

Estate and Capital Gains Taxation: Efficiency and Political Economy Considerations

Saku Aura*

Department of Economics
University of Missouri-Columbia
118 Professional Building
Columbia, MO 65211
auras@missouri.edu

April 27, 2004

Abstract

In this paper a simple dynastic overlapping-generations model with homogeneous agents is used to analyze the optimal use of capital income tax, labor income tax and estate tax. The results of this analysis add to the conventional wisdom about capital income taxation: while it is true that in the long run the estate tax rate should be set to zero, it is also true that other capital income taxation is a usable policy tool even in the steady state.

The other contribution of the paper is the building of a simple dynamic political economy model where the structure of capital taxes is determined. In a median-voter framework with no policy commitment, estate taxation is used too heavily as a capital-tax-revenue-collecting tool relative to the second-best optimum for the social planner.

JEL Codes: H21, H24

Keywords: Capital Income Taxation, Optimal Taxation, Political Economy

*I wish to thank Peter Diamond, Joseph Haslag, Vincenzo Galasso, Emek Basker and seminar participants at University of Wisconsin, Midwest Economics Association Annual Meeting and Missouri Economics Conference for helpful comments on this paper.

1 Introduction

This paper makes two points. The first point is about the optimal long run structure of capital taxes. One of the generally accepted results (due to Chamley 1986) of optimal income tax theory is that if the agents have infinite time-horizon, capital income should not be taxed in the long run. This paper partially challenges this conclusion in an dynastic overlapping-generations setting, where individuals are altruistic towards their offspring. This challenge comes from the fact that there are at least two types of individual capital taxes available for the policy-maker in the real world: estate/bequest/inheritance taxes and capital income taxes levied while individual is alive (such as interest income taxes and capital gains taxes, from now on capital income taxes).¹ When this generalization is taken into account, it is shown that the optimal policy in the long run involves no taxation of estates, but the capital income tax is a desirable policy tool in the long run. Thus the celebrated Summers' argument (Summers 1981) about infinitely compounded deadweight loss of capital income taxation applies only to taxation of estates. The second point of the paper is to illustrate a political economy mechanisms that biases the composition of capital income taxes towards taxation of estates and yields a positive estate tax rate even in the long run.

This first result is noted in the original Chamley article: given two capital income tax tools, only one of them would not be used in the steady-state. Chamley does not provide an analysis of the optimal tax rates in this situation since he dismissed the practicality of using separate tax rates for "capital income used to finance consumption and estates". It is argued here that this is practical and that the qualitative features of the optimal system are not very different from the "stylized" US personal capital income tax.² This paper takes this argument seriously and presents steady state Ramsey-tax rules for optimal labor income and capital income taxes. Interestingly, these Ramsey-rules are shown to be function of the parameters of the demand functions of the currently alive individuals only (so intergenerational

¹In this paper estate, inheritance, transfer tax and bequest tax are used interchangeably. The model presented here is not rich enough to distinguish between inheritance tax and estate tax.

²The current system in the US Federal taxes only realized capital gains and dividends; and leaves the estates untaxed (below certain threshold value). The optimal system in the model would not tax estates in the long run, and would tax dividends and realized capital gains, but would have a deduction for reinvested dividends and capital gains.

linkages do not matter), thus allowing potential empirical implementation.

The immediate policy conclusions of the result yielding zero estate taxes in the long run are limited. As in Chamley's model, it is entirely possible that to get the economy fast to the most desirable steady state a "bang-bang" solution using very high estate taxes (or high estate subsidies) could be called for. For a representative agent model, whether this requires a tax or a subsidy of estates depends on whether the economy starts with too much or too little capital. In a heterogeneous agent extension of the model, the distribution of the ownership of the capital plays also a role for the optimal use of estate taxes during the transition. Thus the main policy lesson of the paper is the potential desirability of other forms of capital taxation in even in the long run and not the estate tax result.

The other point that the paper makes is that in a median voter model without time commitment the outcome of voting over taxation produces the opposite result from our optimal tax program: estates will be taxed highly even in the long run. This result relies on two assumptions: no policy commitment and Markov voting strategies. The intuition behind this result is following: if the commitment period of economic policies is short enough, then most of the voters expect neither to receive nor to leave estates during the period of the effectiveness of the tax rate being currently decided. While most of the paper deals with certain life-time, this argument would also work with uncertain life time and learning about life expectancy environment: in any population there is typically a small minority of people who expect to be part of a intergenerational wealth transfer soon. Thus most young or healthy middle-age individuals should favor higher intergenerational transfer taxes if the period of policy commitment is short enough.

The second prediction of the political economy model is that institutions that make policy commitments possible (such as once-for-all voting, irrevocable legislations, constitutional rules or more general dynamic strategies between generations) reduce this bias in capital income taxation or make it disappear altogether.³

The rest of the paper is organized as follows. Section 2 presents the economic environment. Section 3 derives the normative results on the optimal

³The more general dynamic strategies refer to "folk-theorem" like implicit contracts between generations, where the a generation votes in "cooperative" way with future generations, since it would otherwise be punished by future generations. Galasso and Profeta (2002) provide a review of the use such strategies to explain issues surrounding social security in political economy models.

taxation. Section 4 presents a political economy model of capital income tax determination. Section 5 concludes. The proofs of the results in section 2 are provided in the Appendices.

2 Economic Environment

We consider a overlapping generations model, where a representative individual lives for two periods.

2.1 Individuals:

An individual, who is born on the period t has utility function

$$U^t(c_0^t, l^t, c_1^t, b^t) \equiv u(c_0^t, l^t, c_1^t) + T^t(b^t)$$

where c_0 and c_1 are the first and second-period consumption and l is the labor supply in the first period, b is the bequest that an individual leaves for his off-spring. $T(b)$ measures the (planning) utility value that the donor receives from leaving bequests of a size b . The bequest are assumed to be constrained non-negative. Labor supply is measured negatively using the Debreu sign convention, so a higher negative number in absolute value means more labor supply.⁴

2.2 Dynasties and population growth

The population is growing at rate n and new generation is born every period. The generations born on even-numbered periods form a linked dynasty, while the generations born on the odd-numbered periods form another linked dynasty.⁵

⁴This fairly typical normalization makes the use of standard tools of consumer theory (like Slutsky-equation) more straightforward.

⁵To be perfectly correct, we assume that each generation consists of continuum of individuals and each individual in that continuum is linked to $1+n$ atomistics individuals in offspring generations (to justify the price-taking behavior). This qualification is duly noted here, but in the interest of less cumbersome language the reference to continuum of agents is suppressed from now on.

The dynasty's behavior is assumed to be linked through perfect altruism as in Barro (1974), which means that when the non-negativity constraint of bequests is not binding, the dynasty behaves like it was maximizing a single dynastic utility function.⁶ Thus starting from period 0, the dynasty that has new members being born on the even-numbered periods has a dynastic utility function

$$U_E^0(\{c_0^t, l^t, c_1^t\}_{t \in E}) = \sum_{t \in E} \alpha^t u(c_0^t, l^t, c_1^t),$$

where E is the set of positive even numbers.⁷ Similarly the dynasty that has members born on odd numbered periods has a dynastic utility function

$$U_O^0(\{c_0^t, l^t, c_1^t\}_{t \in O}) = \alpha^{-1} \tilde{u}^{-1}(c_0^{-1}) + \sum_{t \in O} \alpha^t u(c_0^t, l^t, c_1^t),$$

where O is the set of positive odd numbers. The first term of that dynastic utility function refers to the initial old, whose last period consumption is the only planning-utility relevant part of their preferences for our optimal tax theory purposes. Note that the superscript refers always to the generation, not the time period. This means that c_1^t is the second life-period consumption of a generation born at time t . This consumption happens in period $t + 1$.

In our model we assume that the initial young have an endowment A_0 and the initial old an endowment A_{-1} of time zero consumption good. Thus given prices $\{q_0^t, q_1^t, q_w^t\}_{t=0}^{\infty}$ for the consumption while young, old and labor supply respectively at different periods, the dynastic indirect utility functions are $V_E(\{q_0^t, q_1^t, q_w^t\}_{t \in E}, q_0^0 A_0)$ and $V_O(q_1^{-1}, \{q_0^t, q_1^t, q_w^t\}_{t \in O}, q_1^{-1} A_{-1})$, where the last argument of the both indirect utility functions is the wealth term and the first set of arguments are the relevant price terms.⁸ The demand functions

⁶This assumption about dynastic behavior is a strong restriction about the utility of bequests ($T(B)$ -function) for an individual member of the dynasty.

⁷In the specification used consumption enters the utility in per capita terms (this will be important for the budget constraint) and thus the α 's are the product of pure time-preference and potentially a weighting factor $(1 + n)$, depending on whether the dynasties objective is to maximize the discounted sum of life-time utilities of representative (average) members of the generations or the discounted sum of the sums of life-time utilities of the members of the different generations. The qualitative results of the paper are unaffected by this distinction.

⁸ q_1^{-1} is the consumer price of consumption of the initial old.

Table 1: Demographic transitions.

Time-period	“Even-dynasty”	“Odd-dynasty”
t	Young	Old
t+1	Old	Young
t+2	Young	Old
t+3	Old	Young
t+4	Young	Old

and indirect utility functions of the two dynasties depend only even or odd-number generation prices, so we have can normalize one of the even and one of the odd period prices to any given value.

2.3 Production

The economy has a constant returns to scale production technology that can be characterized by the production function

$$y^t = f(k^t, -l^t),$$

where y^t is the output and k^t is capital, both normalized by the number of young agents. The negative sign on the labor input follows from our sign convention for the labor supply. This output can be either consumed (by old or young private agents or by government) or saved to yield capital for the next period. The convention on measuring capital in production is that capital is destroyed in the production, so the production function is the gross production function.

2.4 Taxes

In the analysis of optimal taxation it will be useful to distinguish between two types of tax rates for consumption and labor supply. The dynastic tax rate measures the tax-induced wedge between the marginal rate of substitution and marginal rate of transformation using period 0 consumption as the numeraire for both consumer and producer prices. The within generation tax rate on the other hand uses the first period consumption of the

Table 2: Prices and tax rates. Notation used: $f_k^i = f_k(k^i, -l^i)$, $f_l^i = f_l(k^i, -l^i)$ and depending on t is even or odd either $\hat{q} = q_0^0$ or $\hat{q} = q_1^{-1}$.

Variable	Producer Price	Consumer Price	Dynastic tax rate	Within generation tax rate
c_0^t	$p_0^t = \frac{1}{\prod_{i=1}^t f_k^i} (1+n)^t$	q_0^t	$\frac{p_0^0 q_0^t}{\hat{q} p_0^t} - 1$	0 (numeraire)
c_w^t	$p_w^t = \frac{f_l^t}{\prod_{i=1}^t f_k^i} (1+n)^t$	q_w^t	$\frac{p_0^0 q_w^t}{\hat{q} p_w^t} - 1$	$\frac{p_0^t q_w^t}{q_0^t p_w^t} - 1$
c_1^t	$p_1^t = \frac{1}{\prod_{i=1}^{t+1} f_k^i} (1+n)^t$	q_1^t	$\frac{p_0^0 q_1^t}{\hat{q} p_1^t} - 1$	$\frac{p_0^t q_1^t}{q_0^t p_1^t} - 1$
c_0^{t+2}	$p_0^{t+2} = \frac{1}{\prod_{i=1}^{t+2} f_k^i} (1+n)^{t+2}$	q_0^{t+2}	$\frac{p_0^0 q_0^{t+2}}{\hat{q} p_0^{t+2}} - 1$	$\frac{p_0^t q_0^{t+2}}{q_0^t p_0^{t+2}} - 1$

Notation used: $f_k^i = f_k(k^i, -l^i)$, $f_l^i = f_l(k^i, -l^i)$ and depending on t is even or odd either $\hat{q} = q_0^0$ or $\hat{q} = q_1^{-1}$. A separate line for c_0^{t+2} is included so the generational tax rate on bequests could be stated, i.e. what is the distortion between marginal utility of first-period consumption of two successive generations in a dynasty.

respective generation as the numeraire and is the natural concept to use for the long run analysis. The tax rates, consumer and producer prices of the goods and labor consumed at different times are given in Table 2. As an example for a result of the Table 2, the within generation t labor tax rate is calculated as

$$\left[\frac{\frac{q_w^t}{q_0^t} - \frac{p_w^t}{p_0^t}}{\frac{p_w^t}{p_0^t}} \right] = \frac{p_0^t}{q_0^t} \cdot \frac{q_w^t}{p_w^t} - 1.$$

By altering the consumer prices (after-tax prices) of labor supply and consumption at different periods of life the government can tax capital, bequests and labor at different rates. For the analysis of the optimal taxation, it will be easier to work in terms of after-tax prices q_0^t , q_w^t and q_0^t than in terms of tax rates for capital, labor and bequests.

The main reason for having this non-standard two dynasties structure is to be able to capture the fact that at any given time (period) in a real world, there are individuals who are “not about to die” and are at least partially saving for future consumption and not exclusively to leave estates.

This feature is the driving force of the positive political economy results of the second part of this paper and having two dynasties with alternating birth-periods in a two-period model is the simplest way to capture this concept. None of the substantive results normative optimal tax analysis (except for notation) would change by changing the demographic structure: essentially same results could be derived for a single dynasty where the second period expenditure would be divided between an inter vivos transfer from old agents to young agents and old-age consumption. Also the results would, for the most part, be robust to inclusion of truly multi-period (more than two period) lifetimes and many alternating dynasties.⁹

3 Optimal Taxation

In the representative agent framework, it is assumed that the government has to finance a revenue need by levying any combination of capital taxes, bequest taxes and labor income taxes. The government is assumed to have revenue requirement G per young individual. The government objective is to maximize the following welfare function

$$W = \beta^{-1} \tilde{u}^{-1}(c_0^{-1}) + \sum_{t \in O} \beta^t u(c_0^t, l^t, c_1^t) + \sum_{t \in E} \beta^t u(c_0^t, l^t, c_1^t).$$

That is, the government maximizes the discounted sum of the utilities of the individuals in the two dynasties. Note that the government discount factor is for the moment allowed to be different from the dynasties' discount factor. There are at least two ways to justify why the government should discount the future less than the dynasty itself that has some claim for reasonability:

1. The dynasty's utility function is the planning utility of the zero-period individual. When an individual leaves bequest, those bequest yield positive marginal utility to him, since otherwise he would have not left bequests. The bequest also affects the utility of the offspring. Thus a truly individualistic welfare function should take both of these effects into account and thus count the effect of the bequest both for the donor and the donee.

⁹Although the optimal capital tax rates and wage tax rates would be age-specific, if not constrained to be constant.

2. Constant discounting at any rate not arbitrarily close to one will lead to excessive discounting of the utility of generations far into future. This could be viewed as untenable ethical criterion.¹⁰

The position taken here is that while both of those claims have some intuitive relevance, the central case is assumed to be the one where the dynasty's planning utility coincides with the government's objective: extensions to the case of $\alpha \neq \beta$ are followed when the results presented allow simple provisions for this more general case.

The government is assumed to be able to use taxes on bequests, second-period consumption and labor or equivalently, by choosing the producer prices of first period consumption, second period consumption and labor.¹¹ The government is assumed to be able to follow optimal public debt policy and to be able to commit to a policy in the future. The latter is crucial, since as it is well-known the optimal tax policy would be time inconsistent.

3.1 No estate tax in the long run

In this section we derive the Chamley (1983) result in the context of our model. The approach followed here closely the approach taken in Erosa and Gervais (2001) applied to the specific demographic structure of our model (and the potential discrepancy between government's and dynasties objectives). The dual of governments problem can be written as¹²

¹⁰This is the view taken by Ramsey himself in his original contribution on the economics of savings (1928) when referring to treatment of utility of future generations: "One point should perhaps be emphasised more particularly; it is assumed that we do not discount later enjoyments in comparison with the earlier ones; a practice which is ethically indefensible and arises merely from the weakness of the imagination..." (p. 543).

¹¹With the normalization restriction that one even-period and one odd-period consumer price must be normalized to constants.

¹²In addition to the constraints and choice variables shown here, in principle the social planner chooses asset levels for each dynasty and the division of these assets between physical capital and public debt. The social planner furthermore faces the constraints governing the evolutions of these assets and public debt across generations and the constraint that the bond market must clear. Given government's ability to follow optimal debt policy none of this matters for the analysis and in order to save on notation, these additional variables and constraints are suppressed.

Also by choosing to present this problem as a consumption-labor tax problem we are precluding any taxation of the initial endowments and zero period returns to capital. This arbitrary choice of zero taxes on initial capital is made to make the optimal tax policy question non-trivial.

$$\begin{aligned} & \max_{\{q_0^t, q_1^t, q_w^t, k^t\}_{t=0}^\infty \setminus \{q_0^0, q_1^{-1}, k^0\}} W \\ & \text{subject to} \\ & (\alpha^t \lambda_t)_{t=0}^\infty : (1+n)k^{t+1} = f(k^t, -l^t) - c_0^t - \frac{c_1^{t-1}}{1+n} - G, \end{aligned}$$

where the initial capital level is given by k_0 , $q_0^0 = q_1^{-1} = 1$ (a normalization) and c_0^t, c_1^t and l^t are solutions to the two dynasties' optimization problems. For example, $c_0^0 = c_0^0(\{q_0^t, q_1^t, q_w^t\}_{t \in E}, q_0^0 A_0)$ is the demand for the consumption while young for the even-period dynasty's first generation and $c_1^{-1} = c_1^{-1}(1, \{q_0^t, q_1^t, q_w^t\}_{t \in O}, q_1^{-1} A_{-1})$ is the demand for consumption of the initial old by the odd-period dynasty. Instead of writing the period-specific government budget constraints, we use the set of aggregate resource use constraint (by Walras' law). The multiplier α^t in front of the Lagrange multiplier λ_t is used to normalize the Lagrange multiplier to be a current-value multiplier as opposed to present-value multiplier using the standard vocabulary of dynamic optimization. The choice of α instead of β as a normalizing factor will be useful in deriving the result on optimal taxes when the two differ. It is also worthwhile to note that while the dynastic optimization problem could be easily presented by a dynamic optimization problem, the government problem is not easily rendered to a dynamic optimization problem (because, e.g. the choice of q_w^{100} will affect the consumption and labor supply decisions in period 0). That justifies the presentation of the government's problem as a (countably) infinite-dimensional Lagrangian.¹³

In order to prove the result, it is easier to analyze the problem where the government controls directly the allocation and takes the individual dynastic optimization as a constraint to its problem. The dynastic budget constraint for can be manipulated as follows (using the even-period dynasty as an example):

¹³As in many other dynamic optimal taxation papers, the existence of a steady state to the optimal tax problem is postulated and not proved. Similarly, the existence of solution to government's problem is postulated following the tradition in the field, while acknowledging that the use of demand functions in the constraints might lead the problem to become a non-convex optimization problem.

$$\sum_{t \in E} (q_0^t c_0^t + q_1^t c_1^t + q_w^t l^t) = q_0^0 A_0 \Rightarrow$$

$$\sum_{t \in E} \alpha^t (u_0^t c_0^t + u_1^t c_1^t + u_l^t l^t) = u_0^0 A_0,$$

where e.g. u_l^0 is the partial derivative of the first period utility function with respect to labor and where the individual first-order conditions have been substituted to eliminate consumer prices. Now the governments problem can be restated as:

$$\max_{\{c_0^{t-1}, c_1^t, l^t, k^t\}_{t=0}^{\infty} \setminus \{k^0\}} W$$

subject to

$$(\alpha^t \lambda_t)_{t=0}^{\infty} : (1+n)k^{t+1} = f(k^t, -l^t) - c_0^t - \frac{c_1^{t-1}}{1+n} - G$$

$$(\theta_E) : \sum_{t \in E} \alpha^t (u_0^t c_0^t + u_1^t c_1^t + u_l^t l^t) = u_0^0 A_0$$

$$(\theta_O) : \alpha^{-1} \tilde{u}_0^{-1} c_0^{-1} + \sum_{t \in O} \alpha^t (u_0^t c_0^t + u_1^t c_1^t + u_l^t l^t) = \alpha^{-1} \tilde{u}_0^{-1} A_{-1}$$

Result 1: (No Estate Taxation in the Long Run) If $\alpha = \beta$ then $\frac{q_0^{t+2}}{q_0^t} = \frac{p_0^{t+2}}{p_0^t}$ in the steady state. If $\beta > \alpha$ then $\frac{q_0^{t+2}}{q_0^t} < \frac{p_0^{t+2}}{p_0^t}$.

Proof: See Appendix A.

Thus if the social discount rate (α) is equal to the dynasties' discount rate (β), then in the long run there should be no taxation of bequests. If $\beta > \alpha$ (for the potential reasons discussed earlier) then there should be an estate subsidy in the long run.

3.2 Lifetime Ramsey Formulae

In this section we focus on the case where $\beta = \alpha$ (private discount rate equals the social discount rate). The governments optimization problem can

be stated in terms of the two dynasties' indirect utility functions. In order to derive interesting ‘‘Ramsey Formulae’’ type of results, it is convenient to start the analysis by re-normalizing the consumer prices. In order to do so, we need to formalize the concept of being ‘‘close’’ to the steady state.

Let $\{(c_0^t, c_1^t, l^t, k^t)\}_{t=0}^\infty$ be the vector sequence of the endogenous variables in the economy. Pick $\varepsilon > 0$, where ε is an arbitrarily small number. Define $T(\varepsilon)$ as

$$T(\varepsilon) = \min \{s | t \geq s \implies \|(c_0^{t+2}, c_1^{t+2}, l^{t+2}, k^{t+2}) - (c_0^t, c_1^t, l^t, k^t)\| \leq \varepsilon\}.$$

The analysis that follows should be understood that we have chosen a time period $T = T(\varepsilon)$ as the period when the economy has approximately reached the steady state and proceed ignoring the fact that we are not exactly at the steady state (by choosing a smaller ε , the results become more exact).

A useful way to normalize the consumer prices is to normalize $q_0^T = p_0^T$ and $q_0^{T+1} = p_0^{T+1}$, while keeping the first period consumption as a numeraire for the producer prices (this normalization is why we needed to pick a fixed period T close enough to steady state). From Result 1 it follows that $q_0^t = p_0^t$ for $t \geq T$. For $t \geq T$ labor is taxed from the perspective of an individual who is born at period t if $q_w^t < p_w^t$ and $q_1^t > p_1^t$ means that second-period consumption is taxed.

Result 2: (Modified Lifetime Ramsey Formulae) At the steady state the optimal tax rates are implicitly defined by the following equations

$$\begin{cases} \theta + \varphi \frac{1}{c_0^i} \frac{\partial c_0^i}{\partial I} = \frac{(q_w^i - p_w^i)}{q_w^i} \tilde{\eta}_{c_0^i, q_w^i} + \frac{(q_1^i - p_1^i)}{q_1^i} \tilde{\eta}_{c_0^i, q_1^i} \\ \theta + \varphi \frac{1}{c_1^i} \frac{\partial c_1^i}{\partial I} = \frac{(q_w^i - p_w^i)}{q_w^i} \tilde{\eta}_{c_1^i, q_w^i} + \frac{(q_1^i - p_1^i)}{q_1^i} \tilde{\eta}_{c_1^i, q_1^i} \\ \theta + \varphi \frac{1}{l^i} \frac{\partial l^i}{\partial I} = \frac{(q_w^i - p_w^i)}{q_w^i} \tilde{\eta}_{l^i, q_w^i} + \frac{(q_1^i - p_1^i)}{q_1^i} \tilde{\eta}_{l^i, q_1^i}, \end{cases}$$

where e.g. $\tilde{\eta}_{c_0^i, q_w^i}$ is the compensated demand elasticity of first period consumption with respect to net wage rate, $\frac{\partial c_0^i}{\partial I}$ is the derivative of first period consumption with respect to unearned income (dynasty's starting wealth) and where θ and φ are constants independent of the time-period i .

Proof: See Appendix B.

These are the lifetime Ramsey formulae that apply in the steady state: the optimal tax rates for capital gains and labor can be solved by just using six

compensated elasticities and three income (or wealth) semi-elasticities.¹⁴ All the required parameters could be in principle estimated just using life-time price variation and consumption data. So as in the classic Ramsey rule, checking whether a given tax system satisfies the necessary conditions for the social optimum, one just needs to eliminate parameters (here two, in classical Ramsey one) from a system of equations and to see whether the implied equalities are satisfied.¹⁵

Example 1. *Equal income effects on second period consumption and labor:* Suppose that $\frac{1}{c_1^i} \frac{\partial c_1^i}{\partial I} = \frac{1}{l^i} \frac{\partial l^i}{\partial I}$. The last two equations of the system simplify to the classical Ramsey rule, where the only relevant price changes are the life time price changes. Note that this either requires that the income effects on labor supply and second period consumption are zero or that either leisure or second period consumption is an inferior good.

Example 2. *Within generation utility quasi-linear in first period consumption:* Suppose that $u(c_0, c_1, l) = g(c_0) + h(c_1) + j(l)$, where g , h and j are convex functions. Following similar line of reasoning as in the appendix it is easy to show that $\frac{\partial c_1^i}{\partial I} = \frac{\partial l^i}{\partial I} = \frac{\partial \bar{c}_1^i}{\partial q_w^i} = 0$. The tax rule simplifies to $\frac{(q_1^i - p_1^i)}{q_1^i} \tilde{\eta}_{c_0^i, q_1^i} = \frac{(q_w^i - p_w^i)}{q_w^i} \tilde{\eta}_{l^i, q_w^i}$, or to the “inverse elasticity” rule for the tax rate, which approximately states that the tax rate should be inversely related to the own-price elasticity of demand.¹⁶

Example 3: *Equal income effect on first and second period consumption:* $\frac{1}{c_0^i} \frac{\partial c_0^i}{\partial I} = \frac{1}{c_1^i} \frac{\partial c_1^i}{\partial I}$. This probably is the empirically most plausible special case. Now we have a modified tax rule that states

¹⁴These are referred as semielasticities since a dynasty may start with zero wealth in the first period.

¹⁵By choosing normalizations rightly, it is easy to show that the same equations (with normalization for q_0^i chosen so that $q_0^i = p_0^i$ in the optimum) apply also on the transition path to steady state. This is however, not very useful as a property, since we could only solve the ratio of the labor and capital gains tax based on the equation, the level would depend on the level of estate tax in place. So only in the steady state, where we know that estate tax is zero, can these equations truly be applied.

¹⁶This claim is not exact, since the denominator in the inverse elasticity rule is the consumer price and not the producer price, so the multipliers in front of the elasticities are tax shares of the final price and not tax rates.

Also note how the signing convention for labor supply works here, while the net wage compensated derivative of labor supply is negative, the elasticity is positive (unlike other own-price compensated elasticities, which are negative). This means that there will be a tax on labor ($q_w^i < p_w^i$) and a tax on second period consumption ($q_1^i > p_1^i$).

$$\frac{(q_w^i - p_w^i)}{q_w^i} \left(\tilde{\eta}_{c_0^i, q_w^i} - \tilde{\eta}_{c_1^i, q_w^i} \right) = \frac{(q_1^i - p_1^i)}{q_1^i} \left(\tilde{\eta}_{c_1^i, q_1^i} - \tilde{\eta}_{c_0^i, q_1^i} \right).$$

Assuming that first and second period consumption are Hicksian substitutes, this rule then states that if the compensated elasticity of first and second period consumption with respect to wage rate are equal, then the tax rate on second period consumption should be zero. Furthermore, if the first period consumption is more complementary to labor supply than second period consumption $\left(\tilde{\eta}_{c_0^i, q_w^i} > \tilde{\eta}_{c_1^i, q_w^i} \right)$ then the tax rate on second period consumption should be positive (otherwise there should be saving subsidy). This result is parallel to the classical result of Corlett and Hague (1953) in commodity taxation and the result of Atkinson and Sandmo (1980) in the case of capital income taxation in overlapping generations model with government debt.

Examples 1-3 are given to illustrate the easy applicability of the optimal tax rules. Example 2 provides an example of a situation where the capital gains tax must be strictly positive, since the implied equality (for positive government revenue requirement) implies that both wage and capital gains taxes should be used. Example 3 also, under plausible assumptions about complementarity between consumption and labor supply generates a tax on second period consumption. It is postulated that for many reasonable assumptions about the relevant elasticities the capital gains tax is strictly positive.

3.3 Qualitative properties of the optimal long run taxes

The qualitative properties of optimal steady state taxes have an appealing feature: they resemble an existing real-world structure of capital and income taxes. The optimal system in a two-period world could be implemented as the following tax system:

1. Labor income is taxed.
2. Realized capital gains are taxed in the second period. In a multi-asset world, if assets are sold, then reinvested gains are deducted from the tax base. Dividends and interest payments are taxed, but the amount of that is reinvested is deducted from the tax base.

3. Estates are not taxed and there is a basis step-up at death. This means that the offspring pays capital gains on the difference of the value of the assets at the time of sale and their value at the time of inheritance.

This description comes very closed to the current federal system in the US of capital, labor and estate taxes that applies to most tax-payers (below the threshold of estate taxation). The only difference is the reinvestment deduction: the optimal tax system here calls for only the asset income that is consumed to be taxed.

There are two questions that come immediately mind in practical implementation or relevance of this tax system. The first is the administrative feasibility of the reinvestment reduction. This should not be prohibitively costly, the current tax-system could be amended to capture this feature by withdrawing taxes from capital gains, interest and dividend payments at the source and letting the individual tax payer claim the reinvestment deductions at tax filing time.

The second issue arises from the applicability of the two-period result to more finely divided (and more realistic) individual life-spans. Adding more periods would make the technical analysis either more complicated or the resulting tax system more complicated, while still keeping the validity of the main arguments. With individuals living for more than two periods (and potentially working for more than one period) the optimal tax system would with high likelihood imply labor and capital income tax rates that vary according to the age of the tax-payer. If this is not viewed politically feasible, then the analysis should take into account the constant-rate constraint while deriving the optimal tax formula. While this extension is not followed here, it is easy to see that the main arguments of the paper would not change if this constraint is taken into account: the optimal long run system would still have no estate tax, but would have positive labor and capital income taxes.

3.4 Heterogenous population and redistribution

It is a straightforward extension of the analysis above to extend the situation to handle heterogeneous population (where the types vary according to their productivity) when a demogrant (lump sum amount of money) is paid to each individual in the first period of their lives (so taxation is due to

redistributive motive, and the taxes would be used even if the government has no expenditure requirement). As in the previous analysis the long run optimum does not feature taxation of estates.

The modified Ramsey-rule for taxation of capital and labor income derived above would now be replaced by a modified many-person Ramsey rule (Diamond, 1975), where the relative induced distortion on capital income and labor would also depend on the whether the dynasty with high marginal social utility of income (typically the low productivity type) gets relatively more income from sales of capital in the second period or from first-period labor supply. What is the net effect of a redistributive motive to the relative tax rates on capital income and labor income on the steady state outcome is not clear, since the long-run source of inequality is the differing of labor productivity and not capital income if the estates are taxed at sufficiently high rate along the transition path. The main effect of redistributive motive is that it most likely (if we start from a situation where the high labor productivity dynasty is the one with higher starting wealth, an empirically plausible assumption), then the redistributive motive would lead to a higher taxation of estates along the transition path to correct the wealth distribution.

3.5 Uncertain lifetimes and estate taxation

The optimal income taxation model of this paper relies on one assumption about the bequests, that they are intended and that individuals know with certainty their date of the death. With a slight additions to notation and details, the model can capture the taxation of bequest of when the individuals have perfect access to annuity markets and where they rationally annuitize their wealth to finance their end-of-life consumption and leave only intended bequest. The analysis of partially accidental bequests would complicate the steady-state analysis considerably even when starting from representative agent world (at date zero), since one would need to find an ergodic distribution of bequests to analyze the steady state.¹⁷

The polar opposite case of purely-accidental bequests in a representative agent world is relative simple to handle. In that case, the accidental bequest should be taxed away under most plausible assumptions for purely social insurance reasons. This conclusion holds even in a steady-state model like the

¹⁷It can be shown that such ergodic distribution must have unbounded support.

one presented in this paper and precisely because the unintended bequest are taxed at 100% the analysis of a steady state model remains simple. However, the taxation of unintended bequests becomes a harder question immediately when one allows for heterogeneity of individuals in any dimensions. In their “medium-run” analysis Blumkin and Sadka (2002) show that if other taxes (like labor income tax) are used for redistributive purposes, it is possible that the purely accidental bequests are not taxed at 100% in the social optimum, because of the interaction of the accidental bequest with the redistribution through labor income taxes.¹⁸ As a conjecture it is stated here, that in a steady-state analysis of optimal taxation with heterogeneous population the same results of potentially less than 100% tax on purely accidental bequests should hold. It should be stressed that the main result of this paper, the potential desirability of taxation of other capital income will not change even if some bequests are accidental.

4 Political Equilibrium

In this section we seek to answer the following question: whether a standard political economy model produces as an outcome a tax system resembling the social optimal system derived in the previous section. The answer to this question is negative and this is shown to be driven by lack of policy commitment in a typical voting game. The point of this discussion is not to reinvent the result that without policy commitment the tax system is biased towards taxing capital in general. Instead the main point of this section is to show that the composition of the capital taxes in our political economy model is biased towards taxation of estates even in the long run.

4.1 Model of Political Economy

Our framework of the political economy is following. The economic environment is the model of the previous section with a continuum of representative agents. To present the main result, we start with assumption of inelastic labor supply; this assumption will later be relaxed. The government then uses capital income tax and bequest tax to finance the revenue requirement. We

¹⁸The result of less than 100% estate tax requires assumptions about the relative magnitudes of the elasticities of labor supply of high and low skill individuals evaluated at the social optimum.

assume that the government must meet its budget requirement each period. Note that in this simplified model our optimal taxation analysis would imply that in steady state all the tax revenue should be collected through capital income taxes.

Each period, the government holds a referendum on the rate of estate tax. The rate chosen will be the rate preferred by the median voter. The required capital income tax is then chosen to balance the budget. We also assume that the agents fully understand the structure and are able to predict the tax rates and the whole future time-path of political-economic equilibrium perfectly as a function of today's choices.

Table 3: Timing of events through a life span of an individual.

Time period	Phase	Event
t	1	Birth and inheritance
	2	Referendum and tax determination
	3	Economic choices and consumption
$t + 1$	4	Demographic transition in the other dynasty
	5	Second referendum
	6	Economic choices and consumption
$t + 2$	7	Death and inheritance passed to next generation

The life-span of individual is explained in the Table 3. At phase 1 individual is born and she receives an inheritance from the previous generation. At phase 2 the referendum is held on the applicable estate tax rate applied to the estates left by the current old and the capital income tax on the present young is calculated. At phase 3 savings and estates are chosen and consumption and labor supply occurs. At phase 4 the previous old (from the unrelated dynasties) die and pass their estates are passed to newly born agents. At phase 5 new referendum is held. Phase 5 is identical in setting to phase 2, except now the individual born at time period t is an old individual. At phase 6 she chooses second period consumption and bequests. Consumption (and labor supply by the new young) also happen at phase 6. At phase 7 the individual dies and passes her estate to her heirs.¹⁹

¹⁹The particular timing of the events is chosen so that the recipients of the inheritance don't get to vote. Allowing for the different timing, we can always restore the result by dividing the life-span of individual to more than two periods (and taking a stand on the age of the receipt of inheritances) and assuming a shorter period of policy commitment. These

Assume that initial capital holdings are s_0^0 and s_1^0 , where subscript denotes period of life and superscript calendar time. Let $\{t_b^k\}_{k=0}^\infty$ be the sequence of tax rates (given history up to k) that is most preferred by the respective young voter in each period.

Result: The sequence outcome of the voting game is $\{t_b^k\}_{k=0}^\infty$.

Proof: With positive population growth, the median voter is always a young agent.

This result means that the estate taxes are likely to be high in each period. The reason for this (barring the very unlikely scenario that general equilibrium effects and dynamic feedback between elections overturn this straightforward logic) is very simple: young voters prefer the tax to be paid by someone else (old voters).

4.2 Discussion of the political equilibrium

The main driving assumptions in the political equilibrium in the previous section were the importance of the median voter, one-dimensional policy space and lack of policy commitment. We will discuss each of these below.

4.2.1 Median Voter Assumption

The median-voter model is not a necessary assumption to get the qualitative result of the previous section. In the median voter model the policy is chosen by (repeated) referendum. A standard alternative to the median-voter model is the probabilistic-voting model (e.g. Persson and Tabellini 2002), where two parties compete for votes. Using the most standard assumptions of probabilistic voting (neither party is inherently preferred on average by any group and both young and old are equally mobile as voters), the driving force of the result above (the dominance of the young voters' objective over the old voters' in the outcome) disappears. However, due to the lack of policy commitment, the equilibrium still features positive taxation of estates in the long run. This follows since the social welfare function maximized by the probabilistic voting mechanism would weight the utility of both generations

added complications will still capture the real world fact that the majority of population is not about to receive nor leave bequest with a high probability during any given election cycle.

alive, and the young always prefer to tax estates and not their own savings. Thus under this alternative modelling assumption the stark result of the previous section is lost, but the bias away from the optimal taxation remains.

4.2.2 Two-dimensional policy space: Labor taxes

With elastic labor supply, labor taxes can be included among the available policy tools without introducing lump-sum taxes. This extensions does not change the qualitative conclusions: the bias towards high estate taxation remains.

The median voter model can be extended to handle this situation through the use of Structure-Induced Equilibrium (Persson and Tabellini 2002).²⁰ Without loss of generality we can frame the referenda so that they are held over the tax rate for capital gains and the tax rate for estates. This means that the wage tax rate will be set to balance the budget.

A simple version of structure-induced equilibrium in this situation would satisfy two conditions:

- 1) the tax rate on capital gains is the tax rate preferred by the median voter holding the tax rate for estates constant and
- 2) the tax rate for estates is the tax rate preferred by the median voter holding the tax rate for capital gains constant.

In principle the identity of the median voter in structure-induced equilibrium models can be different depending on the variable that the referendum is referring to. However, in our setup, the median voter is always a young voter since young are a larger group. Young voters like lower capital gains taxes than the old and they like higher estate taxes than the old. Thus, the result survives, young voters can use an estate tax strategy similar to the one-dimensional case.

With probabilistic voting, the remarks are similar the previous case: in this case, the young voter's objective does not completely outweigh the old voter's objective in the outcome, but without policy commitment the young

²⁰An example of a voting process that leads to structure-induced equilibrium is a process where referenda over the two issues (two tax rates here) alternate and the voters vote non-strategically (myopically) on the current issue holding the other variable (tax rate) constant at the current level in forming their preferences.

voters preference for positive estate taxation will still have an influence even in the long run.

4.2.3 Policy commitment

If the sequential voting game is replaced by once-for-all voting in the first period, where the voters actually choose the optimal path of the tax rates from date zero to infinity then the main result of political economy part will change. The result that estates are taxed in the long run in the sequential voting equilibrium is a function of the fact that the tax rates are subject to re-optimization (or change through the political process) every period. While different specifications of the once-for-all voting game would weight the odd and even-period dynasties differently (so the welfare function maximized would have different welfare weights), the end result would be the same as in our optimal tax problem: estates would not be taxed in the long run.

Assuming that the perfect policy commitment is not possible it becomes interesting to speculate on the effect of the length of the policy commitment to the policy outcome. The shorter the time span between election periods becomes (assuming that tax policy is potentially re-optimized every election cycle), the more overwhelmingly is the median voter placed among voters who prefer high estate taxes. Similarly in the probabilistic-voting model the policy objective gets more biased towards high estate taxation when the commitment period shortens. Thus the tendency for high estate taxation should be larger when the tax rates are expected to be revised more often.

5 Conclusions

The main contribution of this paper is the result that optimal taxation includes some taxation of capital income even in the long run in a dynastic model with fully altruistic preferences. The reason for this is that if the tax authority can treat realized capital gains and bequest differently, then the realized capital gains should be taxed even in the long run (except when the realized gains are reinvested). Thus the main policy conclusion of the paper is that beyond the transition path argument for capital income taxation (mostly concerned with wealth distribution) there is a separate economic case to be made for more limited capital income taxation even in the most capital income tax hostile model. Perhaps most interestingly, the argument

for capital income taxation given in this paper has almost nothing to do with wealth redistribution nor has it anything to do with the correcting for dynamic inefficiency due to a sub-optimal level of capital in the economy that can arise in overlapping generations setting.

The political economy part of the model explains why there might be a political bias in democracies to tax estates too heavily compared to other capital income. This is shown to arise due to lack of intertemporal policy commitment. Thus constitutional rules that provide more commitment in tax policy across time (such as supermajority rules to change the relative structure of taxation) could potentially be beneficial by moving the actual tax policy closer to the optimal full-commitment tax policy.

References

- [1] Atkinson, A.B., Sandmo, A., 1980. Welfare Implications of Taxation of Savings. *The Economic Journal*, 90:359.
- [2] Aura, S., Diamond, P., Geanakoplos J, 2002. Savings and Portfolio Choice in a Two-Period Two-Asset Model. *American Economic Review*, 92:4.
- [3] Barro, R., 1974. Are Government Bonds Net Wealth? *The Journal of Political Economy*, 82:6.
- [4] Blumkin, T., Sadka, E., 2003. Estate taxation. *Journal of Public Economics*, (forthcoming).
- [5] Chamley, C., 1986. Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica*, 54(3).
- [6] Corlett, W.J., Hague, D.C., 1953. Complementarity and the Excess Burden of Taxation. *Review of Economic Studies* 21.
- [7] Deaton, A., 1986. Demand Analysis. In: Grilliches, Z., Intriligator, M.(eds.) *Handbook of Econometrics*, Volume III, Elsevier.
- [8] Diamond, P., 1975. A Many-Person Ramsey Tax Rule. *Journal of Public Economics*, 4.
- [9] Erosa, A., Gervais, M., 2001. Optimal Taxation in Infinitely-Lived Agent and Overlapping Generations Models: A Review. *Federal Reserve Bank of Richmond Economic Quarterly*, 87(2).

- [10] Galasso, V., Profeta P. 2002. Politico-Economic Models of Social Security. *European Journal of Political Economy*, 18:1.
- [11] Persson, T., Tabellini G., 2002. *Political Economics - Explaining Economic Policy*. MIT Press.
- [12] Ramsey, F.P, 1928. A Mathematical Theory of Saving. *The Economic Journal*, 38:152.
- [13] Summers, L., 1981. Capital Taxation and Accumulation in a Life Cycle Growth Model. *American Economic Review*, 71:4.

APPENDIX A: Proof of Result 1

Result 1: (No Estate Taxation in the Long Run) If $\alpha = \beta$ then $\frac{q_0^{t+2}}{q_0^t} = \frac{p_0^{t+2}}{p_0^t}$ in the steady state. If $\beta > \alpha$ then $\frac{q_0^{t+2}}{q_0^t} < \frac{p_0^{t+2}}{p_0^t}$.

Proof:

Case 1 ($\alpha = \beta$): The first order condition with respect to k^t is

$$\alpha^{t-1} \lambda_{t-1} (1+n) = \alpha^t \lambda_t f_k^t,$$

where $f_k^t = \frac{\partial f(k^t, -l^t)}{\partial k^t}$ is the marginal productivity of capital. Advancing that by one period and substituting gives:

$$\alpha^t \lambda_t (1+n)^2 = \alpha^{t+2} \lambda_{t+2} f_k^{t+1} f_k^{t+2}.$$

In steady state the endogenous variables are time-invariant modulo 2. This means, that if the “odd-period” and “even-period” dynasties start at different endowment levels, then the steady state could imply a cycling with periodicity 2 of the endogenous variables (so each consecutive members of each dynasty would have the same outcome variables, but this equality does not have to hold between the two dynasties).²¹

The first order conditions with respect to c_0^t and c_0^{t+2} are (without loss of generality we can concentrate on the even-numbered dynasty):

²¹Given that redistribution between even and odd-period dynasties is costly in our model, it is an open question whether the steady state has the property that all the even and odd-numbered dynasties will have same outcome variables.

$$\begin{cases} \beta^t u_0^t + \alpha^t \lambda_t + \theta_E \alpha^t \frac{\partial}{\partial c_0^t} (u_0^t c_0^t + u_1^t c_1^t + u_l^t l^t) = 0 \\ \beta^{t+2} u_0^{t+2} + \alpha^{t+2} \lambda_{t+2} + \theta_E \alpha^{t+2} \frac{\partial}{\partial c_0^{t+2}} (u_0^{t+2} c_0^{t+2} + u_1^{t+2} c_1^{t+2} + u_l^{t+2} l^{t+2}) = 0. \end{cases}$$

In steady state these imply the following law of motion for λ :

$$\lambda_{t+2} - \lambda_t = \frac{\beta^t}{\alpha^t} \left(\frac{\beta^2}{\alpha^2} - 1 \right) u_0^t.$$

If the social discount equals the private discount rate ($\alpha = \beta$), then $\lambda_{t+2} = \lambda_t$ in steady state. Thus the steady state version of the law first-order condition for capital on becomes:

$$\alpha^t \tilde{\lambda} (1+n)^2 = \alpha^{t+2} \tilde{\lambda} f_k^t f_k^{t+2} \implies f_k^t f_k^{t+2} = \frac{(1+n)^2}{\alpha^2},$$

where $\tilde{\lambda}$ is the steady state value of λ for the relevant dynasty.

This is the regular Ramsey condition (the growth theory Ramsey-condition) or the modified golden rule. The steady state dynastic first-order condition

$$\frac{\partial u^t / \partial c_0}{q_0^t} = \alpha^2 \frac{\partial u^{t+2} / \partial c_0}{q_0^{t+2}}$$

yields

$$\frac{q_0^{t+2}}{q_0^t} = \alpha^2.$$

The steady state producer price ratio between c_0^t and c_0^{t+2} is

$$\frac{p_0^{t+2}}{p_0^t} = \frac{\frac{1}{\prod_{i=1}^{t+2} f_k^i} (1+n)^{t+2}}{\frac{1}{\prod_{i=1}^t f_k^i} (1+n)^t} = \frac{(1+n)^2}{f_k^{i+1} f_k^{i+2}} = \alpha^2,$$

where the last equality follows from the modified golden rule.

Case 2 ($\beta > \alpha$): Using the steady state laws of motion for λ it follows immediately that $f_k^t f_k^{t+2} > \frac{(1+n)^2}{\alpha^2}$. Applying this inequality to the ratio of producer prices calculations yields the result.

APPENDIX B: Proof of Result 2

Result 2: (Modified Lifetime Ramsey Formulae) At the steady state the optimal tax rates are implicitly defined by the following equations

$$\begin{cases} \theta + \varphi \frac{1}{c_0^i} \frac{\partial c_0^i}{\partial I} = \frac{(q_w^i - p_w^i)}{q_w^i} \tilde{\eta}_{c_0^i, q_w^i} + \frac{(q_1^i - p_1^i)}{q_1^i} \tilde{\eta}_{c_0^i, q_1^i} \\ \theta + \varphi \frac{1}{c_1^i} \frac{\partial c_1^i}{\partial I} = \frac{(q_w^i - p_w^i)}{q_w^i} \tilde{\eta}_{c_1^i, q_w^i} + \frac{(q_1^i - p_1^i)}{q_1^i} \tilde{\eta}_{c_1^i, q_1^i} \\ \theta + \varphi \frac{1}{l^i} \frac{\partial l^i}{\partial I} = \frac{(q_w^i - p_w^i)}{q_w^i} \tilde{\eta}_{l^i, q_w^i} + \frac{(q_1^i - p_1^i)}{q_1^i} \tilde{\eta}_{l^i, q_1^i}, \end{cases}$$

where e.g. $\tilde{\eta}_{c_0^i, q_w^i}$ is the compensated demand elasticity of first period consumption with respect to net wage rate, $\frac{\partial c_0^i}{\partial I}$ is the derivative of first period consumption with respect to unearned income (dynasty's starting wealth) and where θ and φ are constants independent of the time-period i .

Proof: The government's optimal tax problem can be written as:

$$\max_I V_E(\{q_0^t, q_1^t, q_w^t\}_{t \in E}, q_0^0 A_0) + V_O(q_1^{-1}, \{q_0^t, q_1^t, q_w^t\}_{t \in O}, q_1^{-1} A_{-1})$$

subject to

$$(\beta^t \lambda_t)_{t=0}^\infty : (1+n)k^{t+1} = f(k^t, -l^t) - c_0^t - \frac{c_1^{t-1}}{1+n} - G,$$

where $I = q_1^{-1} \cup \{q_0^t, q_1^t, q_w^t, k^t\}_{t=0}^\infty \setminus \{q_0^T, q_0^{T+1}, k^0\}$ is the set of instruments available for the government.

Without any loss of generality, we can concentrate on the optimal setting of consumer prices for the “even-period” dynasty (the first order conditions for even and odd-period consumer prices will be linked only through capital levels, since the consumer demands depend only on the consumer prices relevant to the dynasty). The first order conditions with respect to the triple (q_0^i, q_1^i, q_w^i) (for $i \notin \{0, T, T+1\}$) are:

$$\begin{aligned}
(q_0^i) : & -\mu_E c_0^i + \sum_{t \in O} \left[\beta^t \lambda_t \left(f_l^t \frac{\partial l^t}{\partial q_0^i} + \frac{\partial c_0^t}{\partial q_0^i} \right) + \frac{\beta^{t+1} \lambda_{t+1} \partial c_0^t}{1+n} \frac{\partial c_0^t}{\partial q_0^i} \right] = 0 \\
(q_1^i) : & -\mu_E c_1^i + \sum_{t \in O} \left[\beta^t \lambda_t \left(f_l^t \frac{\partial l^t}{\partial q_1^i} + \frac{\partial c_0^t}{\partial q_1^i} \right) + \frac{\beta^{t+1} \lambda_{t+1} \partial c_0^t}{1+n} \frac{\partial c_0^t}{\partial q_1^i} \right] = 0 \\
(q_w^i) : & -\mu_E c_w^i + \sum_{t \in O} \left[\beta^t \lambda_t \left(f_l^t \frac{\partial l^t}{\partial q_w^i} + \frac{\partial c_0^t}{\partial q_w^i} \right) + \frac{\beta^{t+1} \lambda_{t+1} \partial c_0^t}{1+n} \frac{\partial c_0^t}{\partial q_w^i} \right] = 0,
\end{aligned}$$

where μ_E is the marginal utility of income and f_l^t is the marginal product of labor at period t . In what follows, we will demonstrate how to modify the first-order condition with respect to q_0^i to yield a ‘‘Ramsey-type’’ first-order condition. The required steps for q_1^i and q_w^i are identical.

1. Substituting for λ s and substituting in producer prices: The first order conditions for the capital levels imply (using repeated substitutions) that $\lambda_t = \lambda_0 \beta^{-t} \frac{1}{\prod_{k=1}^t f_k^i} (1+n)^t$. Using our definition of producer prices, the first-order condition becomes:

$$-\mu_E c_0^i + \lambda_0 \sum_{t \in O} \left[p_w^t \frac{\partial l^t}{\partial q_0^i} + p_0^t \frac{\partial c_0^t}{\partial q_0^i} + p_1^t \frac{\partial c_0^t}{\partial q_0^i} \right] = 0.$$

2. (First standard Ramsey step) Adding and subtracting

$c_0^i \left(\sum_{t \in O} \frac{\partial l^t}{\partial I} + t_0 \frac{\partial c_0^t}{\partial I} + \frac{\partial c_0^t}{\partial I} \right)$ to the equation and then using Slutsky Equation. This yields:

$$\left(\sum_{t \in O} \frac{\partial l^t}{\partial I} + \frac{\partial c_0^t}{\partial I} + \frac{\partial c_1^t}{\partial I} - \mu_E^i \right) c_0^i + \lambda_0 \sum_{t \in O} \left[p_w^t \frac{\partial \tilde{l}^t}{\partial q_0^i} + p_0^t \frac{\partial c_0^t}{\partial q_0^i} + p_1^t \frac{\partial c_1^t}{\partial q_0^i} \right] = 0,$$

where e.g. $\frac{\partial \tilde{l}^t}{\partial q_0^i}$ means the utility compensated (Hicksian) demand derivative of l^t with respect to q_0^i .

3. (Second standard Ramsey step) Adding $\sum_{t \in O} \left[q_w^t \frac{\partial \tilde{l}^t}{\partial q_0^i} + q_0^t \frac{\partial c_0^t}{\partial q_0^i} + q_1^t \frac{\partial c_1^t}{\partial q_0^i} \right] = 0$ to the equation and using symmetry of the Slutsky matrix to arrive

at standard dynastic Ramsey Rule.

$$\begin{aligned} \hat{\theta}c_0^i + \lambda_0 \sum_{t \in O} \left[(q_w^t - p_w^t) \frac{\partial \tilde{l}^t}{\partial q_0^i} + (q_0^t - p_0^t) \frac{\partial \tilde{c}_0^t}{\partial q_0^i} + (q_1^t - p_1^t) \frac{\partial \tilde{c}_0^t}{\partial q_0^i} \right] &= 0 \\ \hat{\theta}c_0^i + \lambda_0 \sum_{t \in O} \left[(q_w^t - p_w^t) \frac{\partial \tilde{c}_0^t}{\partial q_w^t} + (q_0^t - p_0^t) \frac{\partial \tilde{c}_0^t}{\partial q_0^t} + (q_1^t - p_1^t) \frac{\partial \tilde{c}_0^t}{\partial q_1^t} \right] &= 0 \\ \theta &= \frac{\sum_{t \in O} \left[(q_w^t - p_w^t) \frac{\partial \tilde{c}_0^t}{\partial q_w^t} + (q_0^t - p_0^t) \frac{\partial \tilde{c}_0^t}{\partial q_0^t} + (q_1^t - p_1^t) \frac{\partial \tilde{c}_0^t}{\partial q_1^t} \right]}{c_0^i}, \end{aligned}$$

where the $\hat{\theta}$ and θ are constants independent of c_0^i . The last equation (there is a corresponding equation for c_0^t , c_1^t and l^t) is the standard dynastic Ramsey-rule, it calls for equal relative compensated changes in demand due to tax system, as long as the compensated demands are well approximated by first-order Taylor expansion around the no-tax point (assuming also that the producer prices are unchanged by the tax-system).

4. Using the proposition proved in the appendix C this can be rewritten as (for the notation, see appendix C):

$$\begin{aligned} \theta &= \frac{(q_w^i - p_w^i) \frac{\partial \tilde{c}_0^i}{\partial q_w^i} + (q_0^i - p_0^i) \frac{\partial \tilde{c}_0^i}{\partial q_0^i} + (q_1^i - p_1^i) \frac{\partial \tilde{c}_0^i}{\partial q_1^i}}{c_0^i} + \\ &\frac{\frac{\partial c_0^i}{\partial I} \sum_{t \in O \setminus \{i\}} \left[(q_w^t - p_w^t) \left(\frac{\partial \tilde{E}^i}{\partial q_w^t} / \frac{\partial E^i}{\partial I} \right) + (q_0^t - p_0^t) \left(\frac{\partial \tilde{E}^i}{\partial q_0^t} / \frac{\partial E^i}{\partial I} \right) + (q_1^t - p_1^t) \left(\frac{\partial \tilde{E}^i}{\partial q_1^t} / \frac{\partial E^i}{\partial I} \right) \right]}{c_0^i} \\ \implies \theta + \varphi^i \frac{1}{c_0^i} \frac{\partial c_0^i}{\partial I} &= \frac{(q_w^i - p_w^i) \frac{\partial \tilde{c}_0^i}{\partial q_w^i} + (q_0^i - p_0^i) \frac{\partial \tilde{c}_0^i}{\partial q_0^i} + (q_1^i - p_1^i) \frac{\partial \tilde{c}_0^i}{\partial q_1^i}}{c_0^i}, \end{aligned}$$

where φ^i is a generation specific constant.

Optimal tax rules can be characterized by the following sets of equations (for $i \notin \{0, T, T+1\}$):

$$\left\{ \begin{array}{l} \theta + \varphi^i \frac{1}{c_0^i} \frac{\partial c_0^i}{\partial I} = \frac{(q_w^i - p_w^i) \frac{\partial \tilde{c}_0^i}{\partial q_w^i} + (q_0^i - p_0^i) \frac{\partial \tilde{c}_0^i}{\partial q_0^i} + (q_1^i - p_1^i) \frac{\partial \tilde{c}_0^i}{\partial q_1^i}}{c_0^i} \\ \theta + \varphi^i \frac{1}{c_1^i} \frac{\partial c_1^i}{\partial I} = \frac{(q_w^i - p_w^i) \frac{\partial \tilde{c}_1^i}{\partial q_w^i} + (q_0^i - p_0^i) \frac{\partial \tilde{c}_1^i}{\partial q_0^i} + (q_1^i - p_1^i) \frac{\partial \tilde{c}_1^i}{\partial q_1^i}}{c_1^i} \\ \theta + \varphi^i \frac{1}{l^i} \frac{\partial l^i}{\partial I} = \frac{(q_w^i - p_w^i) \frac{\partial \tilde{l}^i}{\partial q_w^i} + (q_0^i - p_0^i) \frac{\partial \tilde{l}^i}{\partial q_0^i} + (q_1^i - p_1^i) \frac{\partial \tilde{l}^i}{\partial q_1^i}}{l^i}. \end{array} \right.$$

Considering this expression for the steady state ($i > T + 1$) simplifies it to:

$$\left\{ \begin{array}{l} \theta + \varphi \frac{1}{c_0} \frac{\partial c_0}{\partial I} = \frac{(q_w - p_w) \frac{\partial \tilde{c}_0}{\partial q_w} + (q_1 - p_1) \frac{\partial \tilde{c}_0}{\partial q_1}}{c_0} \\ \theta + \varphi \frac{1}{c_1} \frac{\partial c_1}{\partial I} = \frac{(q_w - p_w) \frac{\partial \tilde{c}_1}{\partial q_w} + (q_1 - p_1) \frac{\partial \tilde{c}_1}{\partial q_1}}{c_1} \\ \theta + \varphi \frac{1}{l} \frac{\partial l}{\partial I} = \frac{(q_w - p_w) \frac{\partial \tilde{l}}{\partial q_w} + (q_1 - p_1) \frac{\partial \tilde{l}}{\partial q_1}}{l} \end{array} \right.$$

or to

$$\left\{ \begin{array}{l} \theta + \varphi \frac{1}{c_0} \frac{\partial c_0}{\partial I} = \frac{(q_w - p_w)}{q_w} \tilde{\eta}_{c_0, q_w} + \frac{(q_1 - p_1)}{q_1} \tilde{\eta}_{c_0, q_1} \\ \theta + \varphi \frac{1}{c_1} \frac{\partial c_1}{\partial I} = \frac{(q_w - p_w)}{q_w} \tilde{\eta}_{c_1, q_w} + \frac{(q_1 - p_1)}{q_1} \tilde{\eta}_{c_1, q_1} \\ \theta + \varphi \frac{1}{l} \frac{\partial l}{\partial I} = \frac{(q_w - p_w)}{q_w} \tilde{\eta}_{l, q_w} + \frac{(q_1 - p_1)}{q_1} \tilde{\eta}_{l, q_1}, \end{array} \right.$$

where e.g. $\tilde{\eta}_{c_0, q_w}$ is the compensated demand elasticity of first period consumption with respect to net wage rate.

APPENDIX C: Implications of intertemporal separability on demand derivatives

The assumption of intergenerational additive separability has strong implications on the properties of the demand functions (through availability of two-stage budgeting rules). This appendix derives an implication of the additive separability that is used to derive the life-time Ramsey-rules. The

proposition in this appendix is a straightforward extension of the Proposition 8 in Aura, Diamond and Geanakoplos (2002).²²

Assumptions: Let consumer's utility function be $U(\{x^i\}_{n=0}^N) = \sum_{n=0}^N g^n(x^n)$, where $N \leq \infty$ and each x^n is a vector of consumption goods (with potentially differing number of consumption goods for different values of n). Let the prices of the good be given by the vectors given by set of price vectors $\{p^n\}_{n=0}^N$. Let x^{ij} be the j 'th element of the i 'th consumption vector and p^{ij} be the corresponding price. Let the income of the consumer be I . Let $\frac{\partial x^{ij}}{\partial p^{lm}}$ be the Marshallian (uncompensated) demand derivative and let $\frac{\partial \tilde{x}^{ij}}{\partial p^{lm}}$ be the Hicksian (utility compensated) demand derivative.

Proposition: For $i \neq l$, $\frac{\partial \tilde{x}^{ij}}{\partial p^{lm}} = d \frac{\partial x^{ij}}{\partial I}$ and $\frac{\partial \tilde{x}^{ik}}{\partial p^{lm}} = d \frac{\partial x^{ik}}{\partial I}$, where d is a common factor of proportionality.

Proof: Let E^i be the amount of money spent on consumption of goods in the i 'th vector. Let $v^i(p^i, E^i) = \max_{x^i} \{g^i(x^i) | p^i \cdot x^i = E^i\}$ be the "sub-indirect utility" corresponding to consumption of the goods in the i 'th vector. Now the dual of the consumer optimization can be stated as

$$\begin{aligned} & \min_{\{x^n\}_{n \neq i}, E^i} \sum_{n \neq i} p^n \cdot x^n + E^i \\ & \text{subject to} \\ & \sum_{n \neq i} g^n(x^n) + v^i(p^i, E^i) = \bar{u} \end{aligned}$$

Now it follows immediately that $\frac{\partial \tilde{x}^{ij}}{\partial p^{lm}} = \frac{\partial x^{ij}}{\partial E^i} \frac{\partial \tilde{E}^i}{\partial p^{lm}}$ and $\frac{\partial \tilde{x}^{ik}}{\partial p^{lm}} = \frac{\partial x^{ik}}{\partial E^i} \frac{\partial \tilde{E}^i}{\partial p^{lm}}$ where $\frac{\partial \tilde{E}^i}{\partial p^{lm}}$ is the compensated expenditure derivative from the dual problem and $\frac{\partial x^{ij}}{\partial E^i}$ is the conditional demand function from the "sub-utility" maximization. Now if we similarly consider the consumer's primal problem using two state budgeting, we know that we can define income derivative of the expenditure in the goods in the i 'th vector as $\frac{\partial E^i}{\partial I}$. Using the basic rules of calculus for composite functions yields $\frac{\partial x^{ij}}{\partial I} = \frac{\partial E^i}{\partial I} \frac{\partial x^{ij}}{\partial E^i}$ or $\frac{\partial x^{ij}}{\partial E^i} = \frac{\partial x^{ij}}{\partial I} / \frac{\partial E^i}{\partial I}$. Now $\frac{\partial \tilde{x}^{ij}}{\partial p^{lm}}$ and $\frac{\partial \tilde{x}^{ik}}{\partial p^{lm}}$ can be written as $\frac{\partial \tilde{x}^{ij}}{\partial p^{lm}} = \frac{\partial x^{ij}}{\partial I} \left(\frac{\partial \tilde{E}^i}{\partial p^{lm}} / \frac{\partial E^i}{\partial I} \right) = d \frac{\partial x^{ij}}{\partial I}$ and

²²The inclusion of this proof is for completeness. It should be noted that the results presented here are well-known in the demand estimation literature (e.g. Deaton 1986).

$$\frac{\partial \tilde{x}^{ik}}{\partial p^{lm}} = \frac{\partial x^{ik}}{\partial I} \left(\frac{\partial \tilde{E}^i}{\partial p^{lm}} / \frac{\partial E^i}{\partial I} \right) = d \frac{\partial x^{ik}}{\partial I}, \text{ where } d = \left(\frac{\partial \tilde{E}^i}{\partial p^{lm}} / \frac{\partial E^i}{\partial I} \right). \text{ QED.}$$

An immediate corollary of the theorem is that the same type of relationship holds also between the derivatives of the regular demand functions.