correspond to wage rate being dependent on the skill level α. Affecting factor prices could then have a real effect on the relative skill inequality.

5.1 Relationship to No Capital Taxation Results

One of the most important public finance results, due to Chamley (1986) and Judd (1985), says that there should be no capital taxation in the long-run. This result was demonstrated in the representative agent, perfect foresight, framework. There is a special case of the present model that also features those assumptions: it corresponds to the perfect-altruism case with no bequest externalities.

Formally, $v_B = \mu v_C$ and $v_X = 0$, where μ is the rate of time preference. Is it still true that capital accumulation decisions should not be distorted? Because the only source of capital stock are bequests, the question is whether bequeathing decisions should be distorted.

Proposition 4 *Assume that individuals are altruistic ($v_B = \mu v_C$ and $v_X = 0$) and no externalities are present. Then, the marginal estate tax rate is constant and equal to* $-\frac{\kappa}{\rho(1+r)}$.

Proof: It directly follows from Proposition 2, because under perfect altruism $\frac{v_B}{v_C}$ is constant. □

This proposition shows that the optimal estate tax is linear. Before I discuss whether the optimal tax rate is zero, it is interesting to understand why the marginal tax rate is constant. According to the formula of Proposition 2, a nonlinear tax may be optimal for two reasons: externalities (third and fourth term) or redistribution (first term). Of these, the redistributive component is particularly interesting because it has been suggested (Atkeson, Chari and Kehoe, 1999) that the Chamley-Judd no-capital-taxation result remains valid when agents are heterogeneous. However, the reason why it disappears in the current case is that $\frac{\partial}{\partial L}\left\{\frac{v_B}{v_C}\right\} = 0$, which is just an assumption about the utility function that needn't hold in general. This is the Atkinson-Stiglitz logic: the altruistic utility that I consider is separable between leisure and other commodities and therefore distorting bequeathing decision is not useful for redistribution. It has nothing to do with the role that bequests play in capital accumulation.[9]

Is the optimal tax rate zero so that there are no capital distortions? Not necessarily. To see it, assume the extreme case of $\mu = 1$. Then, no capital distortion corresponds to $r = 0$ and no estate taxation corresponds to $\kappa = 0$.

First observe that, were $\kappa = 0$, the optimal income tax formula would be exactly as in the static problem with redistribution taking place.[10] By equation 23, there still should be individual saving distortions as long as $\frac{\partial}{\partial L}\left\{\frac{v_B}{v_C}\right\} \neq 0$ (all other terms in that formula are zero due to the simplifying assumptions made). So, no aggregate distortion does not necessarily entail no individual distortions. In other words, the Atkinson-Stiglitz commodity taxation logic applies even in the long-run.

However, the aggregate distortion may be present as well. Substituting for U_r and U_w in equation 24 (remembering that $v_X = 0$ and $r = 0$) yields $(1 + KF_{KK})\kappa = \int \lambda \frac{1-L}{\alpha} \frac{v_{CL}}{g(\alpha)} \left(BF_{KK} + \alpha(1-L)F_{KN}\right) g(\alpha) d\alpha = 0$. Clearly $\kappa = 0$ when $v_{CL} = 0$. Otherwise, it need not be the case. Recall that $KF_{KK} + NF_{KN} = 0$, $K = \int Bg(\alpha)d\alpha$ and

[9] Atkeson et al. (1999) did not consider informational constraints. For some of their arguments, they assumed that individual types can be observed. They also considered the case when the marginal tax rates are identical for everybody. In that case, the incentive constraints are also absent, and therefore the effect I discuss is of no value.

[10] Note that redistribution still takes place because λ is not uniformly equal to zero. This may be seen as follows. Suppose that it is not true. Because $\frac{dR}{dv} = v_C$, equation 20 implies that $\omega'(v)v_C$ is constant (equal to ρ) in the whole interval where $\lambda = 0$. By Proposition 1, the marginal income tax rate is also constant when $\lambda = 0$. By Proposition 4 the marginal estate tax rate is constant. As shown in the proof of Proposition 5, it is not possible then for the marginal utility of consumption to stay constant, leading to a contradiction.

$N = \int \alpha(1-L)d\alpha$. Therefore, $(1 + KF_{KK})\kappa = \frac{F_{KN}}{N}\text{cov}\left(\lambda\frac{1-L}{\alpha}\frac{v_{CL}}{g(\alpha)}, \frac{B}{K} - \frac{\alpha(1-L)}{N}\right)$. While theoretically possible that this covariance term be zero, nothing in the model forces it to be the case. In fact, with $\kappa = 0$ it would be $\eta = 0$ as well,[11] and redistribution would take place as in the static problem. Only by accident would the solution to the static problem lead to this covariance term being zero.

Why might one want to have capital distortion at the optimum? The direct effect of it is a reduction in consumption down from its "golden rule" level. In a representative agent context there is no other effect of capital distortion, so that it is undesirable. In the current situation, there is however an additional potential benefit: a change in the interest rate may relax incentive compatibility constraints. This is analogous to the effect Naito (1999) concentrates on.[12]

6 Direction of Redistribution

The sign of the multiplier λ represents the direction of redistribution. For example, $\lambda \leq 0$ corresponds to binding incentive constraints of the higher types. In the context of the Mirrlees' model, it was demonstrated under weak regularity conditions that $\lambda \leq 0$ (Mirrlees, 1971; Seade, 1982; Röell, 1985; Brunner, 1995). In particular, all these proofs proceeded by considering marginal changes at an optimum and therefore demonstrated a stronger result than needed: that at any local maximum there is $\lambda \leq 0$. The proposition that follows shows that (under some assumptions) in the *global* optimum the direction of binding incentive constraints (and redistribution) is intuitive: $\lambda \leq 0$. The idea of the

[11] To see it, observe that under constant returns to scale $F_{KN} = -F_{KK}K/N$ and $F_{NN} = -F_{KN}K/N$, so that $U_rF_{KN} + U_wF_{NN} = -(U_rF_{KK} + U_wF_{KN})K/N$. Therefore, formulae 24-25 imply that $\eta = -\kappa K/N$.

[12] Atkeson et al. (1999) do not rule out optimality of capital taxation if there are additional constraints in the problem that depend on the level of capital stock. The incentive constraints depend on the level of capital stock via dependence on the interest rate.

proof is to consider a relaxed problem with only the incentive constraints of the high types imposed. Any feasible allocation in the complete problem is also feasible in the relaxed one. It turns out that at the optimum of the relaxed problem the incentive constraints of the high types must be binding (and thus the incentive constraints of the low types are slack). Consequently, the optimum of this relaxed problem is a feasible allocation for the complete problem so that it must be the global optimum.

In the proof, the following characterization theorem, due to Brunner (1995) is used:

Lemma 1 (Theorem 1 in Brunner (1995)) *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is strictly concave, twice-differentiable function with partial derivatives $f_i > 0$. Assume that good i is non-inferior. Let $\tilde{r}, \tilde{q} \in \mathbb{R}^n$ be consumption bundles such that $\frac{f_k(\tilde{r})}{f_l(\tilde{r})} = \frac{f_k(\tilde{q})}{f_l(\tilde{q})}$ for any $k, l \neq i$, $\frac{f_i(\tilde{r})}{f_l(\tilde{r})} \geq \frac{f_i(\tilde{q})}{u_l(\tilde{q})}$ for any $l \neq i$ and $f(\tilde{r}) \geq f(\tilde{q})$. Then, $f_k(\tilde{r}) \leq f_k(\tilde{q})$ for any $k \neq i$ (when good i is normal, inequalities are strict).*

This result can be easily seen in the two-dimensional case (say, with leisure and consumption): when the utility function is concave, the marginal utilities must fall along the income expansion path. It is also not difficult to show that, when leisure is normal, substitution along the indifference curve toward less leisure and more consumption decreases the marginal utility of consumption. Lemma 1 combines these two effects and generalizes this result to the multi-dimensional case.

I am now ready to demonstrate that the optimal tax system is redistributive.

Proposition 5 *Suppose the utility function v^* satisfies (1) the normality of leisure; and (2) $(v^*_{CC}v^*_{BL} - v^*_{CL}v^*_{BC})\, v_X \geq 0$. Then, in the global optimum, $\lambda \leq 0$.*

Proof: Consider the relaxed problem that imposes only the constraints that the higher types do not mimic lower ones. Formally, the relevant constraint is $\dot{v} \geq \frac{1-L}{\alpha} v_L + x v_X$. Note that when this constraint is binding, the low types will also not deviate. Therefore, if one can demonstrate that this constraint is binding everywhere, the solution to the relaxed problem is the solution to the original one as well. This constraint may be represented in the optimal control problem by setting $\dot{v} = s$ and adding to the Hamiltonian a new term of $\xi\left(\frac{1-L}{\alpha}v_L + xv_X - s\right)$.

There are two new necessary first-order conditions for the optimum: $\lambda = \xi$, and $\xi \leq 0$ (the latter condition reflects that the newly introduced constraint is an inequality). The constraint is binding if $\xi < 0$. Suppose that at the optimum there is no interval with $\xi = 0$. Then, we are set because the constraints of the high types are binding everywhere. Thus, the optimum is also the optimum of the unconstrained problem, with $\lambda = \xi < 0$ everywhere except for the endpoints. So, suppose that it is not the case. Then there is also an interval with $\lambda = 0$, i.e. $\dot{\lambda} = 0$. From the first-order conditions, in the same interval it must be $\beta = 0$ and $\dot{\beta} = 0$. Consequently, everywhere in this interval, $\frac{v_L}{v_C} = \frac{\alpha(w+\eta)}{\rho}$ (by equation 19), $\omega'(v)v_C = \rho$ (by equation 20) and $\frac{v_B+v_X}{v_C} = \frac{1}{1+r} - \frac{\kappa}{(1+r)\rho}$ (from equation 21). Is it possible? Consider the utility function v^* which was defined to be a function of C, B and L (because $X = B$). Observe that $v^*_C = v_C$, $v^*_L = v_L$ and $v^*_B = v_B + v_X$. By assumption, $v^*(\cdot)$ is concave. The utility is increasing in the skill level α by the incentive constraint and the assumption stated in the proposition (proof is in the appendix).[13] We established that the marginal rate of substitution between C and B does not change with α and that the marginal rate of substitution between L and C is increasing in α. By Lemma 1,[14] if leisure is normal, v_C and $v_B + v_X$ must then both be decreasing in α. This is, however, impossible because $\omega'(v)v_C$ may not change. □

The assumption of normality of leisure is not controversial. The second assumption

[13]Note that this is the reason why an analogous approach with only incentive-constraints of the low types does not work. The optimum in that case would be the first-best global optimum, and it is well-known that it should feature decreasing utilities, so that this step of the proof fails.

[14]Note that in Lemma 1 it is assumed that $f_i > 0$. When $v_B + v_X < 0$, however, the lemma may be applied to function $\tilde{v}^*(C, -B, L) = v^*(C, B, L)$ which is increasing in all arguments and satisfies all other assumptions of Lemma 1 whenever v^* satisfies them.

of the proposition is less clear. Its role is to rule out some cases when the utility level could be falling at the optimum with the individual skill level. Neither the first nor the second-order conditions rule out a possibility of utility decreasing in the optimum when an externality is present and it may occur when $\dot{B}$ and v_X are of the opposite signs. In that case, however the meaning of a 'redistributive' policy becomes unclear, and it is not surprising that the direction of redistribution may be affected. There are some simple cases when this assumption is satisfied. In particular, regardless of the sign of the externality, this assumption holds when either leisure is additively separable from other commodities ($v^*_{BL} = v^*_{CL} = 0$) or when bequest is additively separable from other commodities ($v^*_{BL} = v^*_{BC} = 0$),[15] because $v^*_{CC}v^*_{BL} - v^*_{CL}v^*_{BC} = 0$ in these cases. Furthermore, it is straightforward to show that this term is also zero when the utility function has a form of $v^* = f(v)$ where $v(\cdot)$ features either leisure or bequests separable from other commodities. Finally, the result obtains also when $v_X = 0$. Under these assumptions, Proposition 5 states that the optimal policy should be redistributive, meaning that the incentive constraints of higher types are binding.

7 Conclusions

The presence of endogenous wealth inequality is not by itself an argument for the estate taxation. As long as individual and social objectives coincide, a bequest is just another commodity. Consequently, estate taxation is only a supplement to the primary redistributive instrument of income taxation and in some special cases it is not necessary at all. The general equilibrium considerations affect this conclusion in only a minor way: in some circumstances the estate tax may need to be used to affect the factor mix and relative prices in equilibrium, but this effect leads only to a linear tax and is not present

[15] It is also easy to show that it remains to be true if $u_X = \psi u_C + h(X)$ for some constant ψ and a function $h(\cdot)$.

in a special case of leisure being weakly separable from other commodities. Nevertheless, my results show that in the realistic public finance context with informational constraints present, the Atkinson-Stigliz commodity taxation logic applies even in the long-run and when commodity taxation implies introducing saving distortions. The "no capital taxation" theorems should be interpreted as suggesting that factors of production should be set at the efficient level, however it can be achieved by subsidizing some individuals and taxing others, as dictated by redistributional objectives. Finally, even aggregate capital distortions may be useful in cases when they relax incentive constraints, analogously to the static result of Naito (1999).

These conclusions should be additionally refined when social and individual objectives do not coincide. Arguably, this is a realistic case: empirical tests routinely reject the altruistic model leaving room for some "egoistic" bequest motive. Even more importantly, the opponents of the tax often stress social costs of concentration of wealth that are not internalized when people make their decisions. In each of these cases, bequests have an external aspect. As with any externality, the optimal policy has to reflect the Pigouvian correction of the problem. What is more interesting is that inheritances are no longer a "commodity" but rather a type of "income": they are to some extent exogenous to the decisions of individuals. As the result, they are not just a consequence but also a source of inequality. Regardless of the sign of the externality, this calls for a redistributive policy that will attempt to erase these differences. Mathematically, the redistributive components of labor income and estate taxation are symmetric. Each of them should be used to target the corresponding source of inequality and each of them supplements the role of the other instrument when the appropriate separability assumption is not satisfied.

The discussion suggests that high estate tax rates can be justified on the grounds of externalities or by appealing to some form of complementarity between bequests and leisure. However, the claim that one needs estate taxation for redistribution must rest on something more than just the evidence of inequality of wealth.

A Appendix

The second-order condition. *Proof:* The approach is due to Mirrlees (1976). Observe that $X(\alpha)$ is taken as given by the individual and write,

$$w(\alpha) = \max_{\beta} W(y(\beta), B(\beta), C(\beta), \alpha) = \max_{\beta} u\left(X(\alpha) + C(\beta), B(\beta), 1 - \frac{y(\beta)}{\alpha}, X(\alpha)\right),$$

where $y(\beta) = \alpha(1 - L(\beta))$ and $C(\beta)$ is defined appropriately. The dependence on α reflects both its direct effect on the utility level and the effect of $X(\alpha)$. The self-selection requires that $\beta = \alpha$. The second-order condition for individual optimization is given by $\frac{d^2W}{d\beta^2} \leq 0$. Observe that $\frac{dW}{d\beta} = 0$ for any α so that $\frac{d^2W}{d\beta^2} + \frac{dW_\alpha}{d\beta} = 0$. Observe also that $w_{\alpha\alpha} = \frac{d^2W}{d\beta^2} + 2\frac{dW_\alpha}{d\beta} + W_{\alpha\alpha}$. Combining these results yields an alternative expression for the second-order condition, $w_{\alpha\alpha}(\alpha) \geq W_{\alpha\alpha}$. Note that $w_\alpha(\alpha) = W_\alpha$, so that $w_{\alpha\alpha}(\alpha) = W_{\alpha y}\dot{y} + W_{\alpha C}\dot{C} + W_{\alpha B}\dot{B} + W_{\alpha\alpha}$. Note also that it must also be $W_C\dot{C} + W_y\dot{y} + W_B\dot{B} = 0$, so that the second-order condition is, $0 \leq W_{\alpha\alpha} - w_{\alpha\alpha}(\alpha) = \dot{y}W_C\frac{\partial}{\partial\alpha}\frac{W_y}{W_C} + \dot{B}W_C\frac{\partial}{\partial\alpha}\frac{W_B}{W_C}$. □

Proof of Proposition 2. *Proof:* Differentiating the first-order condition for x yields,

$$-\dot{\beta} - \dot{\lambda}v_X - \lambda\dot{v_X} = 0 \tag{26}$$

Note that changing the order of differentiation (using Young's Theorem) yields, $\dot{v_X} = \frac{\partial\dot{v}}{\partial X}$. Substituting for $\dot{\beta}$, $\dot{\lambda}$ and $\dot{v_X}$ in equation (26) yields,

$$0 \;=\; \omega'(v)g(\alpha)v_X + \rho\left(\frac{d\dot{R}}{dv}v_X + \frac{d\dot{R}}{dB}\right) + \lambda\left(\frac{d\dot{v}}{dv}v_X + \frac{d\dot{v}}{dB} - \dot{v_X}\right) + \kappa\frac{g(\alpha)}{1+r}. \tag{27}$$

One can evaluate parts of this expression as follows,

$$\frac{d\dot{R}}{dv}v_X + \frac{d\dot{R}}{dB} = \left(-\frac{v_X}{v_C} + \frac{v_B + v_X}{v_C} - \frac{1}{1+r}\right)g(\alpha) = \left(\frac{v_B}{v_C} - \frac{1}{1+r}\right)g(\alpha),$$

and,

$$\frac{d\dot{v}}{dv}v_X \quad + \quad \frac{d\dot{v}}{dB} - \dot{v_X} = \frac{\partial \dot{v}}{\partial C}\frac{v_X}{v_C} + \left(-\frac{\partial \dot{v}}{\partial C}\frac{v_B + v_X}{v_C} + \frac{\partial \dot{v}}{\partial B} + \frac{\partial \dot{v}}{\partial X}\right) - \frac{\partial \dot{v}}{\partial X} \tag{28}$$

$$= \frac{\partial \dot{v}}{\partial B} - \frac{\partial \dot{v}}{\partial C}\frac{v_B}{v_C} = v_C \frac{d}{d\alpha}\left\{\frac{v_B}{v_C}\right\}, \tag{29}$$

so

$$\frac{v_B}{v_C} - \frac{1}{1+r} = -\frac{v_C \lambda}{\rho g(\alpha)}\frac{d}{d\alpha}\left\{\frac{v_B}{v_C}\right\} - \frac{\omega'(v)v_X}{\rho} - \frac{\kappa}{(1+r)\rho}. \tag{30}$$

Alternatively, the term $v_C \frac{d}{d\alpha}\left\{\frac{v_B}{v_C}\right\} = \frac{\partial \dot{v}}{\partial B} - \frac{\partial \dot{v}}{\partial C}\frac{v_B}{v_C}$ is equal to,

$$\frac{\partial \dot{v}}{\partial B} - \frac{\partial \dot{v}}{\partial C}\frac{v_B}{v_C} = \frac{1-L}{\alpha}v_C \frac{\partial}{\partial L}\left\{\frac{v_B}{v_C}\right\} + \frac{\partial}{\partial X}\left\{\frac{v_B}{v_C}\right\} x, \tag{31}$$

and substituting it into formula (30) yields equation in text. □

Formulae for κ and η. *Proof:* The government is subject to two aggregate constraints:

$$r + \delta \quad = \quad F_K(K^*, N^*), \tag{32}$$

$$w \quad = \quad F_N(K^*, N^*), \tag{33}$$

where K^* and N^* are the aggregate capital stock and labor supply. Variables r and w are slightly uncommon for the optimal control problem, but they can be easily handled. They may be thought of as constant state variables, $\dot{r} = 0$, $\dot{w} = 0$. Leonard and Long (1992) refer to such variables as "control parameters". In addition, factor price constraints form terminal constraints imposed on the state variables K and N. One approach to this problem is to incorporate these constraints as the scrap value function:

$$S(r, w, K^*, N^*, p_1, p_2) = p_1(r + \delta - F_K(K^*, N^*)) + p_2(w - F_N(K^*, N^*))), \tag{34}$$

where p_1 and p_2 are the multipliers on the aggregate constraints. The first-order conditions are,

$$0 \quad = \quad \int \frac{d\mathcal{H}}{dr}\, d\alpha + p_1 = \int \left(\lambda \frac{d\dot{v}}{dr} + \rho \frac{d\dot{R}}{dr} - \kappa \frac{Bg(\alpha)}{(1+r)^2}\right) d\alpha + p_1, \tag{35}$$

$$
\begin{aligned}
0 &= \int \frac{d\mathcal{H}}{dw}\, d\alpha + p_2 = \int \left(\lambda \frac{d\dot{v}}{dw} + \rho \frac{d\dot{R}}{dw} \right) d\alpha + p_2 \,, &(36)\\
\kappa &= -p_1 F_{KK} - p_2 F_{KN} \,, &(37)\\
\eta &= -p_1 F_{KN} - p_2 F_{NN} \,. &(38)
\end{aligned}
$$

Note that $\frac{d\dot{R}}{dr} = \frac{Bg(\alpha)}{(1+r)^2}$ and $\frac{d\dot{R}}{dw} = \alpha(1-L)g(\alpha)$. Furthermore, recall that $K = \int \frac{B}{1+r} g(\alpha)\, d\alpha$ and $N = \int \alpha(1-L)g(\alpha)\, d\alpha$. Use these results to solve for p_1 and p_2 from equations (35) and (36) respectively. Substitution into (37) and (38) yields,

$$
\begin{aligned}
\kappa &= F_{KK} \left(U_r + K \frac{\rho - \kappa}{1+r} \right) + F_{KN}(U_w + N\rho) \,,\\
\eta &= F_{KN} \left(U_r + K \frac{\rho - \kappa}{1+r} \right) + F_{NN}(U_w + N\rho) \,,
\end{aligned}
$$

where $U_r = \int \frac{d\dot{v}}{dr}\, d\alpha$ and $U_w = \int \frac{d\dot{v}}{dw}\, d\alpha$. The first of these equations may be solved for κ (using the identity $NF_{KN} = -KF_{KK}$ due to $F(\cdot)$ being constant returns to scale):

$$
\frac{\kappa}{1+r} = \frac{U_r F_{KK} + U_w F_{KN}}{1 + r + KF_{KK}} - \rho \frac{rKF_{KK}}{(1+r)(1+r+KF_{KK})} \,.
$$

Using this result, the solution for η may be easily obtained (again, using $NF_{KN} = -KF_{KK}$ and $NF_{NN} = -KF_{KN}$),

$$
\eta = \frac{(1+r)\,(U_r F_{KN} + U_w F_{NN})}{1 + r + KF_{KK}} - \rho \frac{rKF_{KN}}{1 + r + KF_{KK}} \,.
$$

□

Proof that $\dot{v} > 0$ in Proposition 5.. *Proof:* To simplify notation note that all derived conditions may be expressed in terms of derivatives of $\omega(v)$. Equivalently, instead of using ω we can take u to be a concave transformation of the utility function. Thus, ω' will be dropped in what follows, without loss of generality. Note that v_C and $v_B + v_X$ (i.e., v_C^* and v_B^*) must stay constant and that $\frac{dv_L}{d\alpha} = w + \eta$. Consequently,

$$
\begin{aligned}
0 &= v_{CC}^* \dot{C} + v_{CL}^* \dot{L} + v_{CB}^* \dot{B} \,,\\
w + \eta &= v_{LC}^* \dot{C} + v_{LL}^* \dot{L} + v_{LB}^* \dot{B} \,,\\
0 &= v_{BC}^* \dot{C} + v_{BL}^* \dot{L} + v_{BB}^* \dot{B} \,,
\end{aligned}
$$

Solve for $\dot{B}$ and note that the denominator is negative by concavity of the utility function. Thus, $sign\dot{B} = sign(v^*_{CC}v^*_{BL} - v^*_{CL}v^*_{BC})$. By assumption then $v_X\dot{B} \geq 0$ and from the incentive constraint $\dot{v} > 0$. □

References

Atkeson, Andrew, V. V. Chari, and Patrick J. Kehoe, "Taxing Capital Income: A Bad Idea," *Federal Reserve Bank of Minneapolis Quarterly Review*, Summer 1999, *23* (3), 3–17.

Atkinson, Anthony B. and Joseph E. Stiglitz, "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, July-August 1976, *6* (1-2), 55–75.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November-December 1974, *82* (6), 1095–1117.

Bernheim, B. Douglas, Andrei Shleifer, and Lawrence H. Summers, "The Strategic Bequest Motive," *Journal of Political Economy*, December 1985, *93* (6), 1045–76.

Brunner, Johann K., "A Theorem on Utilitarian Redistribution," *Social Choice and Welfare*, 1995, *12*, 175–179.

Carroll, Christopher D., "Why Do the Rich Save So Much?," in Joel Slemrod, ed., *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, New York: Harvard University Press and Russell Sage Foundation, 2000.

Chamley, Christopher, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, May 1986, *54* (3), 607–22.

Davies, James B., "Uncertain Lifetime, Consumption, and Dissaving in Retirement," *Journal of Political Economy*, June 1981, *89* (3), 561–77.

Dixit, Avinash K., "Tax Policy in Open Economies," in Alan J. Auerbach and Martin S. Feldstein, eds., *Handbook of Public Economics*, Vol. 1, Amsterdam; New York: North-Holland, 1985, pp. 313–74.

Greenwald, Bruce C. and Joseph E. Stiglitz, "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics*, May 1986, *101* (2), 229–264.

Hurd, Michael D., "Mortality Risk and Bequests," *Econometrica*, July 1989, *57* (4), 779–813.

Judd, Kenneth L., "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, October 1985, *28* (1), 59–83.

Kaplow, Louis, "Arguments For and Against Estate and Gift Taxation," in William G. Gale, James R. Hines Jr., and Joel Slemrod, eds., *Rethinking Estate and Gift Taxation*, Brookings Institution Press, 2001.

Kopczuk, Wojciech, "A Note on Optimal Taxation in the Presence of Externalities," March 2000. University of Michigan, mimeo.

_ , "The Trick is to Live: Is the Estate Tax Social Security for the Rich?," November 2000. University of Michigan, mimeo.

Kotlikoff, Laurence J. and Avia Spivak, "The Family as an Incomplete Annuities Market," *Journal of Political Economy*, April 1981, *89* (2), 372–81.

Leonard, Daniel and Ngo Van Long, *Optimal Control Theory and Static Optimization in Economics*, Cambridge; New York and Melbourne: Cambridge University Press, 1992.

Mirrlees, James A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, April 1971, *38* (114), 175–208.

_ , "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, November 1976, *6* (2), 327–58.

Naito, Hisahiro, "Re-examination of Uniform Commodity Taxes under a Non-Linear Tax System and Its Implication for Production Efficiency," *Journal of Public Economics*, February 1999, *71* (2), 165–188.

Röell, Ailsa A., "A Note on the Marginal Tax Rate in a Finite Economy," *Journal of Public Economics*, November 1985, *28* (2), 267–72.

Sandmo, Agnar, "Optimal Taxation in the Presence of Externalities," *Swedish Journal of Economics*, 1975, *77* (1), 86–98.

Seade, Jesus, "On the Sign of the Optimum Marginal Income Tax," *Review of Economic Studies*, March 1982, *49* (4), 637–43.

Solon, Gary, "Intergenerational Income Mobility in the United States," *American Economic Review*, June 1992, *82* (3), 393–408.

Yaari, Menahem E., "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer," *Review of Economic Studies*, April 1965, *32* (2), 137–150.

Zimmerman, David J., "Regression toward Mediocrity in Economic Stature," *American Economic Review*, June 1992, *82* (3), 409–429.