A Theory of Optimal Capital Taxation*

Thomas Piketty, Paris School of Economics

Emmanuel Saez, UC Berkeley and NBER

October 23^{rd} 2011

Abstract

This paper develops a realistic, tractable normative theory of socially-optimal capital taxation. We present a dynamic model of savings and bequests with heterogeneous tastes for bequests to children and for wealth accumulation per se. We derive formulas for optimal inheritance tax rates expressed in terms of estimable parameters and social preferences. The long-run optimal inheritance tax rate increases with the aggregate steady-state flow of inheritances to output, decreases with the elasticity of bequests to the net-of-tax rate, and decreases with the strength of preferences for leaving bequests. For realistic parameters, the optimal inheritance tax rate should be as high as 50%-60% -or even higher for top bequests- if the government has meritocratic preferences (i.e., puts higher welfare weights on those receiving little inheritance). In contrast to the Atkinson-Stiglitz result, bequest taxation remains desirable in our model even with optimal labor taxation because inequality is two-dimensional: with inheritances, labor income is no longer the unique determinant of lifetime resources. In contrast to Chamley-Judd, optimal inheritance taxation is desirable because our preferences allow for finite long run elasticities of inheritance to tax rates. Finally, we discuss how capital market imperfections and uninsurable idiosyncratic shocks to rates of return can justify shifting one-off inheritance taxation toward lifetime capital taxation, and can account for the actual structure and mix of inheritance and capital taxation.

*Thomas Piketty, Paris School of Economics, piketty@pse.ens.fr; Emmanuel Saez, University of California at Berkeley, Department of Economics, saez@econ.berkeley.edu. We thank Tony Atkinson and Peter Diamond for helpful discussions and comments. We acknowledge financial support from the Center for Equitable Growth at UC Berkeley.

1 Introduction

According to the profession's most popular theoretical models, socially optimal tax rates on capital should be equal to $\tau_K = 0\%$ in the long run–including from the viewpoint of those individuals or dynasties who own no capital at all. This is a very strong result indeed, especially since it is supposed to apply to all forms of capital taxation, either on the capital stock or on the capital income flow, and either at the time of wealth transmission or during one's lifetime. That is, if we were to take this result seriously, then we should be pushing for a 0% inheritance tax rate, a 0% property tax rate, a 0% corporate profit tax rate, a 0% interest and dividend income tax rate, etc., and transfer the resulting revenue loss onto increased tax rates on labor income and/or consumption (or lump sum taxes). Few economists however seem to endorse such a radical policy agenda. Presumably this reflects a lack of faith in the profession's most popular models. This is in our view one of the most important failures of modern public economics.

The objective of this paper is to develop a realistic, tractable, and robust normative theory of socially optimal capital taxation. By realistic, we mean a theory providing normative conclusions that are not fully off-the-mark with respect to the real world (i.e., positive and significant capital tax rates). By realistic, we also mean a theory offering such conclusions for reasons that are consistent with the reasons which - we feel - are at play in the real world. By tractable, we mean that optimal tax formulas should be expressed in terms of estimable parameters and should quantify the various trade-offs in a simple and plausible way. By robust, we mean that our results should not be too sensitive to the exact primitives of the model nor depend on strong homogeneity assumptions for individual preferences. Ideally, formulas should be expressed in terms of estimable "sufficient statistics" such as behavioral elasticities and hence be robust to changes in the underlying primitives of the model. Such an approach has yielded fruitful results in the analysis of optimal labor income taxation (see Piketty and Saez, 2011 for a recent survey).

In our view, the two key ingredients for a proper theory of capital taxation are, first, the importance of inheritance, and, next, the imperfection of capital markets. With zero inheritance (i.e. if life-cycle savings account for the bulk of wealth accumulation), and with perfect capital markets (i.e. if agents can costlessly and risklessly transfer resources across periods at a fixed interest rate r), then we believe that the case for zero capital taxation would indeed be very strong–as in the standard Atkinson-Stiglitz framework. Therefore, our paper proceeds in two steps.

First, we develop a theory of optimal inheritance taxation. We present a dynamic model of savings and bequests with heterogeneous random tastes for bequests to children and for wealth accumulation per se. The key feature of our model is that inequality permanently arises from two dimensions: differences in labor income due to differences in ability, and differences in inheritances due to differences in parental wealth accumulation and transmission. Importantly, top labor earners and top successors are never exactly the same people, implying a non-degenerate trade-off between labor taxation and inheritance taxation. In that context, in contrast to the famous Atkinson-Stiglitz result, bequest taxation remains desirable even with optimal labor taxation because, with inheritances, labor income is no longer the unique determinant of life-time resources. In sum: two-dimensional inequality requires two-dimensional tax policy tools.

We derive formulas for optimal bequest tax rates expressed in terms of estimable parameters and social preferences. The long run optimal bequest tax rate τ_B increases with the aggregate steady-state flow of bequests to output b_y, decreases with the elasticity of bequests to the net-of-tax rate e_B, and decreases with the strength of preferences for bequests. For realistic parameters, the optimal linear inheritance tax rate should be as high as $50\% - 60\%$ if the government has meritocratic preferences (i.e., puts higher welfare weights on those with little inheritance). Because real world inherited wealth is highly concentrated (basically half of the population receives close to zero bequest), our results are very robust to reasonable changes in the social welfare objective. I.e. the optimal tax policy from the viewpoint of those receiving zero bequest is very close to the welfare optimum for bottom 50% bequest receivers.

For top bequests, the optimal inheritance tax rate τ_B can be even larger (say, $70\% - 80\%$), especially if bequest flows are large, and if the probability of bottom receivers to leave a large bequest is small. Therefore our normative model can account for the relatively large bequest tax rates observed in most advanced economies during the past 100 years, especially in Anglo-Saxon countries between the 1930s and the 1980s (see Figure 1). To our knowledge this is the first time that a model of optimal inheritance taxation delivers simple and tractable formulas that can be used to analyze this important part of real world tax policies.

Our model also illustrates the critical importance of perceptions and beliefs systems about wealth inequality and mobility (i.e. individual most preferred tax rates are very sensitive to expectations about bequests received and left), and about the aggregate magnitude of the bequest flow b_y. When b_y is small, say around 5% of national income, as was the case in

Continental Europe during the 1950s-1970s, then bequest taxes should be moderate (new savings should be encouraged in reconstruction periods). But when b_y is high and rising, as has been the case since the 1980s-1990s (b_y is currently about 15% of national income in a country like France, and 20%-25% of disposable income, like in the 19^{th} century), then bequest taxes should be large - so as to reduce the tax burden falling on labor earners.[1]

Second, we show that if we introduce capital market imperfections and uninsurable idiosyncratic shocks to rates of return into our setting, then we can turn our positive optimal inheritance tax results into more general optimal capital tax results.

The basic intuition is the following. From a welfare viewpoint, as well as from an optimal tax viewpoint, what matters is the capitalized bequest $\tilde{b}_{ti} = b_{ti}e^{r_{ti}H}$, not the raw value of bequest b_{ti}. received by a given individual i (with H = generation length, typically 30 years, r_{ti} = rate of return)). But at the time of setting the bequest tax rate τ_B, nobody has any idea about the future rate of return on a given asset is going to during the following 30 years. I.e. nobody knows what $e^{r_{ti}H}$ is going to be. Rates of return are notoriously difficult to predict, and they vary enormously over assets and across individuals. E.g. someone who inherited an apartment in Paris in the 1970s had no idea what the total returns and capitalized value of this asset would be three decades later. Therefore for simple insurance reasons it makes more sense to split the tax burden between one-off bequest taxes τ_B, and lifetime capital taxes τ_K (annual property taxes and/or taxes on flow returns).

We show that if the uninsurable uncertainty about future returns is large, and if the effort elasticity of rates of return is moderate, then the resulting optimal lifetime capital tax rate τ_K can be very high - typically higher than the optimal bequest tax rate τ_B, and labor tax rate τ_L.

This can account for the fact that in modern tax systems the bulk of aggregate capital tax revenues comes from lifetime capital taxes (rather than from inheritance taxes). It is also interesting to note that the countries which experienced the highest top inheritance tax rates also applied the largest tax rates on top incomes, and particularly so on tax capital incomes (see Figures 2-3). This suggests that the policy makers who implemented these policies viewed taxes on large inheritances and large capital incomes as complementary. To our knowledge this is the first time that a model of optimal capital taxation is able to jointly account for this important

[1] The historical evolution and theoretical determinants of the aggregate bequest flow b_y were recently studied by Piketty (2011). Here we build upon this work and extend Piketty's r-vs-g model in order to study normative issues and optimal tax policy.

set of facts.

Of course we do not pretend that these highly progressive policies implemented in the U.S. and the U.K. between the 1930s and 1980s were necessarily optimal. The way we calibrate our optimal tax formulas in the present paper should be viewed as illustrative and exploratory. But at least we offer a theoretical framework which for some parameter values can rationalize such policies - and for some alternative parameter values could also rationalize different policies. Ideally our approach should contribute to a tax debate based more upon empirical estimates, and less upon abstract theoretical results and ideology.

The rest of the paper is organized as follows. Section 2 relates our results to the existing literature. Section 3 presents our basic results on optimal inheritance taxation. Section 4 introduces capital market imperfections and analyzes the consequences for the optimal mix between one-off inheritance taxation and lifetime capital taxation. Section 5 extends our results in a number of directions, particularly elastic labor supply, closed economy, life-cycle saving, population growth, and intergenerational redistribution. Section 6 offers some concluding comments. Most proofs are gathered in the appendix.

2 Relation to existing literature

There are two main results in the literature in support of zero capital income taxation: the Atkinson and Stiglitz (1976) theorem, and the Chamley (1986) and Judd (1985) results. We discuss each in turn and then discuss the more recent literature introducing capital market imperfections in the analysis.

Atkinson-Stiglitz. Atkinson-Stiglitz show that there is no need to supplement the optimal non-linear labor income tax with a capital income tax in a lifecycle model if leisure choice is (weakly) separable from consumption choices and preferences for consumption are homogeneous. In that model, the only source of lifetime income inequality is labor skill and hence there is no reason to redistribute from high savers to low savers (i.e. tax capital income) conditional on labor earnings.[2] This key assumption of the Atkinson-Stiglitz model breaks down in a model with inheritances where inequality in lifetime income comes from both differences in labor income

[2] Saez (2002) shows that this result extends to heterogeneous preferences as well as long as time preferences are orthogonal to labor skills. If time preferences are correlated with labor skills, then taxing saving is a desirable and indirect way to tax ability.

and differences in inheritances received. In that context and conditional on labor earnings, a high level of bequests left is a signal of a high level of inheritances received, which provides a rationale for taxing bequests.

The simplest way to see this point is to consider a model with inelastic and uniform labor income but with differences in inheritances due to parental differences in preferences for bequests. In such a world, labor income taxation is useless for redistribution and taxing inheritances is desirable for redistribution (as long as inheritances are not infinitely elastic to taxation).

This important point has been made by Cremer, Pestieau, and Rochet (2003) in a stylized model with unobservable inherited wealth where they show that taxing capital income becomes desirable. Our model allows the government to directly observe (and hence tax) wealth. Farhi and Werning (2010) also propose a related analysis but consider a model from the perspective of the first generation of donors who do not start with any inheritance (so in effect there is no inequality at all along the inheritance dimension). In this context, bequests should actually be subsidized as they should be untaxed by Aktinson-Stiglitz (ignoring inheritors) and hence should be subsidized when taking into account inheritors. As we shall see, this result is not robust to models where people both receive and leave bequests.

Chamley-Judd. Chamley-Judd show that the optimal capital income tax should be zero in the long-run. This zero long-run result holds for two reasons.

First, and as originally emphasized by Judd (1985), the zero rate results happens because social welfare is measured exclusively from the initial period (or dynasty). In that context, a constant tax rate on capital income creates a tax distortion growing exponentially overtime which cannot be optimal (see Judd 1999 for a clear intuitive explanation). Such a welfare criteria can only make sense in a context with homogeneous discount rates. In the context of inheritance taxation where each period is a generation and where preferences for bequests are very heterogeneous across the population, this strikes us as a particularly bad definition of social welfare. We will adopt instead a definition of social welfare based on long-run equilibrium steady-state utility.[3] We show in appendix how the within generation and across generation redistribution problems can be disconnected using public debt so that there is effectively no loss

[3]In models with uncertainty, using the initial period social welfare criteria leads to optimal policies where inequality grows without bounds (see e.g. Atkeson and Lucas 1992). Obtaining "immiseration" as an optimal redistributive tax policy is obviously absurd and should be interpreted as a failure of the initial period social welfare criteria. Farhi and Werning (2007) show that considering such long-run steady-state equilibrium as we do in this paper eliminates the immiseration results.

of generality in focusing on steady state welfare.

Second, even adopting a long-run steady-state utility perspective, the optimal capital income tax rate is still zero in the standard Chamley-Judd model. This is because the supply side elasticity of capital with respect to the net-of-tax return is infinite in the infinite horizon model with constant discount rate.[4] Our theory leaves this key elasticity as a free parameter to be estimated empirically. Our model naturally nests the Chamley-Judd case when the elasticity is infinite. We think this is a much better approach as there is no compelling empirical evidence for an infinite elasticity.

We should stress that all these authors were aware of these limitations. The basic ideas here are not new. What is new is that we put them together in a simple model and re-arrange the various effects so as to obtain robust, tractable theoretical formulas for optimal tax rates.

Capital Market Imperfections. A number of papers have shown that taxing capital income can become desirable when capital market imperfections are introduced, even in models with no inheritance. Typically, it is good to tax capital as a way to redistribute from those with no credit constraints (the owners of capital) toward those with credit constraints (non-owners of capital). Aiyagari (1995) and Chamley (2001) make this point formally in a model with borrowing constrained infinitely lived- agents facing labor income risk. They show that capital income taxation is desirable when consumption is positively correlated with savings (but do not attempt to compute numerical values for optimal capital tax rates).[5] There are numerical computations such as Conesa, Kitao and Krueger (2009) or Cagetti and DeNardi (2009). Conesa et al. calibrate an optimal tax, OLG model with uninsurable idiosyncratic labor productivity shocks and borrowing constraints, and find $\tau_K = 36\%$ and $\tau_L = 23\%$ in their preferred specification. The main effect seems to be that capital income tax is an indirect way to tax more the old and to tax less the young, so as to alleviate their borrowing constraints. This is an interesting and potentially important effect. But we do not believe that this is the most important explanation for $\tau_K > 0$. If the old vs. young issue was the main issue, then we would probably find other, more direct ways to address it (e.g. age-varying income taxes; some policies, e.g. pension schemes, do depend on age).

[4]This follows from the fact that the net-of-tax rate of return needs to be equal to the Golden rule level in the long run (see appendix B).

[5]This correlation is always positive in the Aiyagari (1995) model with independent and identically distributed labor income, but Chamley (2001) shows that the correlation can be negative theoretically.

Models based upon government time-inconsistency. Yet another way to explain real-world, positive capital taxes is to assume time inconsistency. Zero capital tax results are long run results, but never hold in the short run. I.e. it is always tempting for short-sighted governments to have $\tau_K > 0$ in the short run, even though optimal long run τ_K is equal to 0%. We do not believe however that this is an important part of the explanation as to why we observe positive capital taxes in the real world. E.g. we feel that even long-sighted governments view positive and substantial inheritance tax rates (say, of the kind we have been observing over the past 100 years, see Figure 1 above) as part of a fair and efficient permanent tax system. So in this paper we assume away time inconsistency issues, and we always look at long run optimal tax policies - assuming full commitment.

3 A Theory of Optimal Inheritance Taxation

3.1 Notations and definitions

We consider a small open economy facing an exogenous, instantaneous rate of return on capital $r \geq 0$. To keep notations minimal, we focus upon a simple model with a discrete set of generations $0, 1, .., t, ..$ Each generation has measure one, lives one period (which can be interpreted as H-year-long, where H = generation length, realistically $H \simeq 30$), then dies and is replaced by the next generation. Total population is stationary and equal to $N_t = 1$.

Generation t receives average inheritance (pre-tax) b_t from generation $t-1$ at the beginning of period t. Inheritances go into the capital stock and are invested either domestically or abroad for a "generational" rate of return $R = e^{rH} - 1$. Production in generation t combines labor from generation t and capital to produce a single output good. The output produced by generation t is either consumed by generation t or left as bequest to generation $t+1$. We denote by y_{Lt} the average labor income received by generation t. We denote by c_t the average consumption of generation t and b_{t+1} the average bequest left by generation t to generation $t+1$. We assume that output, labor income and capital income are realized at the end of period. Consumption c_t and bequest left b_{t+1} also take place at the end of the period.[6]

[6] We focus upon this one-period, discrete-time model for notational simplicity only. All results and optimal inheritance tax formulas can be easily extended to a full-fledged, multi-period, continuous-time model with overlapping generations and life-cycle savings. See section 5 below.

Individual i in generation t maximizes utility:

$$\max V_{ti} = V_i(c_{ti}, w_{ti}, \bar{b}_{t+1i}) \quad \text{s.t.} \quad c_{ti} + w_{ti} \leq \widetilde{y}_{ti} = (1-\tau_B)b_{ti}e^{rH} + (1-\tau_L)y_{Lti}$$

With: $\widetilde{y}_{ti} = (1-\tau_B)b_{ti}e^{rH} + (1-\tau_L)y_{Lti}$ = total after-tax lifetime income (combining labor income y_{Lti}, inheritance received b_{ti} and the returns on inheritance received Rb_{ti})

c_{ti} = consumption

w_{ti} = end-of-life wealth = b_{t+1i} = pre-tax raw bequest left to next generation

$\bar{b}_{t+1i} = (1-\tau_B)b_{t+1i}e^{rH}$ = after-tax capitalized bequest left to next generation

τ_B = bequest tax rate, τ_L = labor income tax rate

We derive optimal tax formulas holding for large classes of utility functions V_i, using a sufficient-statistics approach. In order to fix ideas, we focus upon the Cobb-Douglas case:

$$V_i(c, w, \bar{b}) = c^{1-s_i} w^{s_{wi}} \bar{b}^{s_{bi}} \quad (s_{wi} \geq 0, s_{bi} \geq 0, s_i = s_{wi} + s_{bi} \leq 1)$$

This simple form implies that individual i devotes a fraction s_i of his lifetime resources to end-of-life wealth, and a fraction $1 - s_i$ to consumption. Our key results - and in particular our optimal tax formulas - also hold with CES utility functions, and actually with all utility functions $V_i(c, w, \bar{b})$ that are homogenous of degree one (so as to allow for balanced growth paths).[7] In effect, these more general forms imply that the fraction s_i can depend on relative prices (i.e. the inheritance tax rate and the rate of return), rather than being fixed as in the Cobb-Douglas (or log-log) case, where income and substitution effects exactly offset each other.

We use a standard wealth accumulation model with exogenous growth. Output in generation t is given by a constant return to scale production function $Y_t = F(K_t, L_t)$, where K_t is the physical (non-human) capital input and L_t is the human capital input (efficient labor supply). Though this is unnecessary for our results, in order to simplify the notations we assume a Cobb-Douglas production function: $Y_t = K_t^\alpha L_t^{1-\alpha}$.

Aggregate human capital is the sum over all individuals of raw labor supply l_{ti} times labor productivity h_{ti} : $L_t = \int_{i \in N_t} l_{ti} h_{ti} di$. Average productivity h_t is assumed to grow at some exogenous rate $1 + G = e^{gH}$ per generation (with $g \geq 0$): $h_t = h_0 e^{gHt}$. With inelastic labor supply ($l_{ti} = 1$), we simply have: $L_t = N_t h_t = h_0 e^{gHt}$.

Taking as given the "generational" rate of return $R = e^{rH} - 1$, profit maximization implies that the domestic capital input K_t is chosen so that $F_K = R$, i.e. $K_t = \beta^{\frac{1}{1-\alpha}} L_t$ (with $\beta = \dfrac{K_t}{Y_t} =$

[7] See in particular the proof of proposition 2 below.

$\frac{\alpha}{R}$ = domestic generational capital-output ratio).[8] It is important to keep in mind that Y_t is domestic output. In an open economy, Y_t might differ from national income if the domestic capital stock (used for domestic production) differs from the national wealth.

It follows that output $Y_t = \beta^{\frac{\alpha}{1-\alpha}} L_t = \beta^{\frac{\alpha}{1-\alpha}} h_0(1+G)^t$ also grows at rate $1+G = e^{gH}$ per generation. So does aggregate labor income $Y_{Lt} = (1-\alpha)Y_t$. And so do per capita output, capital and labor income y_t, k_t, y_{Lt} ($= Y_t, K_t, Y_{Lt}$ divided by $N_t = 1$). The aggregate economy is on a steady-state growth path where everything grows at rate $1+G = e^{gH}$ per generation.

With $g = 1-2\%$ per year and $H = 30$ years, $1+G = e^{gH} \simeq 1.5-2$. With $r = 3\%-5\%$ per year and $H = 30$ years, $1+R = e^{rH} \simeq 3-4$.

3.2 Steady-state inheritance flows and distributions

The individual-level transition equation for bequest is the following:

$$b_{t+1i} = s_i[(1-\tau_L)y_{Lti} + (1-\tau_B)b_{ti}e^{rH}] \tag{1}$$

In our model–and we believe in the real world–there are three independent factors explaining why different individuals receive different bequests b_{t+1i} within generation $t+1$: their parents received different bequests b_{ti}, and/or earned different labor income y_{Lti}, and/or had different tastes for savings $s_i = s_{wi} + s_{bi}$.[9]

In particular, an important point in our set-up is that taste parameters vary across individuals. E.g. some individuals might have zero taste for wealth and bequest ($s_{wi} = s_{bi} = 0$), in which case they save solely for life-cycle purposes and die with zero wealth ("life-cycle savers"). Others might have taste for wealth but not for bequest ($s_{wi} > 0, s_{bi} = 0$) ("wealth-lovers"), while others might have no direct taste for wealth but taste for bequest ($s_{wi} = 0, s_{bi} > 0$) ("bequest-lovers"). The taste for wealth could reflect direct utility for the prestige or social status conferred by wealth. In presence of uninsurable productivity shocks, it could also measure the security brought by wealth, i.e. its insurance value (so this modeling can be viewed as a reduced form for precautionary saving). The only difference between wealth- and bequest-lovers is that the former do not care about bequest taxes while the latter do.

[8] The annual capital-output ratio is $\beta_{annual} = H \cdot \beta = \alpha(H/R) = \alpha H/(e^{rH}-1) \simeq \alpha/r$ if r is small.

[9] A fourth important factor in the real world is the existence of idiosyncratic shocks to rates of return r_i (which we introduce explicitly in section 4 below). Pure demographic shocks (such as shocks to the age at death of parents and children, number of children, etc.) also play an important role.

In the real world, most individuals are at the same time life-cycle savers, wealth-lovers and bequest-lovers. But the exact magnitude of these various saving motives does vary a lot across individuals, just like other tastes.[10] We allow for any distribution and random process for taste parameters (possibly with some intergenerational persistence). In order to ensure the existence of a unique ergodic steady-state distribution of wealth, we simply require the random process to be drawn from a full support distribution: whatever your parental taste, you always have a positive probability to end up with any taste.[11] This also implies that in each generation there is a positive density of "zero receivers" (i.e. individuals who receive zero bequest, because their parents had zero taste for wealth and bequest).[12]

Assumption 1 *Taste parameters* (s_{wi}, s_{bi}) *are drawn according to a full-support random process: For all* $s_{wti}, s_{bti} \geq 0, s_{wt+1i}, s_{bt+1i} \geq 0, g(s_{wt+1i}, s_{bt+1i}|s_{wti}, s_{bti}) > 0$

(where $s_{wti}, s_{bti} =$ *parental tastes,* $s_{wt+1i}, s_{bt+1i} =$ *children tastes,* $g(.) =$ *density function).*

We denote by $g(s_{wi}, s_{bi})$ *the stationary cross-sectional distribution, and by* $s = E(s_i)$*.*

A special case is no taste memory: for all $s_{wti}, s_{bti}, g(s_{wt+1i}, s_{bt+1i}|s_{wti}, s_{bti}) = g(s_{wt+1i}, s_{bt+1i})$

In the "no taste memory" case (tastes are drawn i.i.d. for each cohort), then by linearity the individual transition equation (1) can be easily be aggregated into:

$$b_{t+1} = s[(1 - \tau_L)y_{Lt} + (1 - \tau_B)b_t e^{rH}] \tag{2}$$

Let us denote the aggregate *capitalized* bequest flow-domestic output ratio by $b_{yt} = \dfrac{e^{rH}B_t}{Y_t} = \dfrac{e^{rH}b_t}{y_t}$. Dividing both sides of equation (2) by per capita domestic output y_t and noting that $b_{t+1}/y_t = b_{yt+1}e^{-(r-g)H}$, we obtain the following transition equation for b_{yt}:

$$b_{yt+1} = e^{(r-g)H}[s(1 - \tau_L)(1 - \alpha) + s(1 - \tau_B)b_{yt}] \tag{3}$$

In order to ensure convergence towards a non-explosive steady-state, we must assume that the average taste for wealth and bequest is not too strong:

[10] For some evidence on the unequal distribution of bequest motives, see Kopczuk and Lupton (2007).

[11] The random process for tastes could also depend on parental productivity shocks or bequest received, providing that we make the appropriate full support assumption (so as to ensure ergodicity).

[12] Another way to obtain the same outcome is to assume that the idiosyncratic shocks to rates of return r_{ti} are such that in each generation there is a positive density of bankrupt individuals (or to assume demographic or health shocks with similar effects). The important point is that in the real world there is always a large set of zero (or near zero) receivers, for a variety of reasons.

Assumption 2 $s(1-\tau_B)e^{(r-g)H} < 1$

In case assumption 2 is violated, then the economy will accumulate infinite wealth as compared to domestic output. At some point it will not be a small open economy any more, and the world rate of return will have to fall in order to restore assumption 2.[13]

In case assumption 2 is satisfied, then $b_{yt} \to b_y = \dfrac{s(1-\tau_L)(1-\alpha)e^{(r-g)H}}{1-s(1-\tau_B)e^{(r-g)H}}$ as $t \to +\infty$. I.e. the aggregate inheritance-output ratio converges, and in steady-state all bequests grow at the same rate at output.

We also need to specify the structure of labor productivity shocks. Individual i in generation t is characterized by a within-cohort normalized productivity parameter θ_{ti}. By definition, we have: $y_{Lti} = \theta_{ti} y_{Lt}$ (with $E(\theta_i) = 1$). Productivity differentials θ_{ti} could come from innate abilities, acquired skills, or sheer luck - and most likely from a complex combination between the three. We assume that productivity shocks also follow a full support, exogenous random process.

Assumption 3 *Productivity parameters θ_i are drawn according to a full-support random process over some interval $[\theta_0, \theta_1]$: for all $\theta_{ti}, \theta_{t+1i}, h(\theta_{t+1i}|\theta_{ti}) > 0$*

(where θ_{ti} = parental productivity, θ_{t+1i} = children productivity, $h(.)$ = density function).

We denote by $h(\theta_i)$ the stationary cross-sectional distribution, with $E(\theta_i) = 1$.

A special case is no labor productivity memory: for all $\theta_{ti}, \theta_{t+1i}, h(\theta_{t+1i}|\theta_{ti}) = h(\theta_{t+1i})$.

Another special case is no labor productivity inequality: $\theta_0 = \theta_1 = 1$.

Finally, we note z_{ti} the within-cohort normalized bequest (i.e. we write down received bequest as the product of normalized times average bequest: $b_{ti} = z_{ti} b_t$), and $\phi_t(z)$ the distribution of normalized bequest within cohort t . Given some initial distribution $\phi_0(z)$, the random processes for tastes and productivity $g(.)$ and $h(.)$ and the individual transition equation (2) entirely determine the low of motion for the distribution of inheritance $\phi_t(z)$ and the joint distribution of inheritance and labor productivity, which we note $\psi_t(z, \theta)$.

Proposition 1 *(a) Under assumptions 1-3, there exists a unique steady-state for the aggregate inheritance flow-output ratio b_y , the inheritance distribution $\phi(z)$ and the joint inheritance-productivity distribution $\psi(z, \theta)$. Whatever the initial conditions, as $t \to \infty$, $b_{yt} \to b_y$, $\phi_t \to \phi$ and $\psi_t \to \psi$.*

[13] Assumption 2 needs to be generalized in order to apply outside the no-taste-memory case. See appendix proof of Proposition 1.

(b) Under "no taste memory" special case, $b_y = \dfrac{s(1-\tau_L)(1-\alpha)e^{(r-g)H}}{1-s(1-\tau_B)e^{(r-g)H}}$

Otherwise, b_y *is also unique but depends on* $g(.)$ *and* $h(.)$ *(no closed form formula).*

(c) Under "no labor inequality" special case, $\psi(z,\theta) = \phi(z)$ *is one-dimensional. Otherwise,* $\psi(z,\theta)$ *is two-dimensional and has full support.*

(d) In all cases, there are "zero bequest receivers" in steady-state: $\phi(0) > 0$.

Proof . The steady-state uniqueness result follows from standard ergodic convergence theorems (see appendix proof for details). **QED**

Two points are worth noting here. First, the aggregate magnitude of inheritance flows relatively to output b_y is a positive function of $r-g$. In societies with high returns and low growth, wealth coming from the past is being capitalized at a faster rate than national income. Successors simply need to save a small fraction of their asset returns to ensure that their inherited wealth grows at least as fast as output. Conversely, with low returns and high growth, inheritance is dominated by new wealth, and the steady-state aggregate inheritance flow is a small fraction of output.

This simple r-vs-g model is able to reproduce remarkably well the observed evolution of aggregate inheritance flows over the past two centuries. In particular, it can explain why inheritance flows were so large in the 19^{th} and early 20^{th} centuries (20%-25% of national income in 1820-1910), and why they are becoming large again in the late 20^{th} and early 21^{st} centuries (about 15% in 2010 in France, up from less than 5% around 1950-1960) (see Figures 4-5). Typically, with $r = 4\% - 5\%$ and $g = 1\% - 2\%$, simple calibrations of the above formula show that the annual inheritance flow b_y can indeed be as large as $20\% - 25\%$ of national income.[14] Available evidence suggests that the French pattern also applies to Continental European countries that were hit by similar growth and capital shocks. For countries like the United States and the United Kingdom, the long-run U-shaped pattern of aggregate inheritance flows was possibly somewhat less pronounced.

Next, one key feature of our model is that inequality is two-dimensional: in steady-state, within each cohort, there will always be some individuals with low inheritance z and high labor productivity θ, and conversely. As we shall see, this explains why the Atkinson-Stiglitz result

[14] E.g. with $r-g = 3\%$, $H = 30$, $\alpha = 30\%$, $s = 10\%$, $\tau_B = \tau_L = 0\%$, then $b_y = 23\%$. With $r-g = 2\%$, then $b_y = 16\%$. With $r-g = 3\%$ and $\tau = 30\%$, then $b_y = 13\%$, but $b_y/(1-\tau) = 19\%$. For detailed simulations using a full-fledged, out-of-steady-state version of this model, with life-cycle savings and full demographic and macroeconomic shocks, see Piketty (2011).

does not hold, and why we need a two-dimensional tax policy tool (τ_B, τ_L) (a redistributive labor income tax is not enough).

3.3 Basic optimal tax formula

We now define our optimal tax problem. We assume that the government faces an exogenous revenue requirement: public good spending must satisfy $G_t = \tau Y_t$ where $\tau \geq 0$ and Y_t is domestic output. For the time being we assume that the government has only two tax instruments: a proportional tax on labor income at rate $\tau_L \geq 0$, and a proportional tax on inheritance at rate $\tau_B \geq 0$. We impose a period-by-period budget constraint: the government must raise from labor income Y_{Lt} and inheritance B_t received by generation t an amount sufficient to cover government spending τY_t for generation t. We assume that government spending happens at the end of the period (in the same way as private consumption). Hence, the period t government budget constraint looks as follows:

$$\tau_L y_{Lt} + \tau_B b_t e^{rH} = \tau y_t$$

$$\text{i.e. :} \quad \tau_L (1-\alpha) + \tau_B b_y = \tau \tag{4}$$

One can interpret the bequest tax $\tau_B b_t e^{rH}$ in two equivalent ways. It could be that the tax is raised on the capitalized value of bequest $b_t e^{rH}$ at the end of the period. It could also be the tax is raised on raw bequest b_t at the beginning of the period, and then the tax revenue is invested by the government at market rate of return r until the end of the period. As long as capital markets are perfect and everybody gets the same return r, these two ways of raising the tax are fully equivalent, and the choice of tax instruments is irrelevant. In section 4 we introduce heterogenous returns and capital market imperfections, which allows us to analyze the optimal mix between inheritance taxation and lifetime capital taxation.

The question that we now ask is the following: what is the tax policy (τ_L, τ_B) maximizing long-run, steady-state social welfare? That is, we assume that the government can commit for ever to a tax policy $(\tau_{Lt} = \tau_L, \tau_{Bt} = \tau_B)_{t \geq 0}$ and cares only about the long-run steady-state distribution of welfare V_{ti}. Under assumptions 1-3, for any tax policy there exists a unique steady-state ratio b_y and distribution $\psi(z, \theta)$. The government chooses (τ_L, τ_B) so as to maximize the following, steady-state social welfare function:[15]

$$SWF = \iint_{z \geq 0, \theta_0 \leq \theta \leq \theta_1} \omega_{z\theta} \frac{V_{z\theta}^{1-\Gamma}}{1-\Gamma} d\Psi(z,\theta) \tag{5}$$

With: $V_{z\theta}$ = average utility level V_i attained by individuals i with normalized inheritance $z_i = z$ and productivity $\theta_i = \theta$

$\omega_{z\theta}$ = social welfare weights as a function of normalized inheritance z and productivity θ

Γ = concavity of the social welfare function ($\Gamma \geq 0$)[16]

A key parameter to answer this question is the long-run elasticity e_B of aggregate inheritance ratio b_y with respect to the net-of-bequest-tax rate $1 - \tau_B$ (letting τ_L adjust to keep budget balance, see equation (4)):

$$e_B = \frac{db_y}{d(1-\tau_B)} \frac{1-\tau_B}{b_y} \tag{6}$$

In general, one might expect $e_B > 0$: with a higher net-of-tax rate $1 - \tau_B$ (a lower tax rate τ_B), agents may choose to devote a somewhat larger fraction of their resources to inheritance, in which case the aggregate, steady-state inheritance ratio will be somewhat bigger. But this could also go the other way, because e_B is defined along a budget balanced steady-state frontier: lower bequest taxes imply higher labor taxes, which in turn make it more difficult for high labor earners to accumulate large bequests.

E.g. with Cobb-Douglas preferences, and in the absence of taste memory, then by substituting $\tau_L(1-\alpha) = \tau - \tau_B b_y$ (equation (4)) into the steady-state formula for b_y (Proposition 1), we obtain:

$$b_y = \frac{s(1-\tau-\alpha)e^{(r-g)H}}{1-se^{(r-g)H}} \tag{7}$$

It follows that $e_B = 0$ in the simplified, one-generation equals one-period Cobb-Douglas model. In the full-fledged overlapping generations and continuous-time model, one can show that if inheritance tends to occur around mid-life, which is the case in practice, then e_B is close to zero.[17] Of course, the Cobb-Douglas form and the no-taste-memory assumption are restrictive, and for general utility functions and random processes for tastes, then e_B could really take any value. We view e_B as a free parameter to be estimated empirically. The important point here is that there is no reason to expect e_B to be infinitely large, unlike in infinite-horizon dynastic models.

[15] This steady-state maximization problem can also be formulated as the limit solution of a dynamic social welfare maximization problem (with a social rate of time preference going to zero). See Appendix B.

[16] If $\Gamma = 1$, then $SWF = \iint_{z \geq 0, \underline{\theta} \leq \theta \leq \bar{\theta}} \omega_{z\theta} \log(V_{z\theta}) d\Psi(z,\theta)$.

[17] See section 5 below.

Throughout this paper we are particularly interested in the zero-bequest-receivers social optimum, i.e. the optimal tax policy from the viewpoint of those who receive zero bequest, and who must rely entirely on their labor income to make their way in life. This corresponds to the case with a linear social welfare function ($\Gamma = 0$) and the following welfare weights: $\omega_{z\theta} = 1$ if $z = 0$, and $\omega_{z\theta} = 0$ if $z > 0$. Since private preferences $V_i()$ are homogenous of degree one, $\Gamma = 0$ implies that the government does not want to redistribute income from high productivity to low productivity individuals - maybe because individuals are viewed as responsible for their productivity parameter θ. In contrast, individuals clearly bear no responsibility at all for their bequest parameter z. Therefore it seems very appealing from a normative viewpoint to try to reduce as much as possible the inequality of lifetime welfare opportunities along the inheritance dimension.[18] Indeed, in the political debate about estate taxation, the left insists on taxing idle wealthy "trust funders" heirs while the right insists on the plight of those who work hard to accumulate a fortune for their family and children only to see it partly confiscated by "death taxes". Hence, it seems reasonable to put more weight on bequest leavers than on bequest receivers. So we start by characterizing this zero-bequest-receivers optimum, which one might also call the "meritocratic Rawlsian optimum":

Proposition 2 ***(zero-bequest-receivers optimum).*** *Under assumptions 1-3, linear social welfare ($\Gamma = 0$), and the following welfare weights: $\omega_{z\theta} = 1$ if $z = 0$, and $\omega_{z\theta} = 0$ if $z > 0$, then*

$$\tau_B = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}} \quad \text{and} \quad \tau_L = \frac{\tau - \tau_B b_y}{1 - \alpha}$$

with $s_{b0} = E(s_{bi}|z_i = 0) =$ the average bequest taste of zero bequest receivers.

Proof . Take a given tax policy (τ_L, τ_B). Consider a small increase in the bequest tax rate $d\tau_B > 0$. Differentiating the government budget constraint, $\tau_L(1 - \alpha) + \tau_B b_y = \tau$, in steady-state $d\tau_B > 0$ allows the government to cut the labor tax rate by:

$$d\tau_L = -\frac{b_y d\tau_B}{1 - \alpha}\left(1 - \frac{e_B \tau_B}{1 - \tau_B}\right) \quad (< 0 \text{ as long as } \tau_B < \frac{1}{1 + e_B})$$

Note that $d\tau_L$ is proportional to the aggregate inheritance-output ratio b_y: with a larger aggregate inheritance flow, a given increase in the bequest tax rate can finance a larger labor tax cut.

[18] Maybe surprisingly, the recent normative literature on fairness, equal opportunity and responsibility has devoted little attention to the issue of inheritance taxation. E.g. Roemer et al. (2003) and Fleurbaey and Maniquet (2006) focus on income taxation. See however the interesting discussion in Fleurbaey (2008, pp.146-148).

An individual i who receives no inheritance ($b_{ti} = 0$) chooses b_{t+1i} to maximize

$$V_i(c_{ti}, w_{ti}, \bar{b}_{t+1i}) = V_i((1 - \tau_L)y_{Lti} - b_{t+1i}, b_{t+1i}, (1 - \tau_B)(1 + R)b_{t+1i}).$$

The first order condition in b_{t+1i} is $V_{ci} = V_{wi} + (1 - \tau_B)(1 + R)V_{\bar{b}i}$ This leads to $b_{t+1i} = s_i(1 - \tau_L)y_{Lti}$ (with $0 \leq s_i \leq 1$) We can define $\nu_i = (1 - \tau_B)(1 + R)V_{\bar{b}i}/V_{ci}$ the share of bequest left for bequest loving reasons ($1 - \nu_i$ is the share left for wealth loving reasons), and $s_{bi} = \nu_i s_i$ the strength of the bequest taste. In the Cobb-Douglas utility case, s_{bi} is simply the fixed exponent in the utility function. In the general homogeneous utility case, s_{bi} may depend on τ_B and $1 + R$.

Using the envelope theorem as b_{t+1i} maximizes utility, the utility change dV_i created by a budget balance tax reform $d\tau_B, d\tau_L$ can be written as follows:

$$dV_i = -V_{ci}y_{Lti}d\tau_L - V_{\bar{b}i}(1 + R)b_{t+1i}d\tau_B$$

$$\text{I.e.:} \qquad dV_i = V_{ci}y_{Lti}d\tau_B \left[\left(1 - \frac{e_B\tau_B}{1 - \tau_B}\right)\frac{b_y}{1 - \alpha} - \frac{1 - \tau_L}{1 - \tau_B}s_{bi} \right]$$

The first term in the square brackets is the utility gain due to the reduction in the labor income tax (as noted above, it is proportional to the aggregate inheritance ratio b_y), while the second term is the utility loss due to reduced net-of-tax bequest left (it is naturally proportional to the bequest taste s_{bi}).

By using the fact that $1 - \tau_L = (1 - \alpha - \tau + \tau_B b_y)/(1 - \alpha)$ (from the government budget constraint), this can be re-arranged into:

$$dV_i = V_{ci}y_{Lti}d\tau_B \frac{1 - \tau_L}{1 - \tau_B} \left[\frac{1 - (1 + e_B)\tau_B}{1 - \alpha - \tau + \tau_B b_y} b_y - s_{bi} \right].$$

Summing up over all zero-bequest-receivers, we get:

$$dSWF \sim d\tau_B \left[\frac{1 - (1 + e_B)\tau_B}{1 - \alpha - \tau + \tau_B b_y} b_y - s_{b0} \right], \text{with} \quad s_{b0} = \frac{\int_{z_i=0} V_{ci}y_{Lti}s_{bi}d\Psi}{\int_{z_i=0} V_{ci}y_{Lti}d\Psi}.$$

Setting $dSWF = 0$, we get the formula: $\tau_B = \dfrac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}}$. **QED.**

Note 1. The proof works with any utility function that is homogenous of degree one (and not only in the Cobb-Douglas or CES cases). In the general case, s_{b0} is the average of s_{bi} over all zero-bequest-receivers, weighted by the product of their marginal utility V_{ci} and of their labor income y_{Lti}. In case $s_{bi} \perp V_{ci}y_{Lti}$ (e.g. in case there is no taste memory, or no labor productivity inequality), and in case the utility functions $V_i()$ are Cobb-Douglas, then s_{b0} is the simple average of s_{bi} over all zero-bequest-receivers: $s_{b0} = E(s_{bi}|z_i = 0)$.

Note 2. The optimal tax formula can be extended to the case $\Gamma > 0$: one simply needs to replace s_{b0} by: $s_{b0} = \frac{\int_{z_i=0} V_{ci} y_{Lti} s_{bi} V_i^{-\Gamma} d\Psi}{\int_{z_i=0} V_{ci} y_{Lti} V_i^{-\Gamma} d\Psi}$. I.e. the formula for s_{b0} needs to be reweighted in order to take into account the lower marginal social utility $V_i^{-\Gamma}$ of zero-receivers with high utility V_i (i.e. zero-receivers with high productivity θ_i). In case the social welfare function is infinitely concave ($\Gamma \to +\infty$), then the planner puts infinite weight on the least productive zero-bequest receivers, so that s_{b0} is the average bequest taste within this group: $s_{b0} = E(s_{bi}|z_i = 0, \theta_i = \theta_0)$. This corresponds to what one might call the "radical Rawlsian optimum".[19]

The optimal tax formula is simple, intuitive, and can easily be calibrated using empirical estimates. First, a higher bequest elasticity e_B unsurprisingly implies a lower τ_B. As $e_B \to +\infty$, $\tau_B \to 0\%$, i.e. one should never tax an infinitely elastic tax base.

More interestingly, a higher bequest flow ratio b_y implies a higher τ_B.

Example 1. *Assume* $\tau = 30\%, \alpha = 30\%, s_{b0} = 10\%, e_B = 0$.
If $b_y = 20\%$, *then* $\tau_B = 73\%$ *and* $\tau_L = 22\%$.
If $b_y = 15\%$, *then* $\tau_B = 67\%$ *and* $\tau_L = 29\%$.
If $b_y = 10\%$, *then* $\tau_B = 55\%$ *and* $\tau_L = 35\%$.
If $b_y = 5\%$, *then* $\tau_B = 18\%$ *and* $\tau_L = 42\%$.

That is, with high bequest flow $b_y = 20\%$, zero receivers want to tax inherited wealth at a higher rate than labor income (73% vs. 22%); with low bequest flow $b_y = 5\%$, they want the opposite (18% vs. 42%).

The intuition is the following. In societies with low b_y (typically because of high g), there is not much tax revenue to gain from taxing bequests. So even zero-receivers do not like bequest taxes too much: it hurts their children without bringing much benefit in exchange. High growth societies care about the future, not about the past. Conversely, in societies with high b_y (typically because of low g), it is worth taxing bequests, so as to reduce labor taxation and to allow people with zero inheritance to live a better life - and in particular to accumulate wealth and leave a bequest (if they so wish).

[19] The most appealing welfare optimum is probably in between the meritocratic and the radical Rawlsian optimum (depending on how much one considers individuals are responsible for their productivity - a complex ethical issue on which this paper has nothing to say).

It is worth noting that the impact of b_y is quantitatively more important than the impact of e_B. That is, behavioral responses matter but not hugely as long as the elasticity is reasonable.[20]

Example 2. *Assume* $\tau = 30\%, \alpha = 30\%, s_{b0} = 10\%, b_y = 15\%$.

If $e_B = 0$, *then* $\tau_B = 67\%$ *and* $\tau_L = 29\%$.

If $e_B = 0.2$, *then* $\tau_B = 56\%$ *and* $\tau_L = 31\%$.

If $e_B = 0.5$, *then* $\tau_B = 46\%$ *and* $\tau_L = 33\%$.

If $e_B = 1$, *then* $\tau_B = 35\%$ *and* $\tau_L = 35\%$.

This is probably the most important lesson of this paper: once one allows the elasticity of capital supply to be a free parameter and to take moderate values (non-infinite), then one can naturally obtain fairly large values for socially optimal bequest tax rates. If we take $b_y = 15\%$ (current French level), then we find that as long as the elasticity e_B is less than one the optimal inheritance tax rate is higher than the optimal labor tax rate. In practice, this bequest elasticity effect e_B is also mitigated by the labor supply elasticity effect e_L , which further reinforces this conclusion (see section 5).

Finally, a higher bequest taste s_{b0} implies a lower τ_B. The key trade-off captured by our theory is that everybody is both a receiver and a giver of bequest (at least potentially). This is why zero receivers generally do not want to tax bequests at 100%. Of course if $s_{b0} = 0$ (zero receivers have no taste at all for leaving bequests), then we obtain $\tau_B = 1/(1 + e_B)$ as a special case: we are back to the classical revenue maximizing rule, and $\tau_B \to 100\%$ as $e_B \to 0$. But as long as $s_{b0} > 0$, we have interior solutions for τ_B, even if $e_B = 0$. In fact, for very high values of s_{b0}, and very low values of b_y, one can even get a negative τ_B, i.e. a bequest subsidy. For plausible parameter values, however, the optimal bequest tax rate τ_B is positive, and generally much larger than the optimal labor tax rate τ_L.[21]

3.4 Alternative social welfare weights

The main limitation of Proposition 2 is that it puts all the weight on the individuals who receive exactly zero bequest (possibly a very small group). But because real world inheritance is highly concentrated (basically half of the population receives very close to zero bequest), our optimal

[20] We leave a proper estimation of e_B to future research. Preliminary computations using time and cross section variations in French inheritance tax rates (e.g. in the French system childless individuals pay a lot more bequest taxes than individuals with children) suggest that e is relatively small (at most $e_B = 0.1 - 0.2$).

[21] See the discussion on bequest subsidies in appendix A2.

tax results are actually very robust to reasonable changes in the social welfare objective. We show this in two steps. First, the above formula can be extended in order to compute the optimal tax rate from the viewpoint of those inheriting $z\%$ of average inheritance:

Proposition 3 ***(z%-bequest-receivers optimum).*** *Under assumptions 1-3 , linear social welfare ($\Gamma = 0$), and the following welfare weights: $\omega_{z\theta} = 1$ for a given $z \geq 0$, and $\omega_{z'\theta} = 0$ if $z' \neq z$, then*

$$\text{(a)} \quad \tau_B = \frac{1 - (1 - \alpha - \tau)s_{bz}/b_y - (1 + e_B + s_{bz})z/\theta_z}{(1 + e_B + s_{bz})(1 - z/\theta_z)} \quad \text{and} \quad \tau_L = \frac{\tau - \tau_B b_y}{1 - \alpha},$$

with $s_{bz} = E(s_{bi}|z_i = z) =$ average bequest taste of z-receivers and $\theta_z = E(\theta_i|z_i = z) =$ average labor productivity of z-receivers (under no labor productivity memory special case: $\theta_z = 1$).

(b) *There is $z^* > 0$ such that $\tau_B > 0$ if and only if $z < z^*$.*

Proof. The proof is essentially the same as for Proposition 2. The formula can be extended to the case $\Gamma > 0$, and to any combination of welfare weights $(\omega_{z\theta})$: one simply needs to replace s_{bz}, z and θ_z by the properly weighted averages $\overline{s}_b$, $\overline{z}$, and $\overline{\theta}$. In case $\Gamma \to +\infty$, then for any combination of positive welfare weights $(\omega_{z\theta})$ (in particular for uniform utilitarian weights: $\omega_{z\theta} = 1 \ \forall z, \theta$) we have: $\overline{s}_b \to s_{b0} = E(s_{bi}|z_i = 0, \theta_i = \theta_0)$ and $\overline{z}/\overline{\theta} \to 0$, i.e. we are back to the radical Rawlsian optimum. See appendix proof for complete details. **QED**.

Unsurprisingly, individuals with higher z want lower bequest taxes. People who receive more than z^* do not want any bequest tax at all. If one cares mostly about the welfare of high receivers, then obviously one should not tax inheritance. Conversely, for individuals with very low z, the formula delivers optimal tax rates that are very close to the meritocratic Rawlsian optimum. One simple way to calibrate the formula is the following. The bottom 50% share in aggregate inherited wealth is typically about 5% (or less), which means that their average z is about 10%. Their average labor productivity θ_z is below 100% (bottom 50% inheritors also earn less than average), but generally not that much below, say at least 50% (which would imply that they are all fairly close to the minimum wage, i.e. that they almost perfectly coincide with the bottom 50% labor earners) and more realistically around 70%. As one can see, this has little impact on optimal tax rates. Inheritance is so concentrated that bottom 50% bequest receivers and zero bequest receivers have welfare maximizing bequest tax rates which are relatively close.

Example 3. *Assume $\tau = 30\%, \alpha = 30\%, b_y = 15\%, e_B = 0.2, s_{bz} = 10\%$.*

If $z = 0\%$, *then* $\tau_B = 56\%$ *and* $\tau_L = 31\%$.

If $z = 10\%$ *and* $\theta_z = 70\%$, *then* $\tau_B = 49\%$ *and* $\tau_L = 32\%$.

If $z = 10\%$ *and* $\theta_z = 50\%$, *then* $\tau_B = 46\%$ *and* $\tau_L = 33\%$.

Our optimal tax formulas show the importance of distributional parameters for the analysis of socially efficient capital taxation. They also illuminate the potentially crucial role of perceptions about distributions. If individuals have wrong perceptions about their position in the various distributions, this can have large impacts on their most preferred tax rate. E.g. with full information all individuals with inheritance below z^* should prefer a positive bequest tax. Interestingly, z^* is generally below one. I.e. people with inheritance above average do not want to tax inheritance at all. In actual fact, the distribution is so skewed that less than 20% of the population has inherited wealth above average. But to the extent that many more people believe to be above average, this might explain why (proportional) bequest taxes can have majorities against them.

In order to further illustrate the role played by distributional parameters, one can also rewrite the optimal tax formula entirely in terms of relative distributive positions:

Corollary 1 ***(z%-bequest-receivers optimum).*** *Under assumptions 1-3, linear social welfare ($\Gamma = 0$), and the following welfare weights: $\omega_{z\theta} = 1$ for a given $z \geq 0$, and $\omega_{z'\theta} = 0$ if $z' \neq z$, then:*

$$\text{(a)} \quad \tau_B = \frac{1 - e^{-(r-g)H}\nu_z x_z/\theta_z - (1+e_B)z/\theta_z}{(1+e_B)(1 - z/\theta_z)} \quad \text{and} \quad \tau_L = \frac{\tau - \tau_B b_y}{1-\alpha},$$

with $x_z = \dfrac{E(b_{t+1i}|z_i = z)}{b_{t+1}}$ $=$ *average bequest left by z-receivers/average bequest left*

$\nu_z = s_{bz}/s_z =$ *share of z-receivers wealth accumulation due to bequest motive*

(b) *If* $x_z \to 0$ *as* $z \to 0$, *then* $\tau_B \to 1/(1+e_B)$ *as* $z \to 0$, *(revenue maximizing tax rate)*

Proof. One simply needs to substitute $(1-\alpha-\tau)s_{bz}/b_y$ by $e^{-(r-g)H}\nu_z x_z/\theta_z - s_{bz}[\tau_B + (1-\tau_B)z/\theta_z]$ in the original formula. See appendix proof for details. **QED.**

By construction, both formulas are equivalent. Whether one should use one or the other depends on which empirical parameters are available. The original formula uses the aggregate inheritance flow b_y (a parameter that is relatively easy to estimate, since it relies only on aggregate data) and the bequest taste s_{bz} (a preference parameter that is relatively difficult to estimate).[22] The alternative formula is based almost entirely on distributional parameters which in principle can be estimated empirically - but require pretty demanding microeconomic data (such as wealth data spanning over two generations).[23] Its main advantage is that it illuminates the key role played by distribution for optimal capital taxation.

In particular, one can see that the optimal tax rate τ_B depends both on z (i.e. the distribution of bequests received) and on x_z (i.e. the distribution of bequests left). In case both distributions are infinitely concentrated, e.g. in case the share of bottom 50% successors in received *and* given bequests is vanishingly small, then the tax rate maximizing the welfare of this group converges towards the revenue maximizing tax rate $\tau_B = 1/(1+e_B)$. This is an obvious but important point: if capital is infinitely concentrated, then from the viewpoint of those who own nothing at all, the only limit to capital taxation is the elasticity effect. If the elasticity e_B is close to 0, then it is in the interest of the poor to tax the rich at a rate τ_B that is close to 100%.

We leave a proper empirical calibration of our optimal tax formula to future research. Here we simply illustrate the crucial role played by the distribution of x_z. If $x_z = 10\%$, i.e. if the children of bottom 50% successors receive as little as what their parents received (relatively to the average), then the optimal bequest tax rate is 77% for an elasticity $e_B = 0.2$ (it would be 95% with a zero elasticity). But if $x_z = 100\%$, i.e. if on average they receive as much as other children, then the optimal bequest tax rate is only 45%. Presumably the real world is in between, say around $x_z = 50\%$.

Example 4. *Assume* $\tau = 30\%, \alpha = 30\%, b_y = 15\%, e_B = 0.2, z = 10\%, \theta_z = 70\%, \nu_z = 50\%, r = 4\%, g = 2\%, H = 30$, so that $e^{(r-g)H} = 1.82$

If $x_z = 10\%$, *then* $\tau_B = 77\%$ *and* $\tau_L = 26\%$.

If $x_z = 50\%$, *then* $\tau_B = 61\%$ *and* $\tau_L = 30\%$.

If $x_z = 100\%$, *then* $\tau_B = 42\%$ *and* $\tau_L = 34\%$.

[22] Due to the relatively low quality of available fiscal inheritance data in most countries, it is actually not that simple to properly estimate b_y. The best way to proceed is to use national wealth estimates, mortality tables, age-wealth profiles and aggregate data on gifts. This is demanding, but this does not require micro data on wealth distributions. See Piketty (2011).

[23] High quality micro data on wealth spanning two generations is rarely available - and when it is available it usually does not include high quality data on labor income (see e.g. the micro data collected in Paris inheritance archives by Piketty et al. (2006, 2011), which can be used to compute x_z, but not θ_z). One can however obtain approximate estimates of the distributions x_z and θ_z using available wealth survey data. Note that the alternative formula also uses the preference parameter ν_z, which to some extent can be evaluated in surveys asking explicit questions about saving motives (and/or by comparing saving behavior of individuals with and without children). Yet another possibility is to set ν_z equal to one in order to get lower bounds for the optimal tax rate.

Note that our framework implicitly double counts welfare arising from bequest planning as bequests enter the utility of donors and enter the budget constraint of donees. When we focus on zero-receivers as in Proposition 2, then there is no double counting in social welfare as inheritances received are never counted. However, for z%-receivers, both bequests received and bequests left are counted. As discussed in the literature (e.g., Cremer and Pestieau, 2004 and Diamond, 2006), double counting raises issues as it can generate "free utility" devices by subsidizing giving and taxing back proceeds. This issue arises in our setting when social welfare weights are heavily tilted toward high z% receivers. Indeed, if $z > \theta_z$, then τ_B is no longer well defined as the government would want an infinite subsidy to bequest: it is always desirable for very high bequest receivers to decrease τ_B and increase τ_L.

In our view, double counting does shape the debate on the proper level of estate taxation: bequest taxes are opposed by both those receiving bequests and those planning to leave bequests, and the views of those voters will in part shape the social welfare objective of the government. In principle, for reasonable welfare criteria that do not put too much weight on high receivers, this issue should not arise. But there is so much uncertainty about the true parameters (not to mention the existence of self-serving beliefs) that it would be naive to expect a consensus to emerge about the proper level of inheritance taxation. Maybe our formulas can at least help to focus the public debate and future empirical research upon the most important parameters.

3.5 Non-linear inheritance taxes

Our basic optimal tax formula can also be extended to deal with non-linear bequest taxes. We now assume that the tax rate τ_B applies only above an exemption $b_t^* > 0$. Most estate or inheritance tax systems adopt such exemptions. The exemption is sometimes very high relative to average in countries such as the United States where less than 1% of estates are taxable, or more moderate as in France where a significant fraction of estates are taxable (typically 10%-20%).[24] Naturally $b_t^* = b^* e^{gHt}$ grows at rate g to ensure a steady state equilibrium. Denoting by B_t^* aggregate taxable bequests (i.e., the sum of $b_t - b_t^*$ across all bequests above b_t^*), the

[24]Note that in any case the fraction of the population paying bequest taxes is generally much less than 50% - a fact that must naturally be related to the high concentration of inherited wealth (the bottom 50% always receives barely 5% of aggregate inheritance, while the top 10% receives over 60% in Europe and over 70% in the U.S.). See Piketty (2011).

government budget constraint becomes

$$\tau_L(1-\alpha)+\tau_B b_y^* = \tau, \tag{8}$$

where $b_y^* = e^{rH} B_t^*/Y_t$ is capitalized taxable bequests over domestic product.

Let us denote by b_t^m the average bequest above b_t^*. That defines the Pareto parameter $a = b_t^m/(b_t^m - b_t^*)$ of the upper tail of the bequest distribution. Let us assume that in steady-state a fraction $p_t^* = p^*$ of individuals leave a bequest above b_t^*. We have $B_t^* = p^* \cdot b_t^* \cdot a/(a-1)$.

As above, we can define the elasticity e_B^* of taxable bequests with respect to $1-\tau_B$

$$e_B^* = \frac{db_y^*}{d(1-\tau_B)} \frac{1-\tau_B}{b_y^*} = a \cdot e^* \tag{9}$$

where e^* is the average elasticity (weighted by bequest size) of individual bequests b_{ti} above b_t^*. Empirical studies can in principle estimate e^* and a is directly observable from tabulated statistics by estate size.

With this nonlinear inheritance tax, we will also have a unique ergodic steady-state. The optimal non linear inheritance tax (for given threshold b^*, and from the viewpoint of zero bequest receivers) can be characterized as follows.

Proposition 4 ***(nonlinear zero-bequest-receivers optimum).*** *Under adapted ergodicity assumptions 1-3, and the following welfare weights:* $\omega_{z\theta} = 1$ *if* $z = 0$*, and* $\omega_{z\theta} = 0$ *if* $z > 0$*, then*

$$\tau_B = \frac{1-(1-\alpha-\tau)s_{b0}^*/b_y^*}{1+e^*+s_{b0}^*} \quad \text{and} \quad \tau_L = \frac{\tau - \tau_B b_y^*}{1-\alpha},$$

with $s_{b0}^* = E[(s_{bi}/s_i)(b_{t+Hi} - b_{t+H}^*)^+ | z_i = 0]/E(\tilde{y}_{ti}|z_i = 0]$ *= strength and likelihood that non-receivers will leave taxable bequests.*

Proof. The proof is similar to Proposition 2 and can be easily extended to the case of z-bequests-receivers. See appendix for details. **QED.**

Three remarks are worth noting. First, if zero-receivers never accumulate a bequest large enough to be taxable, then $s_{b0}^* = 0$, and the formula reverts to the revenue maximizing tax rate $\tau_B = 1/(1+e_B^*) = 1/(1+a \cdot e^*)$.[25] More generally, if zero-receivers have a very small probability to leave a taxable bequest (say, if b^* is sufficiently large), then s_{b0}^* is close to 0, and τ_B is close the revenue maximizing tax rate. This can be easily generalized to small z-receivers (say, bottom

[25] Note that the formula takes the same form as in standard optimal labor income tax theory. See Saez 2001.

50% receivers). If the elasticity is moderate (say, $e_B^* = 0.2$), then this implies the socially optimal inheritance tax rate on large bequests will be extremely high (say, $\tau_B = 70\% - 80\%$).

We believe that this model can help explain why very large top inheritance tax rates were applied in countries like the U.S. and the U.K. between the 1930s and the 1980s (typically around 70%-80%; see Figure 1 above). In particular, the fact that the rise of top inheritance tax rates was less dramatic in Continental Europe (French and German top rates generally did not exceed 30%-40%) seems qualitatively consistent with the fact these countries probably suffered a larger loss in aggregate inheritance flow ratios b_y and b_y^* following WW1 and WW2 capital shocks.[26]

Of course there are many cultural and/or political economy factors (and not just pure normative factors) which have certainly played an important role in order to account for the observed historical evolution of inheritance taxes. For instance, some political scientists have pointed out that bequest tax rates were relatively small pretty much everywhere prior to WW1, and then suddenly rose shortly after the war, which might be due to a "war mobilization" effect.[27] Others have stressed the possibility of a specific, radical U.S. preference for equal opportunity and highly progressive inheritance taxation.[28] All these factors probably mattered, and we certainly do not pretend that our optimal tax formulas alone can account for observed patterns. In particular, our formula point out for the crucial importance of beliefs about wealth inequality and mobility, and ideally one would need to explain where these beliefs come from and why they seem to change over time.[29] But at least our formula offer a theoretical framework which

[26] Note that the German top rate reached 60% in 1946-1948 when it was set by the Allied Control Council, and was soon reduced to 38% in 1949 when the Federal Republic of Germany regained sovereignty over its tax policy. One often stated argument was the need to favor reconstruction and new capital accumulation. See e.g. Beckert (2008).

[27] See Scheve and Stasavage (2011). One limitation of this theory is that if anything the war mobilisation effect should have been bigger in France and Germany than in the US or the UK. A related story would be that the Bolshevik revolution created a serious threat for wealth holders in western countries (it is safer to pay large inheritance taxes than to face a revolution; conversely the fall of the Soviet Union might contribute to explain the ideological shift since 1990). But again a limitation of this theory is that it is unclear why the Soviet threat should have had a bigger impact in the US and the UK.

[28] Apparently many U.S. economists and politicians were shocked to learn in the 1910s-1920s that wealth was almost as concentrated in America as in Old Europe - which might have prompted the rise of confiscatory top inheritance tax rates. See Beckert (2008). However this does not square with the fact that the rise also occurred in the UK.

[29] E.g. the decline in top inheritance tax rates observed in the U.S. and the U.K. since the 1980s might reflect the fact that more individuals now believe that they have a large probability to leave a high bequest (i.e. they believe in large s_{b0}^*). The dynamics of beliefs might itself be influenced by the evolution of inequality (when the rich have a higher share of income and wealth, it might be easier for them to spread their view in the media and via lobbyists and think tanks). This is far beyond the scope of the present paper.

one can use to think about the pros and cons of large top inheritance tax rates.

Second, from a more theoretical standpoint, as b^* grows, there are two options: either s^*_{b0}/b^*_y converges to zero or converges to a positive level. The first case corresponds to an aristocratic society where top bequests always come from past inheritances and never solely from self-made wealth. In that case again, the optimum τ_B should be the revenue maximizing rate. The second case corresponds to a partly meritocratic society where some of the top fortunes are self-made. In that case, even for very large b^*, non-receivers want a tax rate on bequests strictly lower than the revenue maximizing rate. In reality, it is probable that s^*_{b0}/b^*_y declines with b^* as the fraction of self-made wealth likely declines with the size of wealth accumulated. If the elasticity e^* and a are constant, then this suggests that the optimum τ_B increases with b^*. The countervailing force is that aristocratic wealth is more elastic as the bequest tax hits those fortunes several times across several generations, implying that e^* might actually grow with b^*.[30]

Third, one can also ask the question of what is the optimal b^* from the point of view of zero-receivers. Solving for the optimal b^* is difficult mathematically. If the optimal τ_B is zero when $b^* = 0$ (because zero-receivers care a lot of leaving bequests), then it is likely that τ_B will become positive when b^* grows (if society is relatively aristocratic). Then a combination $\tau_B > 0$ and $b^* > 0$ will be better that $\tau_B = 0$ and $b^* = 0$. The trade-off is the following: increasing b^* reduces the tax base b^*_y and hence estate tax revenue (with not much of a boost due to behavioral responses) so this is a negative. The positive is that it reduces s^*_{b0} (probably at a faster rate than b^*_y, allowing for a greater optimal τ_B.

Finally and more generally, real world estate tax systems generally have several progressive rates, and ideally one would like to solve for the full non-linear optimum. Unfortunately there is no simple formula for the optimal nonlinear bequest tax schedule. The key difficulty is that a change in the tax rate in any bracket will end up having effects throughout the distribution of bequests in the long-run ergodic equilibrium. This difficulty does not arise only in the simple case where there is a single taxable bracket. One needs to use numerical methods to solve for the full optimum. We leave this to future research.

[30] This is easily seen in a simple model with homogeneous labor income and individuals who value bequests or do not, and those tastes are iid in the population with no memory. In that case, inheritance level just reflects the number of prior generations in a row who were valuing bequests. The distribution of bequests is Pareto but the elasticity is growing with the size of the estate. See appendix for details.

4 From Inheritance Taxation to Capital Taxation

4.1 Are Inheritance Taxes and Lifetime Capital Taxes Equivalent?

So far we have focused upon optimal inheritance taxation. We have derived optimal tax formulas that can justify relatively large bequest tax rates, providing that the aggregate inheritance flow is sufficiently large. Typically, with inheritance flows b_y around 10%-15% of national income (as observed in today's developed economies, with a gradual upward trend), our formulas suggest that socially optimal bequest tax rates τ_B should be around 40%-60%, or even higher, thereby raising as much as 5%-8% of national income in annual tax revenues.

In the real world, we do observe total revenues from capital taxes of this order of magnitude, or even higher: currently about 9% of GDP in capital taxes in the European Union (out of a total of 39% of GDP in total tax revenues),[31] and about 8% of GDP in capital taxes in the U.S. (out of a total of about 27% of GDP in total tax revenues).[32] However only a small part comes from inheritance taxes - generally less than 1% of GDP. This reflects the fact that bequest tax rates are usually relatively small, except for very top (taxable) estates. E.g. in France the top statutory rate for children successors is currently 45%, but the average effective tax rates on bequest and gifts is around 5%. In today's developed economies, typically in the EU and in the US, the bulk of revenue from capital taxes comes from "lifetime capital taxes" (i.e. capital taxes paid during one's lifetime rather than at the time of wealth transmission). In practice, lifetime capital taxes can fall either on the capital stock (annual property and wealth taxes, with total revenues generally around 1%-2% of GDP) or on the capital income flow (taxes on corporate profits, taxes on rental income, interest, dividend and capital gains, with total revenues typically about 4%-5% of GDP). Empirical simulations in Piketty (2011) show that lifetime capital taxes have had a much larger historical impact than bequest taxes on the magnitude and evolution of aggregate inheritance flows.

Why do we observe so small inheritance taxes and so large lifetime capital taxes? In the simplest version of our model, all forms of capital taxation are equivalent, so the tax mix does

[31] The exact numbers vary from year to year, especially in recessions (total tax burden fell from 39.6% of EU GDP in 2006 and 2007 to 39.3% in 2008 and 38.4% in 2009, while capital taxes fell from 9.3% 2006 and 9.4% in 2007 to 8.9% in 2008 and 7.9% in 2009), but this is secondary here. See Taxation trends in the European Union, 2011 Edition, Eurostat, p.282 (total taxes) and p.336 (capital taxes), GDP-weighted EU 27 averages.

[32] Using OECD tax revenue statistics and summing up inheritance and property taxes, corporate taxes, capital gains taxes and capital income taxes (we attributed 20% of income tax revenues to capital), we find the following numbers for the for the U.S.: total tax revenues fell from 27.8% of GDP in 2006 and 2007 to 26.1% in 2008 and 24.1% in 2009, while capital taxes fell from 9.5% in 2006 and 9.4% in 2007 to 7.7% in 2008 and 7.4% in 2009.

not matter. First, it is equivalent for the government to tax at rate τ_B the capitalized value of bequests tax $b_t e^{rH}$ at the end of the period, or to tax at rate τ_B the raw value of bequest b_t at the beginning of the period, and then invest the tax revenue at market rate of return r until the end of the period. Next, and most importantly, rather than taxing bequests b_t at rate τ_B, it is also equivalent to tax the returns to capital Rb_t at rate τ_K such that:[33]:

$$(1-\tau_B)e^{rH} = 1 + (1-\tau_K)(e^{rH}-1) \quad \text{, i.e.} \quad \tau_K = \frac{\tau_B e^{rH}}{e^{rH}-1} = \frac{\tau_B(1+R)}{R}$$

Example 5. *Assume* $r = 4\%, H = 30$, so that $e^{rH} = 1 + R = 3.3$
If $\tau_B = 20\%$ *then* $\tau_K = 29\%$.
If $\tau_B = 40\%$ *then* $\tau_K = 57\%$.
If $\tau_B = 60\%$ *then* $\tau_K = 86\%$.

That is, it is equivalent to tax bequests at $\tau_B = 40\%$ or to tax capital income flows at $\tau_K = 57\%$ (or $\tau_K = 43\%$ if the we take the equivalent instantaneous tax rate).[34] More generally, any intermediate combination will do. I.e. for any tax mix (τ_B, τ_K), one can define $\overline{\tau}_B = \tau_B + (1-\tau_B)\tau_K \frac{R}{1+R}$. Intuitively, $\overline{\tau}_B$ is the adjusted bequest tax rate (including the tax on the return to bequest). It is equivalent to use any tax mix (τ_B, τ_K) delivering the same $\overline{\tau}_B$.

The reason why we get this general equivalence result between all forms of capital taxes in our simple model is because each generation lives only one period (which we interpret as H-year-long), with consumption taking place entirely at the end of the period. Zero-bequest receivers do save out of their labor income, but their savings go entirely to their children. Therefore taxing the returns to capital has exactly the same effect as taxing bequests (it has the same distributive effect, and the same distortionary effect on consumption vs bequest decisions).

[33] Here it is critical to assume that the utility function $V_{ti} = V(c_{ti}, w_{ti}, \bar{b}_{t+1i})$ is defined over after-tax *capitalized* bequest $\bar{b}_{t+1i} = (1-\tau_B)[1+(1-\tau_K)R]b_{t+1i}$. If V_{ti} were defined over after-tax *non-capitalized* bequest $\bar{b}_{t+1i} = (1-\tau_B)b_{t+1i}$, then zero-receivers would strictly prefer capital income taxes over bequest taxes (in effect $\tau_K > 0$ would allow them to tax positive receivers without reducing their utility from giving a bequest to their own children). However this would amount to tax illusion, so we rule this out.

[34] In the above equation we model the capital income tax τ_K as taxing the full generational return Rb_t all at once at the end of the period. Alternatively one could define τ_K as the equivalent instantaneous capital income tax rate during the H-year period, in which case the equivalence equation would be: $1 - \tau_B = e^{-\tau_K rH}$, *i.e.* $\tau_K = \frac{-\log(1-\tau_B)}{rH}$. Both formulas perfectly coincide for small tax rates and small returns, but differ otherwise. E.g. in the above example, we would have instantaneous $\tau_K = 19\%, 43\%, 76\%$ (instead of generational $\tau_K = 29\%, 57\%, 86\%$). Note that it would also be equivalent to have an annual wealth tax or property tax at rate $\tau_W = r\tau_K$ (with a fixed, exogenous rate of return, annual taxes on capital income flows and capital stocks are equivalent). In the above example, we would have instantaneous $\tau_W = 0.8\%, 1.7\%, 3.0\%$.

Clearly the conclusion would be different in a full-fledged, multi-period model with life-cycle savings.[35] Positive capital income taxes $\tau_K > 0$ would then impose extra distortions on intertemporal consumption decisions (within a given lifetime, including for individuals who do not receive and do not leave any bequest). Following the Atkinson-Stiglitz logic, it would generally be preferable to have $\tau_K = 0$ and to raise 100% of the capital tax revenue via a bequest tax $\tau_B > 0$. Of course, if the intertemporal elasticity of substitution is fairly small (as available estimates suggest), then this extra distortion would also be small, and both tax policies would be relatively close to one another. The point, however, is that in the real world we do observe a strong collective preference in favor of lifetime capital taxes (either stock-based or flow-based) over one-off bequest taxes, so there must be some substantial reasons for this fact. For instance, most individuals seem to prefer to pay an annual property tax equal to 1% of their property value (or 25% of their 4% annual return) during 30 years rather than to pay 30% of the property value all at once at the time they inherit the asset. Why is it so? More generally, the question we ask is the following: what are the extra benefits brought by lifetime capital taxes over bequest taxes which can counterbalance this extra intertemporal distortion and explain why the former are used more heavily than the latter?

There are potentially several factors which can play a role. It could be that real world individuals do not understand well the economic models at work, and/or are subject to various forms of tax illusion (e.g. maybe smaller taxes are less visible than big ones; or maybe people always prefer to pay taxes later in life rather than when they inherit). This can certainly be part of the explanation, especially in light of the extremely low observed levels of effective bequest tax rates (which might well be far below socially optimal levels). But we feel that there must also exist some additional, deeper reasons. One interesting argument put forward by Cremer, Pestieau and Rochet (2003) is that in case a comprehensive bequest tax is informationally or administratively impossible, then it is efficient to use capital income taxation. The problem is that it is not at all obvious that the latter is less informationally demanding than the former.

Here we explore two different mechanisms explaining why lifetime capital taxes are more heavily used than one-off inheritance taxes: the existence of a fuzzy frontier between capital income and labor income flows; and the existence of uninsurable idiosyncratic shocks to rates of return. Each mechanism allows us to explore different aspects of the optimal capital tax

[35] See section 5 below for such an extension.

mix. We certainly do not pretend that these are the only important factors. In particular, other forms of capital market imperfections, such as borrowing constraints, might well play an important role as well.[36]

4.2 Fuzzy frontier between capital and labor income flows

The simplest rationale for taxing capital income is the existence of a fuzzy frontier between capital and labor income flows. For instance, self employed individuals can to a large extent decide which part of their total compensation takes the form of wage income, and which part takes the form of dividends or capital gains. Opportunities for income shifting also exist for a large number of top executives (e.g. via stock options and capital gains). If the gap between the labor income tax rate $\tau_L \geq 0$ and the capital income tax rate $\tau_K \geq 0$ is substantial, then it is likely that many taxpayers will re-arrange their business and compensation package so as to minimize their tax burden. There is extensive empirical evidence that income shifting is a significant issue, and accounts for a large fraction of observed behavioral responses to tax changes.[37] At some level, this fuzzy-frontier problem can be viewed as the consequence of capital markets imperfections. With first-best markets, full financial intermediation and complete separation of ownership and control, there should be no problem to distinguish the returns to capital services from the returns to labor services.

For simplicity, here we assume "full fuziness": individuals can costlessly and limitlessly shift their labor income flows into capital income flows, and vice versa. That is, both income flows are informationally undistinguishable for the tax administration, so both tax rates have to be the same: $\tau_L = \tau_K = \tau_Y$, where $\tau_Y \geq 0$ is the comprehensive income tax rate.[38] Under this assumption, our basic optimal tax formula (Proposition 2) can be easily extended, and the new fiscal optimum can be characterized as follows:

Proposition 5 ***(comprehensive income tax cum inheritance tax).*** *Under the full-fuziness assumption, the zero-bequest-receivers optimum involves a bequest tax τ_B and a comprehensive*

[36] Borrowing constraints can be exacerbated by indivisibility problems and may force successors to quickly sell their property in order to pay large inheritance taxes. Anecdotal evidence suggests that this is an important reason why people dislike inheritance taxes (" death taxes ") and prefer to pay small property taxes and other lifetime capital taxes during 30 years rather than a large bequest tax all at once.

[37] See the recent survey by Saez, Slemrod, and Giertz (2011) for US evidence and Pirttila and Selin (2011) for an analysis of the dual income tax system introduced in Finland in 1993.

[38] Note this rationale for a comprehensive income tax treating equally all forms of income was often mentioned in the original Haig-Simons literature.

income tax $\tau_L = \tau_K = \tau_Y$ *such that:*

$$\tau_B = \frac{\overline{\tau}_B(1+R) - \tau_K R}{(1+R) - \tau_K R} \quad \text{and} \quad \tau_L = \tau_K = \tau_Y = \frac{\tau - \overline{\tau}_B b_y}{1-\alpha},$$

$$\text{With} \quad : \quad \overline{\tau}_B = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}}$$

Proof . The proof is essentially the same as Proposition 2, except that the steady-state government budget constraint is now: $\tau_L(1-\alpha) + \tau_B b_y + (1-\tau_B)\tau_K b_y \frac{R}{1+R} = \tau$. Define $\overline{\tau}_B = \tau_B + (1-\tau_B)\tau_K \frac{R}{1+R}$ the adjusted bequest tax rate (including the tax on the return to bequest). so that $\tau_L = \dfrac{\tau - \overline{\tau}_B b_y}{1-\alpha}$. After-tax, capitalized bequest left to next generation can be rewritten as follows: $\bar{b}_{t+1i} = (1-\tau_B)[1 + (1-\tau_K)R]b_{t+1i} = (1-\overline{\tau}_B)(1+R)b_{t+1i}$.The elasticity e_B of steady-state b_y with respect to $(1-\tau_B)$ (for given τ_K) is the same as the elasticity of b_y with respect to $(1-\overline{\tau}_B)$. To complete the proof, one simply needs to look at small changes in $d\overline{\tau}_B, d\tau_L$ rather than $d\tau_B, d\tau_L$ (see appendix proof for details). **QED**

Example 6. *Assume* $\tau = 30\%, \alpha = 30\%, s_{b0} = 10\%, e_B = 0$, and $r = 4\%, H = 30$, so that $e^{rH} = 1 + R = 3.3$

If $b_y = 20\%$, *then* $\overline{\tau}_B = 73\%$, *so that* $\tau_L = \tau_K = \tau_Y = 22\%$ *and* $\tau_B = 68\%$

If $b_y = 15\%$, *then* $\overline{\tau}_B = 67\%$, *so that* $\tau_L = \tau_K = \tau_Y = 29\%$ *and* $\tau_B = 59\%$

If $b_y = 10\%$, *then* $\overline{\tau}_B = 55\%$, *so that* $\tau_L = \tau_K = \tau_Y = 35\%$ *and* $\tau_B = 41\%$

If $b_y = 5\%$, *then* $\overline{\tau}_B = 18\%$, *so that* $\tau_L = \tau_K = \tau_Y = 42\%$ *and* $\tau_B = -16\%$

The impact of comprehensive income taxation is straightforward: it reduces somewhat the level of socially optimal inheritance tax rates. With plausible inheritance flows, say $b_y \simeq 10 - 20\%$, the socially optimal inheritance tax rate is typically much larger than the socially optimal labor tax rate, so that taxing capital income at the same rate as labor income has a relatively limited impact. For instance, with $b_y = 15\%$ (which is the level prevailing in France in 2010), then taxing capital income at rate $\tau_L = \tau_K = \tau_Y = 29\%$ implies that the bequest tax rate can be reduced from $\tau_B = 67\%$ to $\tau_B = 59\%$ In other words, a comprehensive income tax system reduces the need for inheritance taxation, but not by that much.

For very small inheritance flows, say $b_y = 5\%$ (prevailing in the 1950s-1970s in countries strongly hit by war destructions, such as France and Germany), the effect is much larger. Socially optimal inheritance tax rates are relatively small to start with, and should actually become negative in case we tax capital income at the same rate as labor income. This might explain the large number of exemptions for capital income that were created during the reconstruction period, particularly in countries like France or Germany. When inheritance flows are infinitely small, then one should exempt capital income from taxation in order to encourage new capital accumulation. The problem, however, is that this may lead to income shifting, and that such exemptions tend to persist over time (or actually to extend during the 1990s-2000s, in the context of tax competition), even though the reconstruction period is by now well over and inheritance flows are back to much higher levels.

Of course, whether income shifting is sufficiently massive to justify fully aligned income tax rates ($\tau_L = \tau_K = \tau_Y$) is very much an open empirical issue. Here we assumed full-fuzziness, so that the tax administration is forced to align rates. In practice, only a fraction of the population can easily shift capital into labor income (and vice-versa). This has to be weighted against costs of capital taxation in a model with life-cycle savings. Therefore the resulting optimal tax gap $\Delta\tau = \tau_L - \tau_K \geq 0$ should depend negatively on the fraction of income shifters and positively on the intertemporal elasticity of substitution.[39] Note also that the administrative capability to distinguish between capital and labor income flows and to impose separate tax rates is to some extent endogenous. E.g. it is easier if for the tax administration to observe or estimate capital income if taxpayers file annual wealth declarations in addition to annual income declarations.

4.3 Uninsurable idiosyncratic shocks to rates of return

We now assume away the fuzzy-frontier problem. That is, we assume that the administration can tax at separate rates capital and labor income flows, and we analyze the implications of uninsurable idiosyncratic shocks to rates of return for the optimal tax mix.

The basic intuition is straightforward. From a welfare viewpoint, as well as from an optimal tax viewpoint, what matters is capitalized bequest $\tilde{b}_{ti} = b_{ti} e^{r_{ti} H}$, not raw bequest b_{ti}. But at the time of setting the bequest tax rate τ_B, nobody has any idea about the future rate of return on a given asset is going to during the following 30 years. I.e. nobody knows what $e^{r_{ti} H}$ is going to be. Rates of return are notoriously difficult to predict, and they vary enormously over assets and across individuals. So it makes more sense to charge part of the tax burden via bequest

[39] Alternatively if one assumes a finite elasticity of income shifting with respect to the gap in tax rates, then the optimal tax gap will depend negatively on this elasticity (see Piketty, Saez, Stantcheva (2011)). Here we implicitly assumed an infinite elasticity, so that tax rates have to be exactly equal.

taxation τ_B, and part of the tax burden via lifetime capital taxation τ_K, - possibly a much larger part, in case the uncertainty about future returns is very large.

E.g. take someone who inherited a Paris apartment worth 100,000€(in todays euros) in 1972. At that time nobody could have guessed that this asset would worth one or two millions € by 2012. So instead of charging a very large bequest tax rate at the time of asset transmission, it might be collectively more efficient to charge a moderate bequest tax in 1972, and then to tax the asset continuously between 1972 and 2012, via property taxes and/or rental income taxes.

In order to capture this intuition and solve for optimal tax rates, we assume that individual life-time rates of returns $R_{ti} = e^{r_{ti}H} - 1$ vary across individuals. Let us denote by R the aggregate rate of return across all individuals. We assume that shocks R_{ti} are idiosyncratic so that there is no risk in aggregate.

In the case where R_{ti} is exogenous to the behavior of individuals, then it is clearly optimal for the government to set $\tau_K = 100\%$ to insure individuals against risky returns. In effect, the government is replacing risky individual returns R_{ti} by the aggregate return R , thereby providing social insurance. Standard financial models assume that individuals can insure themselves by diversifying their portfolios but in practice self-insurance is far from complete.

In order to make the problem non trivial (and more realistic), we introduce moral hazard, i.e. we assume that the individual random return $R_{ti}(e_{ti})$. depends on some individual, unobservable effort input e_{ti}. Importantly, we assume that the return conditional on effort remains stochastic so that the government cannot infer individual effort e_{ti} from observing individual capital income and the individual stock of wealth. Without loss of generality, assume a simple linear relationship between the probability R_{ti} to and effort e_{ti}:

$$R_{ti} = \xi e_{ti} + \varepsilon_{ti},$$

where ε_{ti} is a purely random iid component with mean $R_0 \geq 0$. Hence the expected return R is just equal to the product of effort productivity parameter ξ and effort e_{ti}. One can think of e_{ti} as the effort that one puts into portfolio management: how much time one spends checking stock market prices, looking for new investment opportunities, monitoring one's financial intermediaries and finding more performing intermediaries, etc. Parameter ξ measures the extent to which rates of return are responsive to effort. When ξ is close to 0, R_{ti} is almost a pure noise: returns are determined by luck. Conversely when ξ is large (as compared to the mean and variance of ε_{ti}), R_{ti} is determined mostly by effort.

We assume that the effort disutility cost $C(e_{ti})$ is proportional to portfolio size, so that in effect individuals with different levels of inherited wealth end up with the same distribution of returns (and in particular the same average return). That is, we assume $C(e_{ti}) = (1-\tau_B)b_{ti}c(e_{ti})$, where $(1-\tau_B)b_{ti}$ is portfolio size (net-of-tax bequest) and $c(e_{ti})$ is a convex, increasing function of effort.[40]

To simplify further the derivations, we assume that $C(e_{ti})$ enters the utility function as a monetary cost, so that the individual maximization programme and budget constraint look as follows:

$$\max V_{ti} = V(c_{ti}, w_{ti}, \bar{b}_{t+1i}) \quad \text{s.t.} \quad c_{ti}+w_{ti} \leq \widetilde{y}_{ti} = (1-\tau_B)[1+(1-\tau_K)R_{ti}]b_{ti}+(1-\tau_L)y_{Lti}-(1-\tau_B)b_{ti}c(e_{ti})$$

It follows that optimal effort $e_{ti} = e$ is the same for all individuals and is given by:

$$e_{ti} = e \quad \text{s.t.} \quad c'(e) = \xi(1-\tau_K)$$

From this, we can define e_R the elasticity of the aggregate rate of return $R = \xi e$ with respect to the net-of-tax rate $1-\tau_K$.[41] We view e_R as a free parameter, which can really take any value, and which in principle can be estimated empirically. So for instance if ξ is sufficiently small, i.e. if luck matters a lot more than effort in order to get a high return, then e_R can be arbitrarily close to zero. Conversely if ξ is sufficiently large, i.e. if returns are highly responsive to effort, then e_R can be arbitrarily large.[42]

Unsurprisingly, the optimal capital income tax rate τ_K. depends negatively upon the elasticity e_R. If e_R is close to zero, then the government should provide full insurance by taxing capital income at rate $\tau_K = 100\%$. Conversely, if e_R is sufficiently large, then the disincentive effects of taxing capital income are so large that one should have no capital income tax at al ($\tau_K = 0\%$). Unfortunately, there exists no simple closed-form formula for the intermediate case,

[40] It would be interesting to introduce scale economies in portfolio management (i.e. by assuming that cost rises less than proportionally with portfolio size), so as to generate the realistic prediction that higher portfolios tend to get higher returns (at least over some range). We leave this issue to future research.

[41] Alternatively, one could assume non-monetary disutility cost $C(e_{ti})$, so that individuals maximize $U_{ti} = V_{ti} - C(e_{ti})$. If $V_{ti} = V(c_{ti}, w_{ti}, \bar{b}_{t+1i})$ is homogeneous of degree one, we have $V^i = \kappa_i \cdot \tilde{y}_{ti}$, so that optimal effort e_{ti} is given by: $c'(e_{ti}) = \kappa_i \xi(1-\tau_K)$. So e_{ti} varies with individual taste parameters (and also with risk aversion, which needs to be introduced-otherwise idiosyncratic returns shocks do not matter; see appendix). This complicates the analysis and brings little additional insight.

[42] The elasticity e_R also depends on the curvature of the effort cost function. E.g. if $c(e) = e^{1+\eta}/(1+\eta)$, then $e = [\xi(1-\tau_K)]^{1/\eta}$, and $R = R_0 + \xi^{1+1/\eta}(1-\tau_K)^{1/\eta}$.

so one needs to use numerical solutions methods in order to calibrate the optimal tax rate, as is illustrated by the example below.

Proposition 6 ***(optimal capital income tax).*** *With uninsurable idiosyncratic shocks to rates of return, then the zero-bequest-receivers tax optimum involves a bequest tax τ_B, a capital income tax τ_K and a labor income tax τ_L such that:*

(a) If $e_R \rightarrow 0$, then $\tau_K \rightarrow 100\%$, $\tau_B \rightarrow \tau_{B0} = \dfrac{\overline{\tau}_B(1+R) - \tau_K R}{(1+R) - \tau_K R} = \overline{\tau}_B(1+R) - R < \overline{\tau}_B$ and $\tau_L \rightarrow \dfrac{\tau - \overline{\tau}_B b_y}{1-\alpha}$ (with $\overline{\tau}_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}}$)

(b) If e_R is sufficiently small, then $\tau_K > \tau_L$

(c) There exists $\overline{e}_R > 0$ s.t. if $e_R \rightarrow \overline{e}_R$, then $\tau_K \rightarrow 0\%$,, $\tau_B \rightarrow \overline{\tau}_B$ and $\tau_L \rightarrow \dfrac{\tau - \overline{\tau}_B b_y}{1-\alpha} > \tau_K$

(d) If e_R is sufficiently large, then $\tau_K < \tau_L$

Example 7. *Assume* $\tau = 30\%, \alpha = 30\%, s = 10\%, e_B = 0, z = 0\%, \theta_z = 100\%, \nu_z x_z = 50\%, r(\tau_K = 0\%) = 4\%, g = 2\%, H = 30$, so that $e^{(r-g)H} = 1.82$. Those simulations are done with MATLAB assuming $R_0 = 0$, $\xi = 1$ and iso-elastic cost of effort $c(R) = \underline{R} \cdot (R/\underline{R})^{1+1/e_R}/(1+1/e_R)$. See appendix for details.

If $e_R = 0.0$ then $\tau_K = 100\%$, $\tau_B = 9\%$, and $\tau_L = 34\%$.

If $e_R = 0.1$ then $\tau_K = 78\%$, $\tau_B = 35\%$, and $\tau_L = 35\%$.

If $e_R = 0.3$ then $\tau_K = 40\%$, $\tau_B = 53\%$, and $\tau_L = 36\%$.

If $e_R = 0.5$ then $\tau_K = 17\%$, $\tau_B = 56\%$, and $\tau_L = 37\%$.

If $e_R = 1$ then $\tau_K = 0\%$, $\tau_B = 58\%$, and $\tau_L = 38\%$.

These simulations rely on simplifying assumptions and should be viewed as illustrative and exploratory.

In particular, we know very little about what should be considered a reasonable value for the elasticity e_R of the macro rate of return R with respect to the net-of tax rate $1 - \tau_K$. Available macroeconomic evidence shows that aggregate rates of return, factor shares and wealth-income ratios are relatively stable over time and across countries, which -given that taxes vary a lot- would tend to suggest relatively low elasticities e_R (say, $e_R = 0.1 - 0.2$ at most).[43] This would

[43] It could be that when some individuals put higher effort e_{ti}, then the way they obtain higher returns R_{ti} is mostly at the expense of others, i.e. the aggregate R is very little affected. In the extreme case where this is a pure zero-sum game (R fixed), then the relevant elasticity is $e_R = 0$, and the optimal tax rate is $\tau_K = 100\%$. For a model based upon pure rent-seeking elasticities, see Piketty, Saez and Stantcheva (2011).

seem to imply that the optimal capital income tax rate is much larger than the optimal labor income tax rate. E.g. if $e_R = 0.1$ then in our simulations $\tau_K = 78\%$ and $\tau_L = 35\%$.

However our simulations also show that the results are very sensitive to the exact value of e_R . E.g. if $e_R = 0.5$ then capital income should be taxed much less than labor income: $\tau_K = 17\%$, and $\tau_L = 37\%$. This is because in the model a lower return R is not only bad for the capital income tax base: it also has a negative impact on the aggregate steady-state bequest flow b_y.[44]

In addition, these simulations do not take into account the distortionary impact of τ_K on intertemporal consumption allocation along the life-cycle (the magnitude of which depends on the intertemporal elasticity of substitution). A proper empirical calibration should take this into account. We leave this to future research.

The main contribution here is simply to provide a simple conceptual framework which can be used to think about the pros and cons of having $\tau_K > \tau_L$ or $\tau_K < \tau_L$. In particular, we clarify the conditions under which it might be optimal to tax capital income at very high rates.

It is interesting to note that the countries which have experienced very large top inheritance tax rates (particularly the U.S. and in the U.K. between the 1930s and 1980s; see Figure 1 above) also experienced very large top capital income tax rates (see Figures 2-3) In particular, during the 1970s, both the U.S. and the U.K. applied higher top rates on ordinary unearned income (such as capital income) than on earned income (i.e. labor income). One plausible way to account for this fact is to assume that policy makers had in mind a model very close to ours, with a relatively low elasticity of rates of return e_R with respect to effort, and with strongly meritocratic social preferences.[45]

More generally, $\tau_K > \tau_L$ was actually the norm in most income tax systems when the latter were instituted in the early 20th century (generally around 1910-1920). At that time income tax systems typically involved a progressive surtax on all forms of labor and capital income (including imputed rent), and a set of schedular taxes taxing wage income less heavily than interest, dividend, rent or business profit. In the more recent period, it has become more common to have $\tau_K < \tau_L$, via special tax exemptions for various categories of capital income.

[44] This explains why the elasticity $\overline{e}_R$ at which the optimal tax rate τ_K ceases to be positive is finite. See appendix.

[45] Another possible explanation is that U.K. and U.K. policy makers put lower social welfare weights on individuals with high return shocks ("lucky speculators") than on individuals with high labor income (strong meritocratic view). Yet another way to view the problem is to introduce speculative effort into the model (higher effort brings higher returns, but at the expense of others; see above).

But we feel that this mostly reflects a rising concern for international tax competition and tax evasion and the persistent lack of tax coordination (the view is that it is easier to reallocate one's financial portfolio abroad than one's labor income, and that it is harder to apply the residence principle of taxation for capital income; or at least this is a view that became very influential in a number of small open economies, typically in Nordic countries), rather than considerations about the global welfare optimum.

5 Extensions

5.1 Elastic Labor Supply

So far we assumed inelastic labor supply. We now show how the optimal labor and bequest tax rates should be set simultaneously in a model with elastic labor supply.

In order to ensure balanced growth path (and to avoid exploding labor supply), we need to assume a specific functional form for the disutility of labor:

$$U_i = V_i e^{-h_i(l)}, \text{ or, equivalently: } U_i = \log V_i - h_i(l)$$

where l is labor supply and $h_i(.)$ is increasing and convex (and could differ across individuals).

Individual i labor income is $y_{Lti} = v_t \theta_i l_i$ where θ_i is individual productivity (with mean one across the population) and $v_t = v_o e^{gHt}$ is the average wage rate of generation t.[46] We denote by $v_{ti} = (1 - \tau_L) v_t \theta_i$ the net-of-tax wage of individual i.

Individual i chooses b_{t+1i} and l_i to maximize:

$$\log V_i(v_{ti} l_i + (1 - \tau_B)(1 + R) b_{ti} - b_{t+1i}, b_{t+1i}, b_{t+1i}(1 - \tau_B)) - h(l_i)$$

Because V_i is homogeneous of degree one, we have $V_i = \kappa \cdot \tilde{y}_{ti}$ and hence

$$\log V^i - h(l_i) = cte + \log(v_{ti} l_i + \bar{b}_{ti}) - h(l_i),$$

where $\bar{b}_{ti} = (1 - \tau_B)(1 + R) b_{ti}$ is net-of-tax capitalized bequest (i.e. non-labor income). The first order condition for l_i is:

$$h'(l_i) = \frac{v_{ti}}{v_{ti} l_i + \bar{b}_{ti}}$$

[46] As discussed above $v_t = F_L = v_0 (1 + G)^t$ grows at rate $1 + G$ per generation.

Hence (uncompensated) labor supply $l_i = l(v_{ti}, \bar{b}_{ti})$ is a function of the net-wage and non-labor income homogeneous of degree zero. Hence, uniform growth in the wage rate and non-labor income leaves labor supply unchanged. Therefore, we can have a balanced growth path. $l(v_{ti}, \bar{b}_{ti})$ naturally increases with v_{ti} and decreases with $\bar{b}_{ti}$.

The government budget constraint defines τ_L as a function of τ_B as we had before. Consider a small reform $d\tau_B$ and let $d\tau_L$ be the required labor tax rate adjustment needed to maintain budget balance. Differentiating the government budget constraint, we have:

$$d\tau_L y_{Lt} + \tau_L dy_{Lt} + d\tau_B b_t + \tau_B db_t = 0,$$

which can be rewritten as:

$$d\tau_L y_{Lt}\left[1 - \frac{\tau_L}{1-\tau_L}e_L\right] = -d\tau_B b_t\left[1 - \frac{\tau_B}{1-\tau_B}e_B\right],$$

where

$$e_B = \frac{1-\tau_B}{b_t}\frac{db_t}{d(1-\tau_B)} \quad \text{and} \quad e_L = \frac{1-\tau_L}{y_{Lt}}\frac{dy_{Lt}}{d(1-\tau_L)},$$

are the elasticities of bequests and labor income with respect to their net-of-tax rates. Importantly, note that those elasticities are general equilibrium elasticities where both τ_L and τ_B change together to keep budget balance. $d\tau_L > 0$ and $d\tau_B < 0$ discourages labor supply through a reduction in the wage rate and through income effects as inheritances received are larger (Carnegie effect). $d\tau_B > 0$ and $d\tau_L < 0$ discourages bequests through the price effect but indirectly encourages bequests as individuals keep a larger fraction of their labor income.

Proposition 7 ***(zero-bequest-receivers optimum with elastic labor supply).*** *Under adapted ergodicity assumptions 1-3, and the following welfare weights:* $\omega_{z\theta} = 1$ *if* $z = 0$*, and* $\omega_{z\theta} = 0$ *if* $z > 0$*, then*

$$\tau_B = \frac{1 - (1 - \alpha - \tau \cdot (1+e_L))s_{b0}/b_y}{1 + e_B + s_{b0} \cdot (1+e_L)} \quad \text{and} \quad \tau_L = \frac{\tau - \tau_B b_y}{1-\alpha},$$

with $s_{b0} = E(s_{bi}|z_i = 0)$ = *the average bequest taste of zero bequest receivers (weighted by marginal utility×labor income).*

τ_B *increases with* e_L *iff* $\quad \tau(1+e_B) + s_{b0}(1-\alpha) \geq b_y$

If $e_L \to +\infty$ *(infinitely elastic labor supply), then* $\tau_B \to \tau/b_y$ *and* $\tau_L \to 0$

If $e_B \to +\infty$ *(infinitely elastic bequest flow), then* $\tau_B \to 0$ *and* $\tau_L \to \tau/(1-\alpha)$

If $s_{b0} = 0$ (zero-receivers have no taste for bequests), then $\tau_B = 1/(1+e_B)$ is the revenue maximizing rate.

Proof: The proof is similar to that of Proposition 2 (see appendix proof for complete details). **QED**

This formula is similar to the inelastic case except that e_L shows up both in the numerator and denominator. The inequality $\tau(1+e_B) + s_{b0}(1-\alpha) \geq b_y$ is very likely to be satisfied. E.g. if $\tau = 30\%$ and $b_y = 15\%$, it is satisfied even for $e_B = 0$ and $s_{b0} = 0$. That is, a higher labor supply elasticity e_L generally implies a higher bequest tax rate τ_B.

Intuitively, a higher labor supply elasticity makes high labor taxation less desirable, which for given aggregate revenue requirements makes the optimal tax mix tilt more towards bequest taxes (and more generally towards capital taxes in presence of capital market imperfections, which we do not model here in order to illuminate the pure labor supply effect).

If $s_{b0} = 0$, then we obtain again the revenue maximizing rate simple formula $\tau_B = 1/(1+e_B)$. The reason is the following: at $\tau_B = 1/1(1+e_B)$, we have $d\tau_L = 0$ for any small $d\tau_B$. Hence, the labor supply response becomes irrelevant.[47]

The following examples illustrate the quantitative impact of e_L. When both bequests and labor supply are elastic, the planner faces a race between two elasticities. In case labor is more elastic than bequests, then incentive effects reinforce the case for taxing labor income less than bequests. With $b_y = 15\%$ (current French level), it is clear that for reasonable elasticity values one wants to tax labor less than bequests. I.e. one would need very large bequest elasticities - above one - and zero labor supply elasticity to reverse this conclusion.

Example 8. *Assume $\tau = 30\%, \alpha = 30\%, s_{b0} = 10\%, b_y = 15\%$*

If $e_B = 0$ and $e_L = 0$, *then* $\tau_B = 67\%$ *and* $\tau_L = 29\%$.

If $e_B = 0$ and $e_L = 0.2$, *then* $\tau_B = 69\%$ *and* $\tau_L = 28\%$.

If $e_B = 0$ and $e_L = 1$, *then* $\tau_B = 78\%$ *and* $\tau_L = 26\%$.

If $e_B = 0.2$ *and* $e_L = 0$, *then* $\tau_B = 56\%$ *and* $\tau_L = 31\%$.

If $e_B = 0.2$ and $e_L = 0.2$, *then* $\tau_B = 59\%$ *and* $\tau_L = 30\%$.

If $e_B = 0.2$ and $e_L = 1$, *then* $\tau_B = 67\%$ *and* $\tau_L = 29\%$.

[47] This is analogous to the fact that $1/(1+e_L)$ in the revenue maximizing rate in optimal linear labor income taxation even if there are income effects.

5.2 Closed economy

So far we focused upon the small open economy case. I.e. we took as given the world instantaneous rate of return $r \geq 0$ (and the corresponding generational return $1 + R = e^{rH}$).

We now show that our optimal tax formulas also apply to the closed economy case.

In a closed economy, the domestic capital stock K_t is equal to domestic inheritance (i.e. $K_t = B_t$), and the generational rate of return $1 + R_t = e^{r_t H}$ is endogenously determined by the marginal product of domestic capital:

$$R_t = F_K = \frac{\alpha}{\beta_t}$$

With: $\beta_t = \dfrac{K_t}{Y_t} = b_{yt} e^{-r_t H}$= domestic capital-output ratio.

This can be rewritten: $\dfrac{R_t}{1 + R_t} = \dfrac{\alpha}{b_{yt}}$. I.e. closed economies with larger levels of capital accumulation and inheritance flows have lower rates of return.

The rest of the model is unchanged. Under assumptions 1-3, then for any given tax policy (τ_B, τ_L), we again have a unique long run steady-state: $b_{yt} \to b_y$, $R_t \to R$, $\Psi_t \to \Psi$ (Proposition 1). This follows from the fact in the open economy case the long run b_y is an increasing function of the exogenous rate of return R (i.e. long run capital supply is upward sloping). Since the demand for capital is downward sloping, there exists a unique long run rate of return R clearing the capital market: $\dfrac{R}{1 + R} = \dfrac{\alpha}{b_y}$.

The only difference with the open economy case is that a small tax change $d\tau_B > 0$ now triggers long run changes $dR > 0$ and $dv < 0$ (where $v = F_L$ is the wage rate). I.e. higher bequest taxes lead to lower capital accumulation (assuming $e_B > 0$), which raises the marginal product of capital and reduces the marginal product of labor. However the envelope theorem implies that these two effects exactly offset each other at the margin, so that the optimality conditions for (τ_B, τ_L) are wholly unaffected as in the standard optimal tax theory of Diamond and Mirrlees (1971), i.e. we keep the same optimal formulas as before (Proposition 2 and subsequent propositions). The important point is that the elasticity e_B entering the formula is the pure supply elasticity (i.e. not taking into account the general equilibrium effect), and similarly for the elasticity e_L in the case with elastic labor supply.

Proposition 8 ***(zero-bequest-receivers optimum with closed economy).*** *Under adapted ergodicity assumptions 1-3, and the following welfare weights:* $\omega_{z\theta} = 1$ *if* $z = 0$*, and* $\omega_{z\theta} = 0$ *if*

$z > 0$, *then*

$$\tau_B = \frac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}} \quad \text{and} \quad \tau_L = \frac{\tau - \tau_B b_y}{1-\alpha},$$

Proof. See appendix.

5.3 Overlapping generations and lifecycle savings

So far we focused upon a simple discrete time model where each generation lives for only one period (which we interpreted as H-year long, say $H = 30$). We assume that consumption took place entirely at the end of the period, so that in effect there was no lifecycle saving.

We now show that our results and optimal tax formulas can be extended to a full-fledged, continuous time model with overlapping generations and lifecycle savings. As far as optimal inheritance taxation is concerned, we keep the same closed-form formulas for optimal tax rates. Regarding optimal lifetime capital taxation, we keep the same general, qualitative intuitions, but one needs to use numerical methods in order to compute the full optimum.

We assume the following deterministic, stationary, continuous-time OLG demographic structure.[48] Everybody becomes adult at age $a = A$, has one kid at age $H > A$, and dies at age $D > H$. So everybody inherits at age $a = I = D - H > A$. E.g. if $A = 20$, $H = 30$ and $D = 70$, then $I = 40$. If $D = 80$, then $I = 50$. This is a gender free population.

For simplicity we assume zero population growth (at any time t, the total adult population N_t includes a mass one of individuals of age $a \in [A, D]$ and is therefore equal to $N_t = D - A$), and inelastic labor supply (each adult i supplies one unit of labor $l_{ti} = 1$ each period, so aggregate raw labor supply $L_t = N_t h_t = (D-A)h_0 e^{gt}$).

We denote by $\tilde{N}_t$ the cohort receiving inheritance at time t (born at time $t - I$). Each individual $i \in \tilde{N}_t$ solves the following finite-horizon maximization program:

$$\max V_{ti} = V(U_{ti}, w_{tiD}, \bar{b}_{t+Hi})$$

$$s.c. \quad \tilde{c}_{ti} + w_{tiD} \leq \tilde{y}_{ti} = (1-\tau_B)\tilde{b}_{ti} + (1-\tau_L)\tilde{y}_{Lti}$$

With: U_{ti} = utility derived from lifetime consumption flow $(c_{tia})_{A \leq a \leq D}$

[48] In order to obtain meaningful theoretical formulas for inheritance flows (i.e. formulas that can be used with real numbers), we need a dynamic model with a realistic age structure. Models with infinitely lived agents or perpetual youth models will not do, and standard two or three-period OLG models will not do either. Here we follow the continuous-time OLG model introduced by Piketty (2010, sections 5-7 and appendix E; 2011, section 5).

w_{tiD} = end-of-life wealth = b_{t+Hi} = pre-tax bequest left to next generation

$\bar{b}_{t+Hi} = (1-\tau_B)b_{t+Hi}e^{rH}$ = after-tax capitalized bequest left next generation

$\tilde{c}_{ti} = \int_{a=A}^{a=D} c_{tia}e^{r(D-a)}da$ = end-of-life capitalized value of consumption flow c_{tia}

$\tilde{y}_{ti}$ = end-of-life capitalized value of total lifetime resources

$\tilde{b}_{ti} = b_{ti}e^{r(D-I)} = b_{ti}e^{rH}$ = end-of-life capitalized value of received bequest b_{ti}

$\tilde{y}_{Lti} = \int_{a=A}^{a=D} y_{Ltia}e^{r(D-a)}da$ = end-of-life capitalized value of labor income flow y_{Ltia}

τ_B = bequest tax rate, τ_L = labor income tax rate

In the same way as in the discrete-time model, our optimal tax formulas hold for large classes of utility functions V_{ti} and U_{ti}, using a sufficient-statistics approach. Regarding U_{ti}, we assume that it is proportional to $\tilde{c}_{ti}$: $U_{ti} = \mu\tilde{c}_{ti}$. This holds if U_{ti} takes a standard discounted utility form $U_{ti} = [\int_{a=A}^{a=D} e^{-\delta(a-A)}c_{tia}^{1-\gamma}]^{\frac{1}{1-\gamma}}$, as well as for less standard (but maybe more realistic) utility specifications involving for instance consumption habit formation.[49] Regarding V_{ti}, for notational simplicity we again focus upon the Cobb-Douglas case:

$$V(U, w, \bar{b}) = U^{1-s_{bi}-s_{wi}}w^{s_{wi}}\bar{b}^{s_{bi}} \quad (s_{wi} \geq 0, s_{bi} \geq 0, s_i = s_{wi} + s_{bi} \leq 1)$$

This simple form implies that individual i devotes a fraction $s_i = s_{wi} + s_{bi}$ of his lifetime resources to end-of-life wealth, and a fraction $1 - s_i$ to lifetime consumption. Our results again hold with CES utility functions, and actually with all utility functions $V(U, w, \bar{b})$ that are homogenous of degree one

We also need to specify the lifetime structure of labor productivity shocks. To keep notations simple, we assume that that at any time t the average productivity h_t is the same for all cohorts, and that each individual i keeps the same within-cohort normalized productivity $\theta_{ia} = \theta_i$ during his entire lifetime.[50] So we have: $y_{Ltia} = \theta_i y_{Lt}e^{g(a-I)}$. It follows that the end-of-life capitalized value of labor income flows $\tilde{y}_{Lti}$ can be rewritten:

$$\tilde{y}_{Lti} = \theta_i\lambda(D-A)y_{Lt}e^{rH} \tag{10}$$

$$\text{with: } \lambda = \frac{e^{(r-g)(I-A)} - e^{-(r-g)(D-I)}}{(r-g)(D-A)}$$

[49] See Appendix A, proof of proposition 9.

[50] In effect we assume a flat, cross-sectional age-productivity profile at the aggregate level. The λ formula can easily be extended to non flat profiles (e.g. with replacement rate $\rho \leq 1$ above age retirement age $R \leq D$) and to more general demographic structures (e.g. with population growth $n \geq 0$).

Intuitively, λ corrects for differences between the lifetime profiles of labor income flows vs. inheritance flows (dollars received earlier in life are worth more). When labor income flows acrue earlier in life than inheritance flows then $\lambda > 1$ (and $\lambda < 1$ conversely with early inheritance). In practice, inheritance tends to happen around mid-life, and λ is typically very close to one.[51]

The individual-level transition equation for bequest is now the following:

$$b_{t+Hi} = s_i[(1-\tau_L)\widetilde{y}_{Lti} + (1-\tau_B)b_{ti}e^{rH}] \tag{11}$$

In the "no taste memory" case (tastes are drawn i.i.d. for each cohort), then by linearity the individual transition equation can be easily be aggregated into:

$$b_{t+H} = s[(1-\tau_L)\lambda(D-A)y_{Lt}e^{rH} + (1-\tau_B)b_t e^{rH}] \tag{12}$$

The aggregate bequest flow-domestic output ratio is defined by: $b_{yt} = \dfrac{B_t}{Y_t} = \dfrac{b_t}{N_t y_t} = \dfrac{b_t}{(D-A)y_t}$. Dividing both sides of the previous equation by per capita domestic output y_t, we obtain the following transition equation for b_{yt}:

$$b_{yt+H} = e^{(r-g)H}[s(1-\tau_L)\lambda(1-\alpha) + s(1-\tau_B)b_{yt}] \tag{13}$$

In case assumption 2 is satisfied, then $b_{yt} \to b_y = \dfrac{s(1-\tau_L)\lambda(1-\alpha)e^{(r-g)H}}{1-s(1-\tau_B)e^{(r-g)H}}$ as $t \to +\infty$.

I.e. we obtain exactly the same steady-state formula as in the discrete-time, one-period model, except for the correcting factor λ (which in practice is close to one).

Note that b_{yt} is now defined as the cross-sectional, macroeconomic ratio between the aggregate inheritance flow B_t transmitted at a given time t and domestic output Y_t produced at this same time t. This is approximately the cross-sectional inheritance-national income ratio plotted on Figures 4-5.[52]

We impose a cross-sectional government budget constraint:

$$\tau_L Y_{Lt} + \tau_B B_t = \tau Y_t$$

$$\text{i.e.}: \quad \tau_L(1-\alpha) + \tau_B b_y = \tau \tag{14}$$

In the no-taste-memory special case, the steady-state formula for b_y along a budged-balanced path can therefore be rewritten as follows:

[51]For detailed empirical calibrations and theoretical extensions of the λ formula, see Piketty (2010, sections 5-7, and appendix E, tables E5-E10).

[52]See section 3 above. National income is close to domestic output in countries with small net foreign asset position.

$$b_y \simeq \frac{s\lambda(1-\tau-\alpha)e^{(r-g)H}}{1-s[1+(\lambda-1)\tau_B]e^{(r-g)H}} \tag{15}$$

It follows that the long run elasticity e_B of b_y with respect to $1-\tau_B$ is positive if $\lambda < 1$ (inheritance happens earlier in life than labor income receipts, so cutting bequest taxes is good for wealth accumulation), and negative if $\lambda > 1$. If inheritance happens around mid-life, then $\lambda \simeq 1$ and $e_B \simeq 0$. Of course, the Cobb-Douglas form and the no-taste-memory assumption are restrictive, and in general e_B could really take any value, just like in the discrete-time model.

Next, one can easily show that we obtain exactly the same optimal bequest tax formula for the continuous-time model with overlapping generations and lifecycle savings as in the simplified discrete-time model where each generation leaves only one period:[53]

Proposition 9 ***(continuous time model).*** *Under assumptions 1-3, linear social welfare ($\Gamma = 0$), and the following welfare weights: $\omega_{z\theta} = 1$ if $z = 0$, and $\omega_{z\theta} = 0$ if $z > 0$, then*

$$\tau_B = \frac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}} \quad \text{and} \quad \tau_L = \frac{\tau-\tau_B b_y}{1-\alpha}$$

with $s_{b0} = E(s_{bi}|z_i = 0) =$ the average bequest taste of zero bequest receivers.

Proof. The proof is essentially the same as proposition 2. See appendix for details. **QED**

Regarding optimal lifetime capital taxation, the key difference is that with lifecycle savings we now have an extra distortion. That is, positive tax rates on capital income $\tau_K > 0$ distort the intertemporal allocation of consumption $(c_{tia})_{A \le a \le D}$ within a lifetime. The magnitude of the associated welfare cost depends on the intertemporal elasticity of substitution $\sigma = 1/\gamma$ (which might well vary across individuals). As long σ is relatively small, the impact on our optimal capital tax results should be moderate. Unfortunately there does not seem to exist any simple closed-form formula taking these effects into account, so one needs to resort to numerical solutions. We leave this to future research.

In presence of lifecycle savings, one might also want to think about more imaginative optimal tax structures than the conventional taxes on inheritance receipts and labor and capital income flows that we have been considering so far.

[53] If we were to adopt a generational government budget constraint rather than a cross-sectional government budget constraint, then the λ correcting factor would enter the formula for the optimal τ_B. As long as $\lambda \simeq 1$ this would make little difference.

One might first think that the ideal tax system should be based upon total capitalized lifetime resources, i.e. individual $i \in \tilde{N}_t$ should pay tax $\tau_{ti} = \tau(\tilde{y}_{ti})$ (with $\tilde{y}_{ti} = \tilde{b}_{ti} + \tilde{y}_{Lti}$). Intertemporal consumption choices would be undistorted. There are three problems with this. First, one might want to put different welfare weights $\omega_{z\theta}$ on inherited and earned resources (meritocratic social preferences), in which case it is desirable to have separate tax schedules. Next, even with uniform welfare weights on all individuals with similar total resources, it might be desirable to have separate tax schedules because elasticities e_B and e_L are different. Finally, the problem with this solution is that one cannot wait until the end of life in order to compute individual capitalized resources and tax liabilities.

In the absence of idiosyncratic shocks to rates of returns (and assuming away fuzzy frontier problems), one should just charge the relevant bequest tax rate $\tau_B > 0$ at the time of inheritance, and have no capital income tax ($\tau_K = 0$). But with idiosyncratic shocks r_{it} one does not know the right rate of tax at the time of inheritance. With $e_R = 0$, then one should simply set $\tau_K = 100\%$ and replace r_{it} by the average return $r_t = E(r_{it})$, i.e. the government should provide safe savings accounts offering the average, macroeconomic rate of return to everybody. But with $e_R > 0$, one faces a complex trade-off between the e_R distortion and the σ distortion.

Alternatively, one might want to try tax differently the returns to inherited wealth and the returns to life-cycle wealth. In a way this is what existing tax systems attempt to do when they offer preferential tax treatment for particular forms of long term savings (pension funds). One could also try to generalize this by having individual wealth accounts where we recompute the updated capitalized value of inheritance each period and charge the correct extra tax (whether the individual saved or consumed the extra income). But this is fairly complicated, so it might be easier to tax all actual returns, especially if σ is small. These are important issues for future research.

5.4 Population growth

So far we assumed that all individuals had exactly one kid, so that population was stationary: $N_t = 1$. All results can be easily extended to a model with population growth.

I.e. assume that all individuals have on average $1 + N$ kids, so that population grows at rate $1 + N = e^{nH}$ per generation: $N_t = N_0 e^{nHt}$. E.g. if everybody has on average $1 + N = 1.5$ kids (i.e. $2(1 + N) = 3$ kids per couple), then total population rises by $N = 50\%$ by generation, i.e.

by $n = \log(1+N)/H = 1.4\%$ per year (with $H = 30$).

The rest of the model is unchanged. Average productivity h_t is again assumed to grow at some exogenous rate $1 + G = e^{gH}$ per generation: $h_t = h_0 e^{gHt}$. Aggregate human capital $L_t = N_t h_t = N_0 h_0 e^{(n+g)Ht}$. grows at rate $(1+N)(1+G) = e^{(n+g)H}$ per generation. Taking as given the world, generational rate of return $R = e^{rH} - 1$, profit maximization implies that the domestic capital input K_t is chosen so that $F_K = R$, i.e. $K_t = \beta^{\frac{1}{1-\alpha}} L_t$ (with $\beta = \frac{K_t}{Y_t} = \frac{\alpha}{R}$). So output $Y_t = \beta^{\frac{\alpha}{1-\alpha}} L_t = \beta^{\frac{\alpha}{1-\alpha}} N_0 h_0 e^{(n+g)Ht}$ also grows at rate $(1+N)(1+G) = e^{(n+g)H}$ per generation. So does aggregate labor income $Y_{Lt} = (1-\alpha)Y_t$. Per capita output, capital and labor income y_t, k_t, y_{Lt} ($= Y_t, K_t, Y_{Lt}$ divided by N_t.) grow at rate $1 + G = e^{gH}$.

Individuals' preferences over wealth accumulation $V_{ti} = V(c_{ti}, w_{ti}, \bar{b}_{t+1i})$ could well be correlated with their effective number of children $1 + N_{it}$. They could also have unequal tastes for their various children. As long as ergodicity assumptions 1-3 are satisfied, Proposition 1 holds, i.e. there exists a unique steady-state for the aggregate inheritance flow-output ratio b_y and for the joint inheritance-labor distribution $\psi(z, \theta)$. In the "no taste memory" special case, one can easily see that the transition equation for $b_{yt} = \frac{e^{rH} B_t}{Y_t}$. (where $B_t = N_t b_t$ is the aggregate bequest flow received by generation t) now looks as follows:

$$b_{yt+1} = e^{(r-n-g)H}[s(1-\tau_L)(1-\alpha) + s(1-\tau_B)b_{yt}] \tag{16}$$

$$\text{So that:} \quad b_{yt} \to b_y = \frac{s(1-\tau_L)(1-\alpha)e^{(r-n-g)H}}{1 - s(1-\tau_B)e^{(r-n-g)H}}$$

I.e. one simply needs to replace the productivity growth rate g by the sum of population and productivity growth rates $n + g$. In societies with infinitely large population growth (i.e. where individuals have an infinite number of children), inheritance does not get you very far. Wealth gets divided so much between generations that one should rely on new output and large saving rates in order to become rich. The formula and intuition also work for countries with negative population growth (i.e. with $N < 0$), see Germany, Italy, Spain.

Next, one can see from the proof of Proposition 2 that our basic optimal tax formula, as well as all subsequent formulas, are wholly unaffected by the introduction of population growth. It follows that the impact of population growth on socially optimal tax policies is the same as the impact of productivity growth and goes through entirely via its impact on b_y. That is, high population growth countries should tax capital less, because capital accumulation is less inheritance-based and more labor-based and forward looking.

5.5 Dynamic efficiency and intergenerational redistribution

So far we imposed a period-by-period government budget constraint, i.e. we assumed that the government could not accumulate assets nor liabilities. This implies in particular that the government could not directly affect the aggregate level of capital accumulation in the economy, and hence could not address so-called "dynamic efficiency" issues.

In the technical appendix, we show that our results go through even if it can. That is, we allow the government to accumulate assets or liabilities, and we prove that the issue of the optimal capital vs. labor tax mix and the issue of dynamic efficiency and optimal aggregate capital accumulation are to a large extent orthogonal.

More precisely, we prove the following. In the small open economy case, unrestricted accumulation or borrowing by the government naturally leads to corner solutions. If the world rate of return r is larger than the Golden rule rate of return $r^* = \delta + \Gamma g$ (with δ = social rate of time preference and Γ = concavity of social welfare function),[54] then the government should accumulate infinite assets in order to have zero taxes in the long run. Conversely, in case $r < r^*$, the government should borrow indefinitely against future tax revenues. In both cases, the economy would cease to be a small economy at some point. In the closed economy case, the government will accumulate sufficient assets or liabilities to ensure that $r = r^*$, and will then apply the same optimal bequest and labor tax rates as in the case with a period-by-period budget constraint (except for the interest receipt or payment term).[55]

This is an important point, because both issues have sometime been mixed up. I.e. a standard informal argument in favor of small or zero capital taxation in the public debate is the view that there is insufficient saving and capital accumulation at the aggregate level.[56] This argument is flawed, for a number of reasons. First, there is no general presumption that there is too much or too little aggregate capital accumulation in the real world (it can go both ways, depending upon the parameters of the social welfare function). Next, even if we knew for sure that we are in a situation of excessive or insufficient aggregate capital accumulation, there

[54] With positive population growth, the Golden rule becomes $r^* = \delta + \Gamma g + \Gamma' n$ where $0 < \Gamma' < 1$ is the extent to which social welfare takes into account population growth (see appendix).

[55] See appendix B, proposition B3.

[56] E.g. it is hard to make sense of the famous Kaldor (1955) progressive consumption tax proposal without some kind of argument saying that there is too little aggregate saving. Note also that Kaldor proposed to use at the same time a progressive consumption tax and a progressive inheritance tax (this seems consistent with meritocratic welfare weights $\omega_{z\theta}$, but this is not made fully explicit by Kaldor).

would exist other and more efficient policy tools than the capital vs. labor tax mix in order to cope with this. Namely, the government should accumulate assets or liabilities (depending on whether there is too little or too much capital accumulation to start with), with little effect on optimal capital vs. labor tax formulas. Those issues have been addressed by King (1980) in the standard OLG model.

5.6 Uninsurable aggregate shocks to rates of return

It would be interesting to extend our results to aggregate, uninsurable uncertainty about the future rate of return (by definition, uncertainty at the world level is uninsurable). E.g. assume that r_t can take only two values $r_t = r_1 \geq 0$ and $r_t = r_2 > r_1$, keeps the same value for one generation (i.e. during H years), and follows a Markov random process with a switching probability equal to p between generations ($0 < p < 1$). We note: $e^{r_1 H} = 1+R_1 < e^{r_2 H} = 1+R_2$, The rest of the model is unchanged.

The first consequence is that instead of converging towards a unique steady-state inheritance ratio b_y and joint distribution $\psi(z,\theta)$ (Proposition 1), the economy now keeps switching between a continuum of values for b_{yt} and ψ_t. E.g. if the rate of return r_t has been low for an infinitely long time (which happens with an infinitely small probability), then b_{yt} is infinitely close to b_{y1} (the steady-state associated to stationary rate $r_t = r_1$). Similarly, if r_t has been high for an infinitely long time, then b_{yt} is infinitely close to $b_{y2} > b_{y1}$. In between these two extreme values, there is a distribution of b_{yt} in between these two values, depending on how much time the economy has spent with r_1 and r_2 in the recent past.

The second consequence is that socially optimal tax rates $\tau_{Lt}, \tau_{Bt}, \tau_{Kt}$ should now vary over time, and in particular should depend on b_{yt} and R_t. Intuitively, we expect the optimal tax mix to rely more on bequest taxes when the inheritance flow is large, and to rely more on capital income taxes when the rate of return is high. So the existence of aggregate returns shocks should in a way reinforce the results found under idiosyncratic returns shocks (see section 4.3). However it turns out that a complete analytical solution to this problem is relatively complicated. In particular one needs to specify whether we again have a generation-by-generation government budget constraint ($\tau_{Lt}(1-\alpha) + \tau_{Bt} b_{yt} + \tau_{Kt} b_{yt} \frac{R_t}{1+R_t} = \tau$), or whether we allow government to accumulate assets when returns are high and debts when they are low (which might seem natural). We leave this interesting extension to future research.

5.7 Endogenous growth

So far we assumed an exogenous productivity growth rate $g \geq 0$, and looked at how g affects aggregate steady-state bequest flows b_y and optimal tax rates τ_B. One might want to plug in endogenous growth models into this setting. By doing so, one could generate interesting two-way interactions between growth and inheritance.

E.g. with credit constraints, high inheritance flows can have a negative impact on growth-inducing investments (high-inheritance low-talent agents cannot easily lend money to low-inheritance high-talent agents). So high inheritance could lead to lower growth, which itself tends to reinforce high inheritance, as we see below. This two-way process can naturally generate multiple growth paths (with a high inheritance, high rate of return, low wealth mobility, low growth steady-state path, and conversely).[57] Tax policy could then have an impact on long run growth rates, e.g. a higher bequest tax rate might be a way to shift the economy towards a high mobility, high growth path.

The main difficulty with such a model would be empirical calibration. I.e. it is not too difficult to write a theoretical model with borrowing constraints and endogenous growth, but it is hard to find plausible parameters to put in the model. One particular difficulty is that basic cross-country evidence does not seem to bring much support to the view according to which tax policies entail systematic effects on long run growth rates. I.e. developed countries have had very different inheritance tax policies - and more generally very different aggregate tax rates and tax mix - over the past 100 years, but long run growth rates have been remarkably similar (as evidenced by convergence in per capita income and output levels - from Scandinavia to America). This explains why we chose in this paper to focus upon an exogenous growth model. Maybe a better way to proceed would be to keep growth exogenous, and to introduce the impact of borrowing constraints and inheritance on output and income levels. We leave this to future research.

5.8 Tax competition

Throughout this paper we assumed away tax competition. I.e. in the small open economy model we implicitly assumed that capital owners cannot or do not physically move to foreign

[57] See Piketty (1997) for a similar steady-state multiplicity.

countries, and that each country is able to enforce the residence principle of taxation (i.e. if its residents move their capital to foreign countries they still pay the same taxes).

Both hypotheses are highly questionable and rely on strong assumptions about international tax coordination. In particular, in order to properly enforce the residence principle of taxation, one needs extensive cooperation from other countries. E.g. if Germany or France or the U.S. want to tax their residents on the basis of the assets they own in Switzerland, then they need extensive, automatic information transmission from the Swiss tax administration, which they typically do not get. This clearly can put strong constraints on the capital tax rates that a given country can choose.

If we instead assume full capital mobility and tax competition between small open economies (zero international cooperation), then in equilibrium there would be no capital tax at all: $\tau_B = \tau_K = 0\%$. In the context of the Chamley-Judd or Atkinson-Stiglitz models where the optimal capital income tax is zero even absent tax competition, the presence of tax competition is not an issue but rather an additional reason for the government to remove capital taxes. However, in the context of our model where large capital and bequest taxes are desirable, such an uncoordinated tax competition equilibrium would be suboptimal in terms of social welfare. That is, the social welfare in each country would be larger–and, under plausible parameter values, substantially larger–under tax coordination.

6 Conclusion

In this paper, we have developed a tractable normative theory of socially optimal taxation. Our results challenge the conventional zero capital tax results, which in our view rely on ill-suited models. If one assumes from the beginning that there is zero or little inheritance, and that the bulk of wealth accumulation comes from life-cycle savings, then it is maybe not too surprising if one concludes that inheritance taxation is a secondary issue. If one assumes from the beginning that the long run elasticity of saving and capital supply is infinite, then it is maybe not too surprising if one concludes that taxing capital is a bad idea in the long run. Our model removes these assumptions, and shows that the optimal tax mix between labor and capital depends on the various elasticities at play and on critical distributional parameters.

At a deeper level, one of our main conclusions is that the profession's emphasis on $1 + r$ as a relative price is inappropriate, or at least excessive. We do not deny that capital taxation can entail distortions in the intertemporal allocation of consumption. But as long as the intertemporal elasticity of substitution is moderate, this effect is likely to be second order as compared to other effects. In particular, as far as capital taxes are concerned, distributional issues are likely to be first order. We hope our results will contribute to the emergence of more pragmatic debates about capital taxation, based more upon relevant empirical parameters and less upon abstract theoretical results.

A Omitted proofs for the main results

This appendix includes omitted proofs for the formal propositions stated in the main text of the paper.

A.1 Proof of Proposition 1 (convergence result) (section 3)

The four-dimensional, discrete-time stochastic process $X_{ti} = (z_{ti}, \theta_{ti}, s_{wt+1i}, s_{bt+1i})$ is a Markovian process with a state variable b_{yt}. It is governed by the exogenous transition functions $g(s_{wt+1i}, s_{bt+1i}|s_{wti}, s_{bti})$ and $h(\theta_{t+1i}|\theta_{wti})$ and by the following endogenous transition equation for normalized inheritance:

$$z_{t+1i} = s_i(1-\tau_L)(1-\alpha)\theta_{ti} + s_i(1-\tau_B)b_{yt}e^{(r-g)H}z_{ti} \tag{17}$$

The law of motion for b_{yt} is given by:

$$b_{yt+1} = s_\theta(1-\tau_L)(1-\alpha)e^{(r-g)H} + s_z(1-\tau_B)e^{(r-g)H}b_{yt}$$

where $s_\theta = E(s_i\theta_i)$ is the average saving taste weighted by normalized productivity, and $s_z = E(s_i z_i)$ is the average saving taste weighted by normalized inheritance. In the no-taste-memory special case (tastes are drawn iid at each generation), then $s_\theta = s_z = s$ is independent from the distribution.

Assume that the state variable b_{yt} converges towards a given b_y as $t \to +\infty$. Thanks to assumptions 1-3, the Markovian process verifies the following "concavity property": for any relative inheritance positions $0 \leq z_0 < z_1 < z_2$, there exists $T \geq 1$ and $\varepsilon > 0$ such that $proba(z_{it+T} > z_1 \mid z_{it} = z_0) > \varepsilon$ and $proba(z_{it+T} < z_1 \mid z_{it} = z_2) > \varepsilon$. In addition, the transitions are monotonic (i.e. $z_{t+1i}(z_{ti})$ dominates $z_{t+1i}(z'_{ti})$ in the first-order stochastic sense if $z_{ti} > z'_{ti}$). Therefore we can apply standard ergodic convergence theorems to derive the existence of a unique stationary distribution $\phi(z)$ towards which $\phi_t(z)$ converges, independently of the initial distribution $\phi_0(z)$ (see Hopenhayn and Prescott (1992, Theorem 2, p.1397) and Piketty (1997, Proposition 1, p.186)).

To complete the proof, one then needs to ensure that b_{yt} converges towards a unique b_y as $t \to +\infty$. This is trivial in the no-taste-memory special case. In the general case, one needs to generalize assumption 2 in order to ensure global convergence. E.g. one can assume that the taste distribution $g(s_{wt+1i}, s_{bt+1i}|s_{wti}, s_{bti})$ puts zero mass on all $s_i = s_{wi} + s_{bi} > \overline{s}$ (with

$\overline{s}(1-\tau_B)e^{(r-g)H}=1$). This is a relatively strong assumption. It can probably be relaxed to a weaker assumption (e.g. with positive but small mass on $s_i > \overline{s}$). In any case, note that this is relatively secondary for our purposes in this paper. I.e. even if there are multiple steady-state values for b_y , then our optimal tax formulas are valid as long the tax change does not shift the economy towards another b_y steady-state.

A.2 Proof of Proposition 2 (basic optimal tax formula) (section 3)

The proof is given in the main text of the paper (section 3). Here we simply discuss and clarify the conditions under which $\tau_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}} > 0$.

We have: $\tau_B > 0$ iff $b_y > s_{b0}(1-\alpha-\tau)$. Intuitively, if we start from $\tau_B = 0$ and $\tau_L = \tau/(1-\alpha)$, then $s_{b0}(1-\alpha-\tau) = s_{b0}(1-\alpha)(1-\tau_L)$ is the bequest-motive-driven fraction of income that zero-receivers are going to leave to their children; this measures how much τ_B is going to hurt them. On the other hand b_y measures how much fiscal resources the bequest tax is going to bring them in terms of reduced labor tax. So they want to introduce bequest taxation if and only if the latter is larger than the former.

In the Cobb-Douglas, no-taste-memory special case, then $s_{b0} = s$, and $b_y = \dfrac{s(1-\tau-\alpha)e^{(r-g)H}}{1-se^{(r-g)H}}$, so that we get the following the condition for $\tau_B > 0$:

$$\tau_B > 0 \text{ iff } (1+s)e^{(r-g)H} > 1$$

In particular, if $r-g>0$, as is generally the case in the real world, then we always have $\tau_B > 0$ In principle, in case g is sufficiently large as compared to r, then one could get a negative bequest tax, i.e. a bequest subsidy. However, $r-g<0$ would violate the transversality condition, i.e. our steady-state maximization problem could no longer be defined as the limit solution to an intertemporal maximization problem (see Appendix B). So in effect we always have $\tau_B > 0$ in the no-taste-memory case. The only robust way to get $\tau_B < 0$ would be assume relatively peculiar forms of random processes for taste parameters, e.g. one must assume that zero-receivers' bequest taste parameter s_{b0} is larger than the average saving taste s .

Note. The reason why $g >> r \to \tau_B < 0$ in the no-taste-memory formula follows from the fact that with a generational budget constraint $\tau_L y_{Lt} + \tau_B b_t e^{rH} = \tau Y_t$, then a bequest subsidy is equivalent to issuing public debt at the beginning of the period (which is the right thing to when $g >> r$). This possibility would disappear if we were to consider the following, alternative budget constraint: $\tau_L y_{Lt} + \tau_B b_{t+1} = \tau Y_t$ and hence $\tau_L(1-\alpha) + \tau_B b_y e^{-(r-g)H} = \tau$. [58] The basic optimal tax formula for zero-receivers (Proposition 2) would then be:

$$\tau_B^* = \frac{1-(1-\alpha-\tau)s_{b0}/(b_y e^{-(r-g)H})}{1+e_B+s_{b0}} \quad \text{and} \quad \tau_L^* = \frac{\tau - \tau_B b_y e^{-(r-g)H}}{1-\alpha},$$

The generational budget constraint $\tau_L y_{Lt} + \tau_B b_t e^{rH} = \tau Y_t$ is actually better suited for empirical calibrations, since it is closer to the cross-sectional budget balance constraint in the continuous time model with inheritance around mid-life (see section 5). I.e. if inheritance flows are received on average at the same time as labor income flows by a given generation, then the right way to model the budget balance case (zero intergenerational redistribution) is the generational budget constraint. However for the sake of completeness it is interesting to analyze the implications of the alternative budget constraint.

First, if $r > g$ then $\tau_B^* < \tau_B$.Intuitively, with $r > g$, bequests $b_{t+1} = b_t e^{gH}$ left by generation t are smaller than capitalized bequests $b_t e^{rH}$ received by generation t , so from the viewpoint of zero-receivers it is less desirable to tax bequest. Conversely, if $r < g$ then $\tau_B^* > \tau_B$.

Next, in the Cobb-Douglas, no-taste-memory special case, then by substituting $\tau_L(1-\alpha) = \tau - \tau_B b_y e^{-(r-g)H}$ into the steady-state formula for b_y (Proposition 1), we obtain:

$$b_y = \frac{s(1-\tau-\alpha)e^{(r-g)H}}{1-se^{(r-g)H}+s(e^{(r-g)H}-1)\tau_B} \tag{18}$$

It follows that $e_B = \dfrac{db_y}{d(1-\tau_B)}\dfrac{1-\tau_B}{b_y} > 0$ if $r > g$, and that $e_B < 0$ if $r < g$. Intuitively, here is what is going on. In the initial formulation, we had $e_B = 0$: with a generational budget constraint, the τ_B vs τ_L tax mix had no impact on the generational transition equation and on steady-state b_y. With the alternative formulation, and with $r > g$, then a higher reliance on bequest taxes is bad from the viewpoint of the generational budget constraint.

Finally, note that : $\tau_B^* > 0$ iff $b_y e^{-(r-g)H} > (1-\alpha-\tau)s_{b0}$. In the Cobb-Douglas, no-taste-memory special case, one can see that we always have $\tau_B^* > 0$, even if $g >> r$. In particular, in case $g \to +\infty$ and $b_y \to 0$, then $\tau_B^* \to 0$, but never becomes negative.

A.3 Proof of Proposition 3 (alternative welfare weights) (section 3).

The proof is the same as for Proposition 2 (see section 3), except that we now consider an individual i who receives positive bequest $b_{ti} = z_i b_t$, and with total after-tax lifetime income

[58] The same issue arises in optimal Ramsey taxation in a life-cycle model. This issue is discussed very clearly in King (1980).

$\widetilde{y}_{ti} = (1-\tau_B)(1+R)b_{ti} + (1-\tau_L)y_{Lti}$. Individual i chooses b_{t+1i} to maximize

$$V_i(\widetilde{y}_{ti} - b_{t+1i}, b_{t+1i}, (1-\tau_B)(1+R)b_{t+1i}).$$

The first order condition is again $V_{ci} = V_{wi} + (1-\tau_B)(1+R)V_{\bar{b}i}$ This leads to $b_{t+1i} = s_i \widetilde{y}_{ti}$ (with $0 \leq s_i \leq 1$) We can again define $\nu_i = (1-\tau_B)(1+R)V_{\bar{b}i}/V_{ci}$ the share of bequest left for bequest loving reasons, and $s_{bi} = \nu_i s_i$ the strength of the overall bequest taste.

The difference with the zero-receiver case is that the utility change dV_i created by a budget balance tax reform $d\tau_B, d\tau_L$ now includes an extra term:

$$dV_i = -V_{ci}y_{Lti}d\tau_L - V_{\bar{b}i}(1+R)b_{t+1i}d\tau_B - V_{ci}(1+R)b_{ti}(1+e_B)d\tau_B$$

The extra term corresponds to the extra tax paid on received bequest b_{ti}. This term includes a multiplicative factor $1+e_B$, because steady-state received bequest $b_{ti} = z_i b_t$ is reduced by $db_{ti} = -e_B z_i b_t d\tau_B/(1-\tau_B)$ (note that the elasticity is the same at any given normalized level z_i of the inheritance distribution).

Using the fact that $(1+R)b_{ti} = z_i b_y y_t$, this can re-arranged into:

$$dV_i = V_{ci}y_{Lt}d\tau_B \left[\left(1 - \frac{e_B\tau_B}{1-\tau_B}\right)\frac{\theta_i b_y}{1-\alpha} - \left(\frac{1-\tau_L}{1-\tau_B}\theta_i + \frac{z_i b_y}{(1-\alpha)}\right)s_{bi} - (1+e_B)\frac{z_i b_y}{(1-\alpha)}\right]$$

The first term in the square brackets is the utility gain due to the reduction in the labor income tax, the second term is the utility loss due to reduced net-of-tax bequest left, and the third term is the utility loss due to reduced net-of-tax bequest received. By using the fact that $1-\tau_L = (1-\alpha-\tau+\tau_B b_y)/(1-\alpha)$ (from the government budget constraint), this can further be re-arranged into:

$$dV_i = \frac{V_{ci}y_{Lt}d\tau_B}{(1-\tau_B)(1-\alpha)}\left[(1-(1+e_B)\tau_B)\, b_y\theta_i - (1-\alpha-\tau+\tau_B b_y)s_{bi}\theta_i - (1+e_B+s_{bi})(1-\tau_B)z_i b_y\right]$$

Summing up dV_i over all z-bequest-receivers, we get:

$$dSWF = \frac{V_{cz}\theta_z y_{Lt}d\tau_B}{(1-\tau_B)(1-\alpha)}\left[(1-(1+e_B)\tau_B)\, b_y - (1-\alpha-\tau+\tau_B b_y)s_{bz} - \frac{(1+e_B+s_{bz})(1-\tau_B)zb_y}{\theta_z}\right]$$

with $s_{bz} = \dfrac{\int_{z_i=z} V_{ci}\theta_i s_{bi}d\Psi}{\int_{z_i=z} V_{ci}\theta_i d\Psi}$, $\theta_z = \dfrac{\int_{z_i=z} V_{ci}s_{bi}\theta_i d\Psi}{\int_{z_i=z} V_{ci}s_{bi}d\Psi}$ and $V_{cz} = \dfrac{\int_{z_i=z} V_{ci}\theta_i d\Psi}{\theta_z}$

Setting $dSWF = 0$, we get the formula:

$$\tau_B = \frac{1 - (1-\alpha-\tau)s_{bz}/b_y - (1+e_B+s_{bz})z/\theta_z}{(1+e_B+s_{bz})(1-z/\theta_z)}$$

Note 1. This proof is a direct generalization of the proof of proposition 2 and also works with any utility function that is homogenous of degree one (and not only in the Cobb-Douglas or CES cases). In the general case, s_{bz} is the average of s_{bi} over all z-bequest-receivers, weighted by the product of their marginal utility V_{ci} and of their labor productivity θ_i, and θ_z is the average of θ_i over all z-bequest receivers, weighted by the product of their marginal utility V_{ci} and of their bequest taste s_{bi} In case $s_{bi} \perp V_{ci} y_{Lti}$ (e.g. in case there is no taste memory, or no labor productivity inequality), and in case the utility functions $V_i()$ are Cobb-Douglas, then s_{bz} is the simple average of s_{bi} over all z-bequest-receivers, and θ_z is the simple average of θ_i over all z-bequest receivers : $s_{bz} = E(s_{bi}|z_i = z)$ and $\theta_z = E(\theta_i|z_i = z)$.

Note 2. The optimal tax formula can be extended to the case $\Gamma > 0$, and to any welfare weights combination $(\omega_{z\theta})$. I.e. summing up dV_i over the entire distribution $\Psi(z,\theta)$, we have:

$$\begin{aligned}
dSWF &= \frac{\overline{V}_c \overline{\theta} y_{Lt} d\tau_B}{(1-\tau_B)(1-\alpha)} \left[(1-(1+e_B)\tau_B)\, b_y - (1-\alpha-\tau+\tau_B b_y)\overline{s}_b - \frac{(1+e_B+\overline{s}_b)(1-\tau_B)\overline{z} b_y}{\overline{\theta}} \right] \\
\text{with} &: \quad \overline{s}_b = \frac{\int\int_{z\geq 0,\theta_0\leq\theta\leq\theta_1} \omega_{z\theta} V_c \theta s_b V^{-\Gamma} d\Psi}{\int\int_{z\geq 0,\theta_0\leq\theta\leq\theta_1} \omega_{z\theta} V_c \theta V^{-\Gamma} d\Psi}, \ \overline{\theta} = \frac{\int\int_{z\geq 0,\theta_0\leq\theta\leq\theta_1} \omega_{z\theta} V_c \theta s_b V^{-\Gamma} d\Psi}{\int\int_{z\geq 0,\theta_0\leq\theta\leq\theta_1} \omega_{z\theta} V_c s_b V^{-\Gamma} d\Psi}, \\
\overline{z} &= \frac{\int\int_{z\geq 0,\theta_0\leq\theta\leq\theta_1} \omega_{z\theta} V_c z s_b V^{-\Gamma} d\Psi}{\int\int_{z\geq 0,\theta_0\leq\theta\leq\theta_1} \omega_{z\theta} V_c s_b V^{-\Gamma} d\Psi} \text{and } \overline{V}_c = \frac{\int\int_{z\geq 0,\theta_0\leq\theta\leq\theta_1} V_c \theta V^{-\Gamma} d\Psi}{\overline{\theta}}
\end{aligned}$$

Setting $dSWF = 0$, we get the formula:

$$\tau_B = \frac{1 - (1-\alpha-\tau)\overline{s}_b/b_y - (1+e_B+\overline{s}_b)\overline{z}/\overline{\theta}}{(1+e_B+\overline{s}_b)(1-\overline{z}/\overline{\theta})}$$

Note that for any combination of positive welfare weights $(\omega_{z\theta})$ (in particular for uniform utilitarian weights: $\omega_{z\theta} = 1 \ \forall z,\theta$), then as $\Gamma \to +\infty$, we have: $\overline{s}_b \to s_{b0} = E(s_{bi}|z_i = 0, \theta_i = \theta_0)$ and $\overline{z}/\overline{\theta} \to 0$, i.e. we are back to the radical Rawlsian optimum.

Note 3. The derivation of τ_B presented here and in the proof of Proposition 2 neglects the fact the steady-state distribution $\Psi(z,\theta)$ may also change in response to a small tax change $d\tau_B$, which could affect social welfare SWF via an extra term $dSWF$. However we feel that this extra term is not justified from a normative viewpoint (it is bizarre to let the planner affect the distribution of characteristics used to define social welfare), so we prefer in the general case to define the welfare weights ($\omega_{z\theta}$) "ex post" (i.e. taking as given the steady-state distribution $\Psi(z,\theta)$ induced by the optimal policy) The question of endogenous welfare weights raises complex conceptual issues and would deserve more attention in future research. In any case, note that this makes no difference as long as we assume no labor productivity memory (θ_{ti} randomly drawn for each generation t according to some fixed distribution $h(\theta)$) *and* we consider z-receivers optima (i.e. $\omega_{z\theta}=1$ for a given $z \geq 0$, and $\omega_{z'\theta}=0$ if $z' \neq z$) - since the sum $dSWF$ is computed over a fixed distribution $\Psi(z,\theta)=h(\theta)$. More generally, empirical evidence suggests that endogenous distribution effects are not very large - at least for the bottom segments of the distribution that are relevant for social welfare computations (i.e. the bottom 50% share in inherited wealth appears to be less than 5%-10% in every country and time period for which we have data, irrespective of the wide variations in bequest tax rates), so this extra term seems unlikely to have a large impact on socially optimal tax rates.

Proof of Corollary 1. The distributional formula can be derived in two alternative ways.

(i) First, starting from the original formula (Proposition 3), one can simply substitute $(1-\alpha-\tau)s_{bz}/b_y$ by $e^{-(r-g)H}\nu_z x_z/\theta_z - s_{bz}[\tau_B + (1-\tau_B)z/\theta_z]$. and obtain immediately the distributional formula:

$$\tau_B = \frac{1 - e^{-(r-g)H}\nu_z x_z/\theta_z - (1+e_B)z/\theta_z}{(1+e_B)(1-z/\theta_z)}$$

This substitution comes from the following algebra. I.e. consider an individual i receiving bequest $b_{ti} = z_i b_t$, and leaving bequest $b_{t+1i} = x_i b_{t+1}$. So we have:

$$b_{t+1i} = s_i[(1-\tau_L)\theta_i y_{Lt} + (1-\tau_B)z_i b_t e^{rH}] = x_i b_{t+1}$$

In steady-state we have $b_{t+1} = e^{gH} b_t = e^{-(r-g)H} b_y y_t$. Therefore the equation can be rearranged into:

$$s_{bi}[(1-\tau_L)(1-\alpha)\theta_i + (1-\tau_B)z_i b_y] = e^{-(r-g)H}\nu_i x_i b_y$$

Substituting $(1-\tau_L)(1-\alpha) = 1-\alpha-\tau+\tau_B b_y$, multiplying both sides by V_{ci} and summing up over all individuals with $z_i = z$, this gives:

$$\begin{aligned}
(1-\alpha-\tau)s_{bz}/b_y &= e^{-(r-g)H}\nu_z x_z/\theta_z - s_{bz}[\tau_B + (1-\tau_B)z/\theta_z] \\
\text{with} \quad : \quad s_{bz} &= \frac{\int_{z_i=z} V_{ci}\theta_i s_{bi} d\Psi}{\int_{z_i=z} V_{ci}\theta_i d\Psi},\ \theta_z = \frac{\int_{z_i=z} V_{ci} s_{bi}\theta_i d\Psi}{\int_{z_i=z} V_{ci} s_{bi} d\Psi}, \\
x_z &= \frac{\int_{z_i=z} V_{ci}\theta_i \nu_i x_i d\Psi}{\int_{z_i=z} V_{ci}\theta_i \nu_i d\Psi} \text{and } \nu_z = \frac{\int_{z_i=z} V_{ci}\theta_i \nu_i d\Psi}{\int_{z_i=z} V_{ci}\theta_i d\Psi}
\end{aligned}$$

(ii) Alternatively, one can return to the equation $dV_i = -V_{ci} y_{Lti} d\tau_L - V_{\bar{b}i}(1+R)b_{t+1i}d\tau_B - V_{ci}(1+R)b_{ti}(1+e_B)d\tau_B$. By substituting $b_{t+1i} = x_i b_{t+1} = x_i e^{gH} b_t$ and $y_{Lt} d\tau_L = -b_t e^{rH}(1 - \frac{e_B\tau_B}{1-\tau_B})d\tau_B$, we get:

$$dV_i = V_{ci} b_t e^{rH} d\tau_B \left[\left(1 - \frac{e_B\tau_B}{1-\tau_B}\right)\theta_i - e^{-(r-g)H}\frac{\nu_i x_i}{1-\tau_B} - (1+e_B)z_i \right]$$

Summing up over all individuals with $z_i = z$, this gives:

$$\begin{aligned}
dSWF &= V_{cz} b_t e^{rH} d\tau_B \left[\left(1 - \frac{e_B\tau_B}{1-\tau_B}\right)\theta_z - e^{-(r-g)H}\frac{\nu_z x_z}{1-\tau_B} - (1+e_B)z \right] \\
\text{i.e. } \tau_B &= \frac{1 - e^{-(r-g)H}\nu_z x_z/\theta_z - (1+e_B)z/\theta_z}{(1+e_B)(1-z/\theta_z)}
\end{aligned}$$

(iii) Finally, note that depending on the available parameters, one might prefer to express the optimal tax formula in yet another equivalent way. Namely, in the original formula (Proposition 3) one can replace s_{bz} by $s_{bz} = s \cdot x_z \cdot \nu_z/\pi_z$.[59] In words, the fraction of total resources specifically left for bequest motives s_{bz} by z%-inheritance receivers is equal to the product of fraction of total aggregate resources left (s), average bequest left by z-receivers/average bequest left (x_z), the share of z-receivers wealth accumulation due to bequest motive (ν_z), and divided by average total resources of z-receivers/average total resources (π_z).[60] We then get the following formula:

$$\tau_B = \frac{1 - \frac{s\cdot x_z \cdot \nu_z}{\pi_z b_y}(1-\alpha-\tau) - (1 + e_B + \frac{s\cdot x_z\cdot\nu_z}{\pi_z})z/\theta_z}{(1+e_B+\frac{s\cdot x_z\cdot \nu_z}{\pi_z})(1-z/\theta_z)}$$

By construction, all these formulas are fully equivalent.

[59] With: $\pi_z = E(\tilde{y}_{ti}|z_i = z)/\tilde{y}_t$ = average total resources of z-receivers/average total resources; and: $s ==$ $b_{t+1}/\tilde{y}_t$ = aggregate steady-state saving rate (bequests/lifetime resources).

[60] $s = b_{t+1}/\tilde{y}_t$ = aggregate steady-state saving rate (bequests/lifetime resources). In the no-taste-memory special case, $\pi_z = E(\pi_i|z_i = z)$ (with $\pi_i = \tilde{y}_{ti}/\tilde{y}_t$) = average total resources of z-receivers/average total resources. In the general case, $\pi_z = \frac{\int_{z_i=z} V_{ci}\theta_i\pi_i d\Psi}{\int_{z_i=z} V_{ci}\theta_i d\Psi}$ = average of π_i weighted by the product $V_{ci}\theta_i$.

A.4 Proof of Proposition 4 (non-linear inheritance taxes) (section 3).

The proof is similar to the proof of Proposition 2.

Consider a small increase in the bequest tax rate $d\tau_B > 0$ above b^*. In steady-state this allows the government to cut the labor tax rate by:

$$d\tau_L = -\frac{b_y^* d\tau_B}{1-\alpha}\left(1 - \frac{e^*\tau_B}{1-\tau_B}\right)$$

(< 0 as long as $\tau_B < 1/(1+e^*)$).

Consider an agent i with zero received bequest ($b_{ti} = 0$) and with total resources $\tilde{y}_{ti} = (1-\tau_L)\tilde{y}_{Lti}$. We have:

$$d\tilde{y}_{ti} = -\tilde{y}_{Lti} d\tau_L = \tilde{y}_{Lti}\frac{b_y^*[1-(1+e^*)\tau_B]}{1-\alpha}\frac{d\tau_B}{1-\tau_B}.$$

Replacing $1-\tau_L$ by $(1-\alpha-\tau+\tau_B b_y^*)/(1-\alpha)$, we have:

$$d\tilde{y}_{ti} = \tilde{y}_{ti}\frac{b_y^*[1-(1+e^*)\tau_B]}{1-\alpha-\tau+\tau_B b_y^*}\frac{d\tau_B}{1-\tau_B}$$

(> 0 as long as $\tau_B < 1/(1+e^*)$).

Agent i divides his lifetime resources $\tilde{y}_{ti}$ into lifetime consumption $\tilde{c}_{ti}$ and end-of-life wealth $w_{ti} = b_{t+1i}$ by maximizing $V_{ti} = V(c_{ti}, w_{ti}, b_{t+1i} - \tau_B(b_{t+1i} - b_{t+1}^*)^+)$. Using the envelope theorem, a change in $d\tau_B$ keeping $\tilde{y}_{ti}$ constant leads to a utility loss equal to $-V_b(b_{t+1i} - b_{t+1}^*)^+ d\tau_B$. The utility loss naturally is zero if the individual does not leave a bequest greater than b_{t+1}^*. The utility loss coming from $d\tilde{y}_{ti}$ is $V_{ci}d\tilde{y}_{ti}$.

For individuals leaving bequests above b_{t+H}^*, the first-order condition is $V_{ci} = V_{wi} + (1-\tau_B)(1+R)V_{bi}$, and one can again define $s_i = b_{t+1}/\tilde{y}_{ti}$ the fraction of life-time resources individual i devotes to wealth accumulation. Then, we can define: define $s_{wi} = s_i V_{wi}/V_{ci}$ and $s_{bi} = s_i(1-\tau_B)(1+R)V_{bi}/V_{ci}$. Hence, we have:

$$dV_i = V_{ci}d\tilde{y}_{ti} - V_{bi}(b_{t+1i} - b_{t+1}^*)^+ d\tau_B = V_{ci}\left[d\tilde{y}_{ti} - \frac{s_{bi}}{s_i}(b_{t+1i} - b_{t+1}^*)^+ \frac{d\tau_B}{1-\tau_B}\right],$$

$$dV_i = V_{ci}\frac{d\tau_B}{1-\tau_B}\left[\tilde{y}_{ti}\frac{b_y^*[1-(1+e_B^*)\tau_B]}{1-\alpha-\tau+\tau_B b_y^*} - \frac{s_{bi}}{s_i}(b_{t+1i} - b_{t+1}^*)^+\right].$$

Summing up over all zero-bequest-receivers, we get:

$$dSWF = \frac{d\tau_B}{1-\tau_B}\left[\frac{b_y^*[1-(1+e_B^*)\tau_B]}{1-\alpha-\tau+\tau_B b_y^*}\int_{z_i=0} V_{ci}\tilde{y}_{ti}d\Psi - \int_{z_i=0} V_{ci}\frac{s_{bi}}{s_i}(b_{t+1i} - b_{t+1}^*)^+ d\Psi\right],$$

Introducing

$$s_{b0}^* = \frac{\int_{z_i=0} V_{ci} \frac{s_{bi}}{s_i} (b_{t+1i} - b_{t+1}^*)^+ d\Psi}{\int_{z_i=0} V_{ci} \tilde{y}_{ti} d\Psi},$$

We have:

$$dSWF = \frac{d\tau_B}{1-\tau_B} \int_{z_i=0} V_{ci} \tilde{y}_{ti} d\Psi \left[\frac{b_y^* [1 - (1+e_B^*)\tau_B]}{1 - \alpha - \tau + \tau_B b_y^*} - s_{b0}^* \right],$$

Setting $dSWF = 0$, we get:

$$\tau_B = \frac{1 - (1 - \alpha - \tau)(s_{b0}^*/b_y^*)}{1 + e_B^* + s_{b0}^*} \quad \text{and} \quad \tau_L = \frac{\tau - \tau_B b_y^*}{1 - \alpha}.$$

A.5 Proof of Proposition 5 (fuzzy capital-labor frontier) (section 4)

Start from a tax mix (τ_B, τ_L, τ_K) satisfying the full-fuzziness constraint: $\tau_L = \tau_K = \tau_Y$, and consider a small change $d\tau_B > 0$ This allows the government to cut the labor tax rate by $d\tau_L$ s.t.:

$$d\tau_L = -\frac{b_y d\overline{\tau}_B}{1-\alpha} \left(1 - \frac{e_B \overline{\tau}_B}{1 - \overline{\tau}_B} \right) \quad \text{with} \quad d\overline{\tau}_B = d\tau_B (1 - \tau_K \frac{R}{1+R}) + (1 - \tau_B) d\tau_K \frac{R}{1+R} \quad \text{and} \quad d\tau_K = d\tau_L$$

An individual i who receives no inheritances ($b_{ti} = 0$) chooses b_{t+1i} to maximize

$$V_i((1-\tau_L) y_{Lti} - b_{t+1i}, b_{t+1i}, (1 - \overline{\tau}_B)(1+R) b_{t+1i}).$$

Therefore, using the envelope theorem as b_{t+1i} is optimized, the utility change dV_i created by $d\tau_B, d\tau_L$ is such that

$$dV_i = -V_{ci} y_{Lti} d\tau_L - V_{bi} b_{t+1i} d\overline{\tau}_B$$

Utility maximization first order condition in b_{t+1i} is $V_{ci} = V_{wi} + (1 - \overline{\tau}_B)(1+R) V_{bi}$ which leads to $b_{t+1i} = s_i \tilde{y}_{ti}$. We can then define $s_{bi} = s_i (1 - \overline{\tau}_B)(1+R) V_{bi}/V_{ci}$ (in the Cobb-Douglas utility, s_{bi} is a fixed exponent in the utility function, in the general homogeneous utility, s_{bi} depends on $\overline{\tau}_B$). Hence, we have $V_{bi} = (s_{bi}/s_i) V_{ci} / [(1 - \overline{\tau}_B)(1+R)]$ and

$$dV_i = V_{ci} \left[y_{Lti} (1 - \tau_L) \frac{-d\tau_L}{1 - \tau_L} - \frac{d\overline{\tau}_B}{1 - \overline{\tau}_B} \tilde{y}_{ti} s_{bi} \right].$$

Replacing $1 - \tau_L$ by $(1 - \alpha - \tau + \overline{\tau}_B b_y)/(1-\alpha)$, and $-d\tau_L$ by $b_y d\overline{\tau}_B (1 - \overline{\tau}_B (1+e_B))/[(1-\alpha)(1-\overline{\tau}_B)]$, we have:

$$dV_i = \frac{V_{ci} \tilde{y}_{ti} d\overline{\tau}_B}{1 - \overline{\tau}_B} \left[\frac{1 - (1+e_B)\overline{\tau}_B}{1 - \alpha - \tau + \overline{\tau}_B b_y} b_y - s_{bi} \right].$$

Summing up over all zero-bequest-receivers, and setting $dSWF = 0$, we get the formula: $\overline{\tau}_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e+s_{b0}}$. The corrected formula for τ_B then follows directly from $\overline{\tau}_B = \tau_B + (1-\tau_B)\tau_K \frac{R}{1+R}$ and $\tau_K = \tau_L = \dfrac{\tau - \overline{\tau}_B b_y}{1-\alpha}$.

Note. In the statement and proof of Proposition 5, we assume implicitly that the bequest tax is raised at the beginning of the period (with tax revenue invested by the government at rate R), while the capital income tax is raised at the end of the period on the return to net-of-tax bequest. Alternatively, one could assume that all taxes are raised at the same time (namely, at the end of the period: the bequest tax is raised at the end of the period on the capitalized value of bequests, and the capital income tax is an extra tax on the return to bequests). This is closer in spirit to the continuous time model, but this would be somewhat strange in the fuzzy-frontier context (since it would amounts to tax the return twice). Anyway, the formula would be changed as follows. The government budget constraint would be: $\tau_L(1-\alpha) + \tau_B b_y + \tau_K b_y \frac{R}{1+R} = \tau$. The adjusted bequest tax rate would be defined as: $\overline{\tau}_B = \tau_B + \tau_K \frac{R}{1+R}$, so that $\tau_L = \dfrac{\tau - \overline{\tau}_B b_y}{1-\alpha}$. After-tax, capitalized bequest left to next generation can be rewritten as follows: $\bar{b}_{t+1i} = [(1-\tau_B)(1+R) - \tau_K R]b_{t+1i} = (1-\overline{\tau}_B)(1+R)b_{t+1i}$. The formula for optimal $\overline{\tau}_B$ would be the same as before ($\overline{\tau}_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}}$). Same thing for $\tau_K = \tau_L = \dfrac{\tau - \overline{\tau}_B b_y}{1-\alpha}$ The only difference would be that we now have: $\tau_B = \overline{\tau}_B - \tau_K \frac{R}{1+R}$. As compared to the previous formula, we get lower bequest tax rates (this is simply because they apply to capitalized bequests rather than to raw bequests). E.g. in the example given in section 4.2, we would get $\tau_B = 58\%, 47\%, 31\%, -11\%$ (instead of $\tau_B = 68\%, 59\%, 41\%, -16\%$).

A.6 Proof of Proposition 6 (optimal capital income tax) (section 4)

TO BE COMPLETED

The proof follows immediately from a simple continuity result. I.e. with $e_R = 0$, then for any positive risk aversion level it is optimal to have full insurance ($\tau_K = 100\%$). So for e_R arbitrarily close to 0, then τ_K is arbitrarily close to 100%. The same continuity reasoning applies to $e_R = \overline{e}_R$ and $\tau_K = 0\%$. Note that $\overline{e}_R$ is finite because a lower return R is not only bad for the capital income tax base: it also has a negative impact on the aggregate steady-state bequest flow b_y

In order to solve the model numerically in the intermediate case, we need to specify the form of risk aversion. Of course risky returns are detrimental only if individuals are risk averse. A simple, albeit extreme, way to capture risk aversion is to posit that bequests leavers consider the worst possible scenario case where their heir will receive the worst possible return. Let us assume that the worst possible negative shock for ε_{it} is equal to $-\varepsilon_0 < 0$. We assume ε_0 to be exogenous and finite so that net capitalized bequests left are always positive even in the worst case scenario. For simplicity we also assume $R_0 = 0$ and $\xi = 1$.

Hence individual i choose b_{t+1i} to maximize

$$V^i[(1-\tau_B)[1+(\varepsilon_{ti}+R)(1-\tau_K)-c(R)]b_{ti}+(1-\tau_L)y_{Lti}-b_{t+1i}, b_{t+1i}, (1-\tau_B)b_{t+1i}(1+(R-\varepsilon_0)(1-\tau_K)-c(R))]$$

Recall that R is such that $c'(R) = 1-\tau_K$. We naturally assume that ε_{ti} is already realized when choosing b_{t+1i}. Assuming the worst possible return $R - \varepsilon_0$ is a useful short-cut to capture risk aversion for risky returns. In general, one could have used a concave utility and expectations and we could have defined $R - \varepsilon_0$ as the certainty equivalent rate of return. However, in that general case, ε_0 would depend on the complete structure of the model (including all tax rates), making the formulas much less tractable.

The first order condition for b_{t+1i} is such that

$$V_c^i = V_w^i + V_{\tilde{b}}^i(1-\tau_B)(1+(R-\varepsilon_0)(1-\tau_K)-c(R)) \quad \text{hence} \quad \nu_i = V_{\tilde{b}}^i(1-\tau_B)(1+(R-\varepsilon_0)(1-\tau_K)-c(R))/V_c^i$$

We also make the Cobb-Douglas utility assumption and assume that s_i is orthogonal to θ_i and z_i (no memory case). In that case, the first order condition in b_{t+1i} defines:

$$b_{t+1i} = s^i \cdot [(1-\tau_B)(1+(1-\tau_K)(\varepsilon_{ti}+R)-c(R))b_{ti} + (1-\tau_L)y_{Lti}]$$

which aggregates to

$$b_{t+1} = s \cdot [(1-\tau_B)(1+(1-\tau_K)R-c(R))b_t + (1-\tau_L)y_{Lt}]$$

The government budget constraint is

$$\tau_L y_{Lt} + \tau_B b_t \cdot [1+(1-\tau_K)R-c(R)] + \tau_K b_t R = \tau Y_t$$

where Y_t is defined such that $(1-\alpha)Y_t = y_{Lt}$. Here, we assume that the bequest tax is raised on capitalized bequests net of capital income taxes and net of costs to earn return R. As we shall see, this is the natural assumption to obtain a simple expression for b_t as it implies:

$$b_{t+1} = s \cdot [(1+R-c(R))b_t + y_{Lt} - \tau Y_t] \quad \text{and} \quad b_t = \frac{s(1-\alpha-\tau)Y_t}{1+G-s(1+R-c(R))}$$

which shows that b_t does not depend on τ_B (for fixed τ) so that $e_B = 0$ and depends upon τ_K only through R. We denote e_B^R the elasticity of b_t with respect to R. In the general case (not Cobb-Douglas and with potential memory, we still have R a function of τ_K only but b_t now depends in a complex way on both τ_K and τ_B (for a given τ), which complicates the formulas.

We derive the optimum for zero receivers. For zero receivers, the utility is:

$$V^i[(1-\tau_L)\theta_i y_{Lt} - b_{t+1i}, b_{t+1i}, (1-\tau_B)b_{t+1i}(1+(R-\varepsilon_0)(1-\tau_K)-c(R))]$$

Optimum τ_B. Consider a small reform $d\tau_B, d\tau_L$ that leaves the government budget constraint unchanged. As $e_B = 0$ and R depends solely on τ_K, we have $db_t = dR = 0$ and hence

$$-d\tau_L y_{Lt} = d\tau_B b_t \cdot [1 + (1-\tau_K)R - c(R)]$$

For zero receivers, the effect on utility is

$$dV^i = -d\tau_L y_{Lt}\theta_i - d\tau_B x_i b_{t+1}(1+(R-\varepsilon_0)(1-\tau_K)-c(R))V^i_{\bar{b}}$$

Using the definition of $\nu_i = V^i_{\bar{b}}(1-\tau_B)(1+(R-\varepsilon_0)(1-\tau_K)-c(R))/V^i_c$, we have

$$dV^i = d\tau_B b_t \cdot [1+(1-\tau_K)R - c(R)]V^i_c \left[\theta_i - \frac{\nu_i x_i}{1-\tau_B}\frac{1+G}{1+(1-\tau_K)R-c(R)}\right]$$

Therefore, the optimum τ_B for zero-receivers is such that:

$$\tau_B = 1 - \frac{\bar{\nu x}}{\bar{\theta}}\frac{1+G}{1+(1-\tau_K)R-c(R)}$$

This formula is the same as the standard formula in Proposition 2 with $e_B = 0$ but with the rate of return R replaced with the net-rate of return $(1-\tau_K)R - c(R)$. Naturally, with $\tau_K > 0$ and costs of getting return R, the net-return is less than the gross return R and hence τ_B is smaller relative to proposition 2.

Optimum τ_K. Consider a small reform $d\tau_K, d\tau_L$ that leaves the government budget constraint unchanged. We have (as $c'(R) = 1-\tau_K$):

$$-d\tau_L y_{Lt} = \tau_B db_t \cdot [1+(1-\tau_K)R - c(R)] + \tau_K db_t R + d\tau_K b_t(1-\tau_B)R + \tau_K b_t dR$$

As b_t depends on τ_K only through R, we have

$$\frac{1-\tau_K}{b_t}\frac{db_t}{d(1-\tau_K)} = e_B^R \cdot e_R \quad \text{with} \quad e_B^R = \frac{R}{b_t}\frac{db_t}{dR}$$

which implies that

$$-d\tau_L y_{Lt} = d\tau_K b_t R \left[1 - \tau_B - \frac{\tau_K}{1-\tau_K} e_R (1+e_B^R) - \frac{\tau_B e_R e_B^R}{(1-\tau_K)R}[1+(1-\tau_K)R - c(R)]\right]$$

For zero receivers, the effect on utility is

$$dV^i = -V_c^i d\tau_L y_{Lt}\theta_i - d\tau_K (1-\tau_B) x_i b_{t+1}(R-\varepsilon_0) V_b^i$$

$$dV^i = -V_c^i d\tau_L y_{Lt}\theta_i - d\tau_K V_c^i \frac{1+G}{1+(R-\varepsilon_0)(1-\tau_K) - c(R)}(R-\varepsilon_0)\nu_i x_i b_t$$

$$\frac{dV^i}{V_c^i \theta_i d\tau_K b_t R} = 1 - \tau_B - \frac{\tau_K e_R (1+e_B^R)}{1-\tau_K} - \frac{\tau_B e_R e_B^R [1+(1-\tau_K)R - c(R)]}{(1-\tau_K)R} - \frac{\frac{\nu_i x_i}{\theta_i}(1+G)\frac{R-\varepsilon_0}{R}}{1+(R-\varepsilon_0)(1-\tau_K) - c(R)}$$

which leads to the fairly complex optimal tax formula for τ_K:

$$\frac{\tau_K}{1-\tau_K} e_R (1+e_B^R) = 1 - \tau_B \left[1 + e_R e_B^R \frac{1+(1-\tau_K)R - c(R)}{(1-\tau_K)R}\right] - \frac{\frac{\bar{\nu x}}{\bar{\theta}}(1+G)\frac{R-\varepsilon_0}{R}}{1+(R-\varepsilon_0)(1-\tau_K) - c(R)}$$

If $e_R = 0$, then $\tau_K = 100\%$ and $\tau_B = 1 - \frac{\bar{\nu x}}{\bar{\theta}}(1+G)$

If $e_R > 0$, then $\tau_K < 100\%$ and τ_B decreases.

Note: with rent-seeking, if e_R is zero sum game, then trivially $e_R = 0$ and $\tau_K = 100\%$. The optimal tax formula is trivial because zero receivers do not get any extra-return capital income (hence we do not need to keep track of welfare effects at all). See Piketty, Saez and Stantcheva (2011).

These formulas can be solved numerically using MATLAB. In the simulation results presented in section 4.3 we assumed: $\varepsilon_0 = 0.6 \cdot R(\tau_K = 0)$.

A.7 Proof of Proposition 7 (elastic labor supply) (section 5)

With elastic labor supply, the most natural formulation for the government budget constraint is

$$\tau_L y_{Lt} + \tau_B b_t = \tau \bar{Y}_t,$$

where $\bar{Y}_t$ is an exogenous reference income (which grows at rate $1+G$ and independent of τ_B, τ_L). Otherwise the revenue requirements would vary with labor supply, which seems strange.[61]

[61] With inelastic labor supply, we could use actual domestic output Y_t which was independent of taxes.

It is also useful to introduce $\bar{\tau} = \tau \bar{Y}_t / Y_t$, the tax to output ratio (which is now endogenous) to rewrite the government budget constraint as:

$$\tau_L(1-\alpha) + \tau_B b_y = \bar{\tau},$$

We have:

$$U^i = \log V^i[(1-\tau_L)\theta_i v_t l_i + (1-\tau_B)(1+R)b_{ti} - b_{t+1i}, b_{t+1i}, b_{t+1i}(1-\tau_B)] - h(l_i)$$

Hence, using the envelope theorem as l_i and b_{t+1i} are optimized, we have:

$$dU^i = \frac{V_c^i}{V^i}\left[-d\tau_L y_{Lti} - d\tau_B(1+R)b_{ti} + (1-\tau_B)(1+R)db_{ti}\right] - \frac{V_b^i}{V^i} b_{t+1i} d\tau_B,$$

Using that $db_{ti} = -e_B^i b_{ti} d\tau_B/(1-\tau_B)$, $V_b^i = (s_{bi}/s_i)V_c^i/(1-\tau_B)$, and $b_{t+1i} = s_i \tilde{y}_{ti}$ we have:

$$dU^i = \frac{V_c^i}{V^i(1-\tau_B)}\left[-d\tau_L y_{Lti}(1-\tau_B) - d\tau_B(1+R)(1-\tau_B)(1+e_B^i)b_{ti} - \tilde{y}_{ti}s_{bi}d\tau_B\right],$$

$$dU^i = \frac{V_c^i d\tau_B}{V^i(1-\tau_B)}\left[-\frac{d\tau_L}{d\tau_B}\frac{1-\tau_B}{1-\tau_L}y_{Lti}(1-\tau_L) - (1+R)(1+e_B^i)(1-\tau_B)b_{ti} - \tilde{y}_{ti}s_{bi}\right],$$

Using the link between $d\tau_L$ and $d\tau_B$: $y_{Lt}d\tau_L(1-\tau_L(1+e_L))/(1-\tau_L) = -b_t d\tau_B(1-\tau_B(1+e_B))/(1-\tau_B)$, we have:

$$dU^i = \frac{V_c^i d\tau_B}{V^i(1-\tau_B)}\left[\frac{b_t}{y_{Lt}}\frac{1-(1+e_B)\tau_B}{1-(1+e_L)\tau_L}y_{Lti}(1-\tau_L) - (1+R)(1+e_B^i)(1-\tau_B)b_{ti} - \tilde{y}_{ti}s_{bi}\right],$$

We can use $b_y = b_t/Y_t = b_t(1-\alpha)/y_{Lt}$ and $(1-\alpha)\tau_L = \bar{\tau} - \tau_B b_y$ to get:

$$dU^i = \frac{V_c^i d\tau_B}{V^i(1-\tau_B)}\left[\frac{b_y[1-(1+e_B)\tau_B]}{1-\alpha-(1+e_L)(\bar{\tau}-\tau_B b_y)}y_{Lti}(1-\tau_L) - (1+R)(1+e_B^i)(1-\tau_B)b_{ti} - \tilde{y}_{ti}s_{bi}\right],$$

For zero receivers, we have $b_{ti} = 0$ and hence:

$$dU^i = \frac{V_c^i d\tau_B}{V^i(1-\tau_B)}\left[\frac{b_y[1-(1+e_B)\tau_B]}{1-\alpha-(1+e_L)(\bar{\tau}-\tau_B b_y)}\tilde{y}_{ti} - \tilde{y}_{ti}s_{bi}\right],$$

Setting $dSWF = 0$ for zero receivers, and defining

$$s_{b0} = \frac{\int_{z_i=0}(V_c^i/V^i)y_{Lti}s_{bi}d\Psi}{\int_{z_i=0}(V_c^i/V^i)y_{Lti}d\Psi},$$

we obtain:

$$0 = \frac{b_y[1-(1+e_B)\tau_B]}{1-\alpha-(1+e_L)(\bar{\tau}-\tau_B b_y)} - s_{b0}.$$

Rearranging, we obtain the formula in the proposition. The second part is straightforward.

A.8 Proof of Proposition 8 (closed economy) (section 5).

TO BE COMPLETED

With a closed economy, the most natural formulation for the government budget constraint is again: $\tau_L y_{Lt} + \tau_B b_t = \tau \bar{Y}_t$ (see proof of Proposition 7).

A.9 Proof of Proposition 9 (continuous-time model) (section 5).

The proof is the same as proposition 2, except that the time subscript t now denotes the time at which cohort $\widetilde{N}_t$ inherits. We simply need to show that $U_{ti} = \mu \tilde{c}_{ti}$ holds for various utility specifications. We consider two different possible specifications for utility function U_{Ci}:

$$U_{ti} = \left[\int_{a=A}^{a=D} e^{-\delta(a-A)} c_{tia}^{1-\gamma} \right]^{\frac{1}{1-\gamma}} \quad \text{(specification :: 1)},$$

with δ = rate of time preference.

γ = elasticity of marginal utility of consumption (=coefficient of relative risk aversion)

$\sigma = 1/\gamma$ = intertemporal elasticity of substitution

$$U_{Ci} = \left[\int_{a=A}^{a=D} e^{-\delta(a-A)} (\frac{c_{tia}}{q_{tia}})^{1-\gamma} \right]^{\frac{1}{1-\gamma}} \quad \text{(specification 2)},$$

with q_{tia} = individual consumption habit stock

Specification 1 corresponds to the standard discounted utility model. Specification 2 is less standard but in our view more realistic: it incorporates habit formation into the utility function (which one can also interpret as a concern for relative status or relative consumption), in the spirit of Carroll et al. (2000) (more on this below). Our results can also be extended to more general utility functions, e.g. a mixture of the two.

Specification 1. Under specification 1, standard first-order conditions imply that individual i chooses a consumption path $c_{tia} = c_{tiA} e^{g_c(a-A)}$growing at rate $g_c = \sigma(r - \delta)$ during his lifetime. The utility value $\mathrm{U_{Ci}}$ of this consumption path is given by:

$$U_{ci} = \left[\int_{a=A}^{a=D} e^{-\delta(a-A)} c_{tia}^{1-\gamma} da \right]^{\frac{1}{1-\gamma}} = \mu_c c_{tiA}$$

$$\text{with} \quad \mu_c = \left(\frac{1 - e^{-(\delta-(1-\gamma)g_c)(D-A)}}{\delta - (1-\gamma)g_c} \right)^{\frac{1}{1-\gamma}}.$$

Note that with $g_c = \sigma(r-\delta)$, we have $\delta - (1-\gamma)g_c = r - g_c$. So μ_c can also be rewritten:

$$\mu_c = \left(\frac{1-e^{-(r-g_c)(D-A)}}{r-g_c}\right)^{\frac{1}{1-\gamma}}.$$

The end-of-life capitalized value of individual i consumption flow $\tilde{c}_{ti}$ is given by:

$$\tilde{c}_{ti} = \int_{a=A}^{a=D} e^{r(D-a)} c_{tia} da = \tilde{\mu} c_{tiA},$$

$$\text{with} \quad \tilde{\mu} = e^{r(D-A)} \frac{1-e^{-(r-g_c)(D-A)}}{r-g_c}.$$

Therefore we have: $U_{ci} = \mu \tilde{c}_{ti}$

$$\text{with} \quad \mu = \frac{\mu_c}{\tilde{\mu}} = \left(\frac{1-e^{-(r-g_c)(D-A)}}{r-g_c}\right)^{\frac{\gamma}{1-\gamma}} e^{-r(D-A)} \quad \text{and} \quad g_c = \sigma(r-\delta)$$

So in effect the continuous-time maximization program can be re-written as a two-period maximization program:

$$\max V_{ti} = V(\mu \tilde{c}_{ti}, w_{tiD}, \bar{b}_{t+Hi})$$

$$\text{s.c.} \quad \tilde{c}_{ti} + w_{tiD} \leq \tilde{y}_{ti} = (1-\tau_B)\tilde{b}_{ti} + (1-\tau_L)\tilde{y}_{Lti}.$$

In the Cobb-Douglas case ($V(U, w, \bar{b}) = U^{1-s_{bi}-s_{wi}} w^{s_{wi}} \bar{b}^{s_{bi}}$), the μ term disappears, and we simply have: $\tilde{c}_{ti} = (1-s_i)\tilde{y}_{ti}$ and $w_{tiD} = b_{t+Hi} = s_i \tilde{y}_{ti}$ (with $s_i = s_{wi} + s_{bi}$).

In the CES case ($V(U, w, \bar{b}) = [(1-s_{wi}-s_{bi})U^{1-\bar{\gamma}} + s_{wi} w^{1-\bar{\gamma}} + s_{bi} \bar{b}^{1-\bar{\gamma}}]^{\frac{1}{1-\bar{\gamma}}}$), or in the general case with degree-one-homogeneity ($\forall \Lambda \geq 0, V(\Lambda U, \Lambda w, \Lambda \bar{b}) = \Lambda V(U, w, \bar{b})$), the μ term does not disappear, but the point is that it does not depend on tax rates, so in effect it cancels out from the first-order condition for optimal tax rates.

Specification 2. One unrealistic feature of specification 1 (making it ill-suited for empirical calibrations) is that it implies that countries with faster growth should save less. This is because the utility-maximizing consumption growth rate $g_c = \sigma(r-\delta)$ is independent from the economy's growth rate g, so in effect with high g and high expected lifetime income $\tilde{y}_{ti}$ young agents borrow a lot against future growth (i.e. they set c_{tiA} far above their current earnings y_{LtiA}). In practice consumption seems to track down income much more closely. The advantage of specification 2 is precisely that the habit formation term $q_i(a)$ provides a simple and plausible way to deliver consumption growth paths more in line with income growth. For notational simplicity we assume $q_i(a) = e^{qa}$ and consider the two following cases:

- case 2a: $q = \frac{\delta+\gamma g-r}{1-\gamma}$ (so that the utility-maximizing consumption growth rate is always exactly equal to the economy's growth rate: $g_c = g$)

- case 2b: $q = \frac{\gamma g}{1-\gamma}$ (so that: $g_c = g + \sigma(r-\delta)$)

In case 2a, the economy's saving rate is fully independent of its growth rate and of the rate of return, and is solely determined by the taste-for-wealth and taste-for-bequest parameters. In case 2b, utility maximizing consumption paths do react to changes in r, but in a reasonable way (i.e. with consumption growth rates around the economy's growth rate). This provides two useful benchmark points to which the results obtained under specification I can be compared. Our results could be extended to other intermediate specifications, as well as to more elaborate models with endogenous habit stock dynamics, such as those developed by Caroll et al. (2000), which can under adequate assumptions lead to the conclusion that countries with high growth rates save more (if anything, this seems more in line with observed facts than the opposite conclusion).

One can see that under both specifications 2a and 2b, U_{ci} can be written: $U_{ci} = \mu\tilde{c}_{ti}$, with $\mu = \dfrac{\mu_c}{\tilde{\mu}}$ given by the same formulas as before, except that one needs to replace $g_c = \sigma(r-\delta)$ by $g_c = g$ (case 2a) or $g_c = g + \sigma(r-\delta)$(case 2b).

B Extensions to dynamic efficiency and intergenerational redistribution

Our optimal tax results can be extended in order to analyze the interaction between optimal capital taxation and dynamic efficiency issues. These extensions are summarized in the main text of the paper (see section 5.3). Here we provide the formal statements and proofs.

B.1 Intertemporal social welfare function

We first need to properly specify the intertemporal social welfare function. In the main text of the paper, we solved a steady-state social welfare maximization problem. That is, we assumed that the government attempts to maximize the following, steady-state social welfare function (see section 3):

$$SWF = \iint_{z\geq 0,\underline{\theta}\leq\theta\leq\bar{\theta}} \omega_{z\theta}\frac{V_{z\theta}^{1-\Gamma}}{1-\Gamma}d\Psi(z,\theta) \tag{19}$$

With:

$V_{z\theta}$ = utility level V_i attained by individuals with normalized inheritance z and productivity θ

$\omega_{z\theta}$ = social welfare weights as a function of normalized inheritance z and productivity θ[62]

Γ = concavity of social welfare function ($\Gamma \geq 0$)[63]

$\Psi(z,\theta)$ = steady-state joint distribution of normalized inheritance z and productivity θ

We now consider the following intertemporal, infinite-horizon social welfare function:

$$SWF = \sum_{t=0,1,..,+\infty} \frac{\tilde{V}_t}{(1+\Delta)^t} = \sum_{t=0,1,...,+\infty} \tilde{V}_t e^{-\delta H t}$$

With: $1+\Delta = e^{\delta H}$ = social rate of time preference[64]

$\tilde{V}_t$ = social welfare of generation t , which can be written as follows:

$$\tilde{V}_t = \iint_{z\geq 0, \underline{\theta}\leq\theta\leq\bar{\theta}} \omega_{tz\theta} \frac{V_{tz\theta}^{1-\Gamma}}{1-\Gamma} d\Psi_t(z,\theta)$$

This intertemporal social welfare function might not be well defined (i.e. the intertemporal sum might be infinite). In order to ensure that it is well defined, we need to put constraints on parameters. First, note that the utility level $V_{tz\theta}$ attained by generation-t individuals with normalized inheritance z and productivity θ grows at the same rate as per capita output (for any z,θ). To see this, note that $V_{tz\theta}$ is given by the following maximization programme:

$$V_{tz\theta} = \max V(c_{tz\theta}, w_{tz\theta}, \bar{b}_{t+1z\theta}) \quad \text{s.t.} \quad c_{tz\theta} + w_{tz\theta} \leq \tilde{y}_{tz\theta} = (1-\tau_B) z b_t e^{rH} + (1-\tau_L)\theta y_{Lt}$$

As $t \to +\infty$, and under assumptions 1-3, $b_{yt} = B_t e^{rH}/Y_t = b_t e^{rH}/y_t \to b_y$, so that after-tax income $\tilde{y}_{tz\theta} \to [(1-\tau_B)zb_y + (1-\tau_L)\theta(1-\alpha)]y_t$. I.e. $\tilde{y}_{tz\theta}$ grows proportionally to per capita output $y_t = Y_t/N_t = y_0 e^{gHt}$. Since the utility function $V(c,w,\bar{b})$ is homogenous of degree one, it follows that all utility levels $V_{tz\theta}$ also grow at instantaneous rate g (i.e. at generational rate $1+G = e^{gH}$): as $t \to +\infty$, $V_{tz\theta} = \mu\tilde{y}_{tz\theta} \to \mu_{z\theta} y_t$. So $\Gamma \geq 0$ can be viewed as a parameter measuring the concavity of the social planner's preferences with respect to income.

[62]More generally, the welfare weights $\omega_{z\theta}$ could depend not only on normalized inheritance z and productivity θ, but also on other individual characteristics, such as taste parameters. This would complicate the notations without affecting any of the results.

[63]If $\Gamma = 1$, then $SWF = \iint_{z\geq 0, \underline{\theta}\leq\theta\leq\bar{\theta}} \omega_{z\theta} \log(V_{z\theta}) d\Psi(z,\theta)$.

[64]In the same way as for growth rates $1+G = e^{gH}$, population growth rates $1+N = e^{nH}$, rates of return $1+R = e^{rH}$, we use capital letters for generational rates and small letters for instantaneous rates: we note $1+\Delta = e^{\delta H}$ the generational social rate of time preference, and δ the corresponding instantaneous social rate of time preference. E.g. if $\delta = 1\%$ and $H = 30$ years, then $\Delta = 35\%$, i.e. from the social planner's viewpoint the welfare of next generation matters 35% less than the welfare of the current generation.

Next, a natural constraint to put on welfare weights $\omega_{tz\theta}$ is that their sum grows at rate $(1-\Gamma')n$, where n is the instantaneous population growth rate (possibly zero), and $\Gamma' \in [0,1]$ can be thought of as a parameter measuring the concavity of the social planner's preferences with respect to population size:

$$\iint_{z\geq 0,\underline{\theta}\leq\theta\leq\bar{\theta}} \omega_{tz\theta} d\Psi_t(z,\theta) = N_t^{1-\Gamma'} = N_0 e^{(1-\Gamma')nHt}$$

In case $\Gamma' = 0$, then this means that the sum of welfare weights grows at the same rate as population, so that in a sense the planner puts equal weight on each individual - whether they belong to small or large cohorts - so that larger cohorts matter more in terms of social impact. This is sometime called the "Benthamite" case in the literature: the planner cares about the total quantity of welfare, supposedly like Jeremy Bentham. Conversely, in case $\Gamma' = 1$, the sum of welfare weights is constant over time, i.e. the planner does not care about population size per se: he only cares about average welfare of each cohort (or on the normalized distribution of welfare within each cohort), and puts equal total weight on each cohort - irrespective of their size. This is the so-called "non-Benthamite" case.[65] Both approaches do have some merit - and so does the intermediate formulation with $\Gamma' \in [0,1]$. In this paper, we do not take a strong stand on this complex ethical issue. Nor do we take a strong stand about the income concavity parameter $\Gamma \geq 0$ (we later return to this normative discussion).

Our point here is simply that intertemporal social welfare SWF is well defined (non-infinite) if and only if:

Assumption 4 $\quad \delta + \Gamma g + \Gamma' n > g + n$

So for instance in case $\Gamma \geq 1$ (i.e. if the social welfare function is at least as concave as the *log* function), and $\Gamma' = 1$ (or $n = 0$), then this condition is satisfied for any $\delta > 0$, including for infinitely small rates of time preference ($\delta \to 0$). In our view, this is a desirable property for a social welfare function. We see no strong normative reason why δ should be strictly positive, i.e. why the social planner should put higher welfare weight on the present or nearby generation than on future or distant generations. As $\delta \to 0$, the social planner cares almost exclusively about the long-run, steady-state social welfare, which implicitly is the only case we considered so far.

[65] On Benthamite vs. non-Benthamite social welfare functions, see e.g. Blanchard and Fischer (1989, Chapter 2, pp. 39-45, notes 4 and 13).

B.2 Period-by-period government budget constraint

In the main text of the paper, we assumed the following period-by-period budget constraint:

$$\tau_{Lt}Y_{Lt} + \tau_{Bt}B_t e^{rH} = \tau Y_t \quad \text{i.e.}: \quad \tau_{Lt}(1-\alpha) + \tau_{Bt}b_{yt} = \tau$$

In Proposition 2 (and subsequent propositions), we solved for the stationary tax policy $(\tau_{Lt} = \tau_L, \tau_{Bt} = \tau_B)_{t\geq 0}$ maximizing steady-state social welfare. If we maintain the period-by-period budget constraint, but now consider intertemporal social welfare maximization, our results can be extended as follows.

First, if we allow for non-stationary tax policies $(\tau_{Lt}, \tau_{Bt})_{t\geq 0}$, then unsurprisingly it will generally be desirable to have higher bequest tax rates τ_{Bt} early on and then to let τ_{Bt} decline over time. This simply comes from the fact that the short run elasticity of the bequest flow is smaller than the long run elasticity. Indeed the elasticity of the initial bequest flow b_0 is equal to zero: capital in on the table and can be taxed at no efficiency cost.

Next, and more interestingly, as $t \to +\infty$, then $\tau_{Bt} \to \tau_B(\delta)$, and $\tau_{Lt} \to \tau_L(\delta)$, and as $\delta \to 0$, these limit tax rates converge towards the steady-state welfare optimum.

Proposition 10 . *With a period-by-period government budget constraint, and infinitely small rates of time preference ($\delta \to 0$), then the tax policy sequence $(\tau_{Lt}, \tau_{Bt})_{t\geq 0}$ maximizing intertemporal social welfare converges towards the steady-state welfare optimum:*

(1) As $t \to +\infty$, then $\tau_{Bt} \to \tau_B(\delta)$, and $\tau_{Lt} \to \tau_L(\delta)$

(2) As $\delta \to 0$, then $\tau_B(\delta) \to \tau_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}}$ and $\tau_L(\delta) \to \tau_L = \dfrac{\tau - \tau_B b_y}{1-\alpha}$

Proof. The proof is the same as Proposition 2. The formula for s_{bo} with $\Gamma > 0$ is given in proof of Proposition 2 (see appendix A2 above). If the social rate of time preference is vanishingly small ($\delta \to 0$), then the intertemporal maximization programme is equivalent to the steady-state maximization programme (in effect the planner puts infinite weight on the long run). **QED**.

B.3 Intertemporal government budget constraint

We now introduce the intertemporal government budget constraint. We start with the open economy case. The government can freely accumulate assets or liabilities at a given world rate of return $1 + R = e^{rH}$ We again assume an exogenous public good requirement $G_t = \tau Y_t$ each period (with $\tau \geq 0$). We assume zero initial government assets ($A_0 = 0$). The intertemporal government budget constraint can be written as follows:[66]

$$\sum_{t=0,1,..,+\infty} (\tau_{Lt} Y_{Lt} + \tau_{Bt} B_t e^{rH}) e^{-rHt} = \sum_{t=0,1,...,+\infty} \tau Y_t e^{-rHt}$$

Noting $\overline{\tau}_t = \tau_{Lt}(1-\alpha) + \tau_{Bt} b_{yt}$ the average effective tax rate imposed on generation t total resources (with $\overline{\tau}_t \geq 0$), this can be rewritten as follows:

$$\sum_{t=0,1,..,+\infty} \overline{\tau}_t Y_t e^{-rHt} = \sum_{t=0,1,...,+\infty} \tau Y_t e^{-rHt}$$

This budget constraint might not be well defined (i.e. the intertemporal sum might be infinite). For the sum to be well-defined, we must assume the standard transversality condition, according to which the rate of return r should be larger than the economy's growth rate $g + n$:

Assumption 5 $\quad r > g + n$

In case this assumption is not satisfied, i.e. in case $r < g + n$, then the net present value of future domestic output and tax revenue flows is infinite, so that the government would like to borrow indefinitely against future resources in order to finance current consumption. In principle, this should make the world net asset position decline in the long run (i.e. at some point the domestic economy would cease to be small), so that ultimately the world rate of return (the world marginal product of capital) should rise so as to restore $r > g + n$.

Given a tax policy sequence $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$, the net asset position A_t of the government at time t is equal to the capitalized value of previous primary surpluses or deficits: $A_{t+1} = (1 + R)A_t + (\overline{\tau}_t - \tau)Y_t$. The ratio between net government assets and domestic output $a_t = A_t/Y_t$ can be written as follows:

$$a_{t+1} = e^{(r-g-n)H} a_t + (\overline{\tau}_t - \tau) e^{-(g+n)H} \text{ i.e. } a_t = \sum_{s=0,1,..,t} (\overline{\tau}_s - \tau) e^{(r-g-n)H(t-s)}$$

[66] Here we assume a constant r. When r varies over time, one needs to replace e^{-rt} by $e^{-\int_{s=0}^{t} r_s ds}$. Note that the intertemporal budget constraint integral converges iff $r > n + g$ (if $r < n + g$ then the net present value of future output flows and government spending is infinite). More on this below.

Take any tax policy sequence $(\tau_{Bt}, \tau_{Lt})_{t\geq 0}$ satisfying the budget constraint and converging towards some asymptotic tax policy (τ_B, τ_L) as $t \to +\infty$. Under assumptions 1-3, $b_{yt} \to b_y$, and $\overline{\tau}_t \to \overline{\tau} = \tau_L(1-\alpha) + \tau_B b_y$ Then two cases can happen:[67]

(i) Either the government runs a long run primary deficit: $\bar{\tau} \leq \tau$. This deficit is financed by the returns to the government assets accumulated through initial primary surpluses: as $t \to +\infty$, $a_t \to \overline{a} \geq 0$

(ii) Or the government runs a long run primary surplus: $\bar{\tau} \geq \tau$. This surplus is used to finance the interest payments on the government debt accumulated through initial primary deficits: as $t \to +\infty$, $a_t \to \overline{a} \leq 0$.

In both cases, the long run government budget constraint and net government asset position can be written as follows:

$$\tau_L(1-\alpha) + \tau_B b_y + \bar{R}\overline{a} = \overline{\tau} + \overline{R}\overline{a} = \tau$$

$$\text{I.e.} \quad \overline{a} = \frac{\tau - \bar{\tau}}{\bar{R}}$$

Where $\bar{R} = e^{rH} - e^{(g+n)H} = 1 + R - (1+G)(1+N) = R - G - N - GN$

Intuitively, $\bar{R}$ is the rate at which the government can consume its asset returns while insuring that assets keep up with economic growth (or equivalently the rate at which the government should reimburse its debt to avoid exploding debt).[68]

B.4 Open economy

The key question is the following: in the long run, will the government choose to accumulate positive assets or debt? In the open economy case, the answer depends entirely on the level of the world rate of return $r \geq 0$ In case r is sufficiently large (as compared to the planner's rate of time preference and concavity parameters), then intertemporal social welfare is maximized by accumulating positive assets. In case r is sufficiently small, then intertemporal social welfare is maximized by accumulating debt. More precisely, it all depends upon whether r is larger or smaller than the so-called modified Golden rule rate of return $r^* = \delta + \Gamma' n + \Gamma g$. I.e. the optimal policy sequence $(\tau_{Bt}, \tau_{Lt})_{t\geq 0}$ can be characterized as follows:

[67] Here we neglect exploding asset accumulation paths ($a_t \to +\infty$ or $a_t \to -\infty$), which in effect are ruled out by the assumptions $\overline{\tau}_t \geq 0$ and $\overline{\tau}_t \leq 1$ (see below).

[68] In a continuous time model, this rate would simply be $\overline{r} = r - g - n$.

Proposition 11 ***(Open economy intertemporal optimum).*** *Define the modified Golden rule rate of return r^* as follows: $r^* = \delta + \Gamma' n + \Gamma g$. Three cases can happen:*

(1) *If $r > r^*$, then the social planner chooses to have all tax payments in the short run (current or nearby generations). I.e. the planner chooses to impose high taxes in the short run, zero taxes in the long run ($\tau_{Bt} \to 0$ and $\tau_{Lt} \to 0$) and to accumulate sufficiently large positive assets to finance public good provision: $a_t \to \overline{a} = \dfrac{\tau}{\overline{R}} > 0$*

(2) *If $r < r^*$, then the social planner chooses to postpone tax payments to the long run (distant generations). I.e. the planner chooses to impose low taxes in the short run, revenue-maximizing taxes in the long run ($\tau_{Bt} \to \overline{\tau}_B$ and $\tau_{Lt} \to \overline{\tau}_L$) and to accumulate maximal debt compatible with financing public good provision: $a_t \to a_0 = \dfrac{\tau - \overline{\tau}}{R} < 0$*

(3) *If $r = r^*$, then any positive or negative government asset position can be a social optimum (depending on the initial condition and the parameters). For any given optimum $\overline{a} \geq 0$ or $\overline{a} \leq 0$, then for infinitely small rates of time preference, the tax policy sequence $(\tau_{Lt}, \tau_{Bt})_{t \geq 0}$ maximizing intertemporal social welfare converges towards the steady-state welfare optimum. I.e. as $t \to +\infty$, $\tau_{Bt} \to \tau_B(\delta)$, and $\tau_{Lt} \to \tau_L(\delta)$, and as $\delta \to 0$, $\tau_B(\delta) \to \tau_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}}$ and $\tau_L(\delta) \to \tau_L = \dfrac{\tau - \tau_B b_y - \overline{a}/\overline{R}}{1-\alpha}$.*

Proof. Consider the case $r > r^*$, and assume that $\overline{\tau}_t = \tau_{Lt}(1-\alpha) + \tau_{Bt} b_{yt} \to \overline{\tau} > 0$. Consider a small change whereby the planner raises the tax rate at time t to $\overline{\tau}_t = \overline{\tau} + d\tau$ and reduces the tax rate at some future time $s > t$ to $\overline{\tau}_s = \overline{\tau} - d\tau'$ Budget balance requires: $e^{rH(s-t)} d\tau = e^{gh(s-t)} d\tau'$. If $r > r^*$, one can immediately see that the induce change in social welfare $dSWF$ is positive.

TO BE COMPLETED

Note that the only reason why we get finite asset accumulation is because we constrain the government to use very simple policy instruments. I.e. once the government has accumulated enough assets to finance public good provision, there is no point accumulating additional assets, since the government cannot distribute negative taxes (i.e. we assume $\overline{\tau}_t \geq 0$, and we also assume away lump sum transfers). Otherwise in case $r > r^*$, the planner would accumulate infinite assets ($a_t \to +\infty$). In effect, the economy would cease to be a small open economy any more. So if we are interested in the full social optimum, it makes more sense to look the closed economy case.

B.5 Closed economy

In the closed economy case, the domestic capital stock K_t is equal to the sum of private and government assets (i.e. $K_t = B_t + A_t$), and the generational rate of return $1 + R_t = e^{r_t H}$ is equal to the marginal product of capital:

$$R_t = F_K = \frac{\alpha}{\beta_t}$$

with: $\beta_t = \dfrac{K_t}{Y_t} = b_{yt} e^{-r_t H} + a_t =$ domestic capital-ouput ratio

One can show that the government will accumulate assets until the point where the modified Golden Rule condition is satisfied: $r_t = r^* = \delta + \Gamma' n + \Gamma g$. (or, equivalently, $1 + R_t = e^{r_t H} = 1 + R^* = e^{r^* H} = (1+\Delta)(1+N)^{\Gamma'}(1+G)^{\Gamma}$). That is, the government will accumulate assets until the point where $\beta_t = \beta^* = \dfrac{\alpha}{R^*}$. Will this involve the accumulation of positive government assets or the accumulation of public debt? Both cases can happen, depending on parameters. The important point is that the socially optimal level of capital accumulation and the market equilibrium level of capital accumulation depend on completely independent parameters, so this can really go both ways. T

On the one hand the socially optimal level β^* depends on the parameters of the social welfare function δ, Γ', Γ . A more patient planner will accumulate more capital. A more concave planner will accumulate less capital (infinite concavity = intergenerational Rawlsian = no need to leave any capital to next generations since they will be richer than us; of course one problem if we do that is that growth might itself disappear).[69]

On the other hand, the market equilibrium depends on the parameters of private preferences. If the taste for bequest and wealth is very small, then there will little capital accumulation. Conversely, if the taste for bequest and wealth is very large, then there will be a lot of capital accumulation. In a full fledged life cycle model, demographic parameters would also matter

[69] In the famous Stern (2007) vs. Nordhaus (2007a, 2007b) controversy about the proper social discount rate r^*, both parties agreed about $\delta = 0.1\%$ (Stern views this as an upper bound of the probability of earth crash; Nordhaus is unenthusiastic about what he views as an excessively low and "prescriptive" value, but does not seriously attempt to put forward ethical argument for a bigger δ) and $g = 1.3\%$ (on the basis of observed per capita growth rates in the long run), but strongly disagreed about Γ: Stern picked $\Gamma = 1$, so that $r^* = 1.4\%$, implying a very large net present value of future environmental damages and an urgent need for immediate action; Nordhaus picked $\Gamma = 3$, so that $r^* = 4.0\%$, implying a more laissez-faire attitude. As argued by Sterner and Persson (2008), a surprising feature of this debate is that from a cross-sectional redistribution perspective, $\Gamma = 1$ implies relatively low inequality aversion and government intervention (probably less than Stern would support), while $\Gamma = 3$ implies relatively large inequality aversion and government intervention (which Nordhaus would probably not support). One way to make the various positions internally consistent would be to introduce one supplementary parameter, namely the long run relative price of the environment in a two-good growth model (see Guesnerie 2004).

(Modigliani triangle formula), and one could end up with very small or very large capital accumulation, depending on the specifics. The general point is that there is really no reason in general to expect the market equilibrium to deliver more or less capital accumulation than the social optimum: it could really go both ways.[70]

But wherever it goes, the point is that this is essentially orthogonal to the issue of optimal tax policy:

Proposition 12 ***(Closed economy intertemporal optimum).*** *The optimal government policy looks as follows:*

(i) First, the optimal government asset and debt policy is chosen so as to satisfy the modified Golden rule: $r_t \to r^* = \delta + \Gamma' n + \Gamma g$. *The capital-output ratio converges towards the corresponding level:* $\beta_t = b_{yt}e^{-r_t H} + a_t \to \beta^* = \dfrac{\alpha}{R^*} = \dfrac{\alpha}{e^{r^* H} - 1}$.*This will involve government assets* ($a_t \to \overline{a} > 0$) *in case private agents tend to under-accumulate capital in this economy* ($b_y e^{-rH} < \beta^*$), *and government debt* ($a_t \to \overline{a} < 0$) *in case they tend over-accumulate assets* ($b_y e^{-rH} > \beta^*$). *Both cases can happen, depending on parameter values for private preferences and social welfare function.*

(ii) Next, for infinitely small rates of time preference, the tax policy sequence $(\tau_{Lt}, \tau_{Bt})_{t\geq 0}$ *maximizing intertemporal social welfare converges towards the steady-state welfare optimum. I.e. as* $t \to +\infty$, $\tau_{Bt} \to \tau_B(\delta)$, *and* $\tau_{Lt} \to \tau_L(\delta)$, *and as* $\delta \to 0$, $\tau_B(\delta) \to \tau_B = \dfrac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}}$ *and* $\tau_L(\delta) \to \tau_L = \dfrac{\tau - \tau_B b_y - \overline{a}/\overline{R}}{1 - \alpha}$.

Proof. The proof is essentially the same as proposition 2.

TO BE COMPLETED

[70] The central point made by Diamond (1965) was exactly this: in a general OLG model, one could very well get over-accumulation or under-accumulation of capital.

References

Aiyagari, S. Rao. 1995. "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting." *Journal of Political Economy* 103(6): 1158-75.

Atkeson, Andrew and Robert E. Lucas, "On Efficient Distribution with Private Information," *Review of Economic Studies* 59, 427-453.

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

Beckert, Jens, *Inherited Wealth,* Princeton University Press, 2008, 382p.

Blanchard Olivier and Stanley Fisher, *Lectures on Macroeconomics*, MIT Press, 1989, 650p.

Cagetti, Marco, and Mariacristina De Nardi. 2009. "Estate Taxation, Entrepreneurship, and Wealth." *American Economic Review*, 99(1): 85–111.

Carroll C., J. Overland, and D.N. Weil, "Saving and Growth with Habit Formation", *American Economic Review* 90(3), 2000, 341-355

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

Chamley, Christophe. 2001 "Capital Income Taxation, Wealth Distribution and Borrowing Constraints," *Journal of Public Economics* 79, 55–69.

Conesa, J. C., S. Kitao, and D. Krueger. 2009. "Taxing Capital? Not a Bad Idea after All!." *American Economic Review*, 99(1): 25–48.

Cremer, Helmuth, and Pierre Pestieau. 2004. "Wealth Transfer Taxation: A Survey of the Theoretical Literature." in S.C. Kolm and J. Mercier Ythier, eds, *Handbook on Altruism, Giving and Reciprocity,* North Holland, Amsterdam.

Cremer, Helmuth, Pierre Pestieau, and Jean-Charles Rochet. (2003) "Capital Income Taxation when Inherited Wealth is not Observable", *Journal of Public Economics*, 87 (2003) 2475–2490.

Diamond, Peter, "National Debt in a Neoclassical Growth Model", *American Economic Review* 55(5), 1965, 1126-1150

Diamond, Peter, "Optimal Tax Treatment of Private Contributions for Public Goods with and without Warm Glow Preferences," *Journal of Public Economics*, 90(4-5), May 2006, 897-919.

Diamond, Peter A., and James A. Mirrlees, "Optimal Taxation and Public Production I: Production Efficiency." *American Economic Review,* 61(1): 8–27.

Diamond, Peter, and Emmanuel Saez "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, Fall 2011.

Farhi, Emmanuel and Iván Werning, "Inequality and Social Discounting", *Journal of Political Economy* 2007

Farhi, Emmanuel and Iván Werning. "Progressive Estate Taxation," *Quarterly Journal of Economics* , 2010, 125(2), 635-673.

Fleurbaey, Marc, and François Maniquet, "Fair Income Tax", *Review of Economic Studies* 2006, vol.73, pp.55-83

Fleurbaey, Marc, *Fairness, Responsability and Welfare*, Oxford University Press, 2008

Guesnerie, Roger, "Calcul économique et développement durable", *Revue économique*, vol.55(5), 2004, pp.363-382

Hopehnayn, Hugo and Edward Prescott, "Stochastic Monotonicity and Stationary Distributions for Dynamic Economies", *Econometrica*, 1992, vol.60(6), pp.1387-1406

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

Judd, Kenneth L. 1999. "Optimal taxation and spending in general competitive growth models," *Journal of Public Economics*, 71: 1-26.

Kaldor, Nicholas, *An Expenditure Tax*, Unwin University Books, 1955, 248p.

King, Mervyn "Savings and Taxation", in G. Hughes and G. Heal, eds., *Public Policy and the Tax System* (London: George Allen Unwin, 1980), 1-36.

Kopczuk, Wojciech and J. Lupton, "To Leave or Not to Leave: The Distribution of Bequest Motives", *Review of Economic Studies*, 2007, 74(1), pp.207-235.

Nordhaus, William, "Critical Assumptions in the Stern Review on Climate Change", *Science*, vol.317, 2007a, pp.201-202

Nordhaus, William, "A Review of the Stern Review on the Economics of Climate Change", *Journal of Economic Literature*, vol.45, 2007b, pp.686-702

Piketty, Thomas, "The Dynamics of the Wealth Distribution and the Interest Rate with Credit-Rationing", *Review of Economic Studies*, 1997, vol.64(2), pp.173-189

Piketty, Thomas, "On the Long-Run Evolution of Inheritance: France 1820-2050", Working Paper, Paris School of Economics, 2010, 421p.

Piketty, Thomas, "On the Long-Run Evolution of Inheritance: France 1820-2050", *Quarterly Journal of Economics*, 2011, vol.126(3), pp.1071-1131

Piketty, Thomas, Gilles Postel-Vinay and Jean-Laurent Rosenthal, "Wealth Concentration in a Developing Economy: Paris and France, 1807-1994", *American Economic Review,* 2006, vol.96(1), pp.236-256

Piketty, Thomas, Gilles Postel-Vinay and Jean-Laurent Rosenthal, "Inherited vs Self-Made Wealth - Theory and Evidence from a Rentier Society (Paris 1872-1937)", *Working Paper,* 2011

Piketty, Thomas, and Emmanuel Saez, "Optimal Labor Income Taxation", 2011, in preparation *Handbook of Public Economics*, Volume 5, Elsevier.

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva, "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities", Working Paper, 2011

Pirttila, Jukka, and Hakan Selin. "Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993." *Scandinavian Journal of Economics,* 113(1), 2011, 120–144.

Roemer, John et al., "To What Extent Do Fiscal Systems Equalize Opportunities for Income Acquisition Among Citizens?", *Journal of Public Economics*, 2003, vol.87, pp.539-565

Saez, Emmanuel, "Using Elasticites to Derive Optimal Tax Rates", *Review of Economic Studies*, 2001

Saez, Emmanuel "The Desirability of Commodity Taxation under Nonlinear Income Taxation and Heterogeneous Tastes," *Journal of Public Economics*, 83, 2002, 217-230.

Saez, Emmanuel, Joel Slemrod, and Seth Giertz, "The Elasticity of Taxable Income with respect to Marginal Tax Rates: A Critical Review", *Journal of Economic Literature*, forthcoming, 2011

Scheve, Kenneth and David Stasavadge, "Democracy, War and Wealth: Lessons from Two Centuries of Inheritance Taxation", Working Paper, NYU, 2011

Stern, Nicholas, *The Economics of Climate Change: The Stern Review*, Cambridge University Press, 2007, 712p.

Sterner T., and U.M. Persson, "An Even Sterner Review: Introducing Relative Prices into the Discounting Debate", *Review of Environmental Economics & Policy*, vol.2, 2008, pp.61-76

Figure 1: Top Inheritance Tax Rates 1900-2011

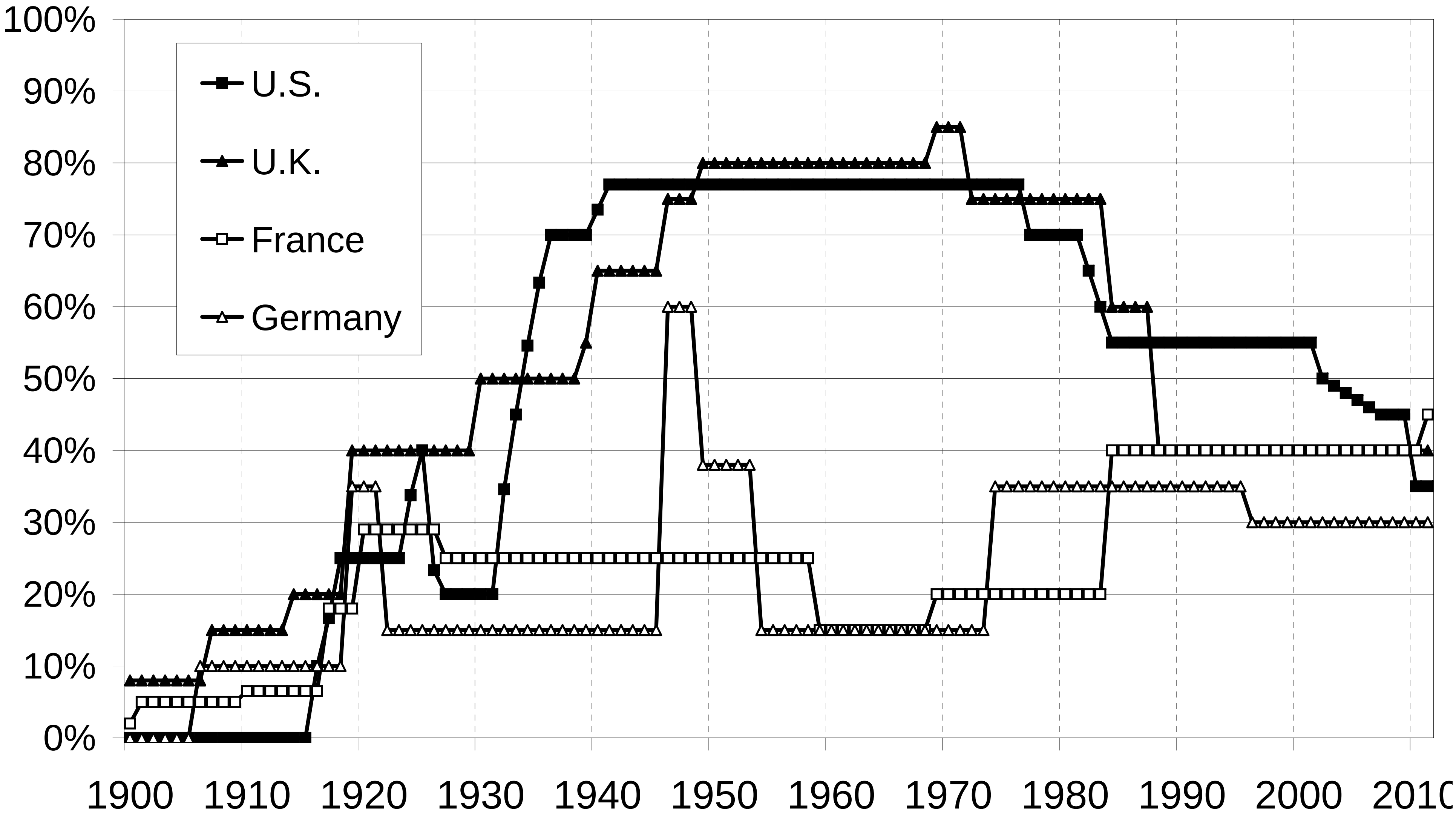


Figure 2: Top Income Tax Rates 1900-2011

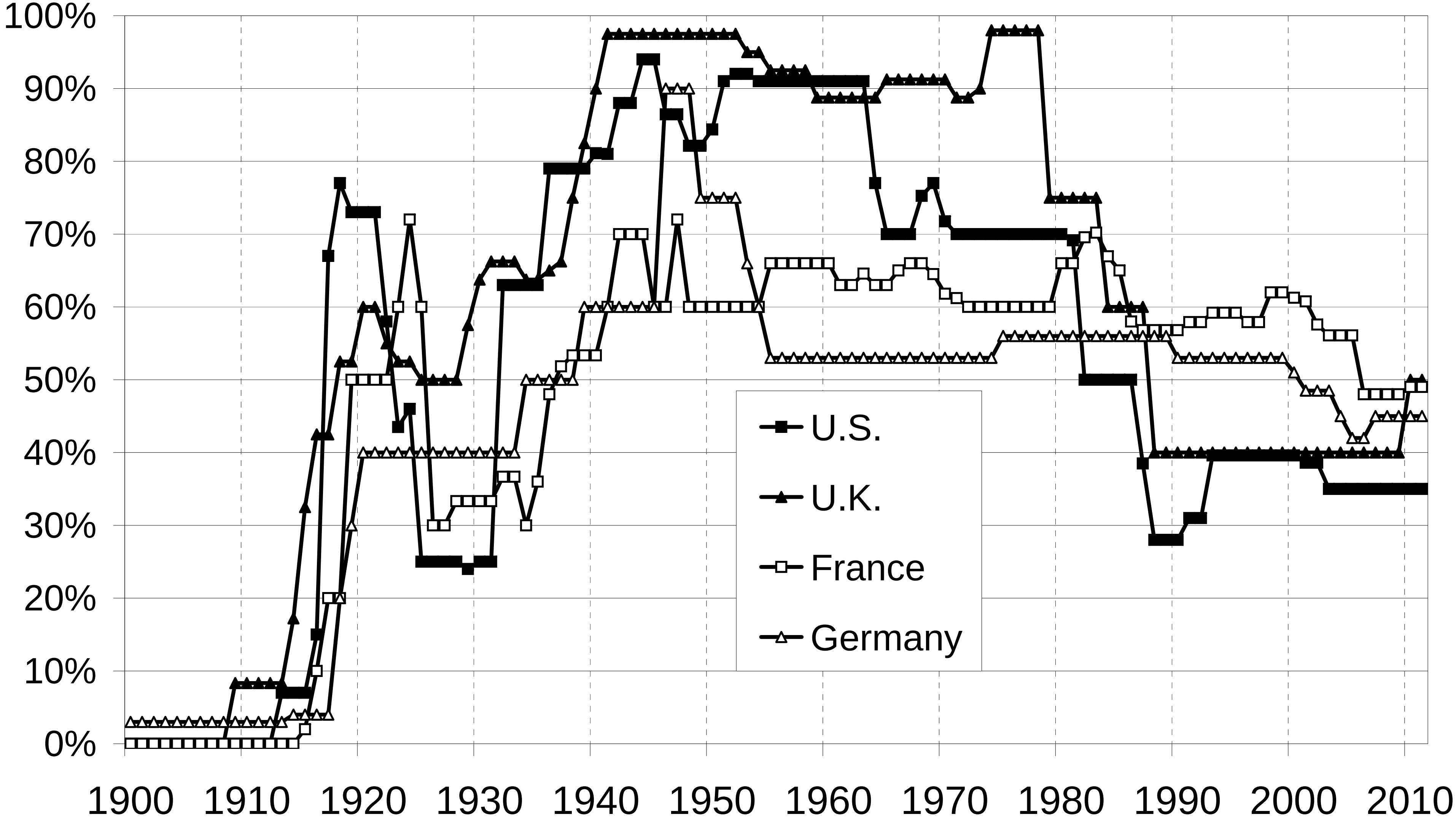


Figure 3: Top Income Tax Rates: Earned (Labor) vs Unearned (Capital)

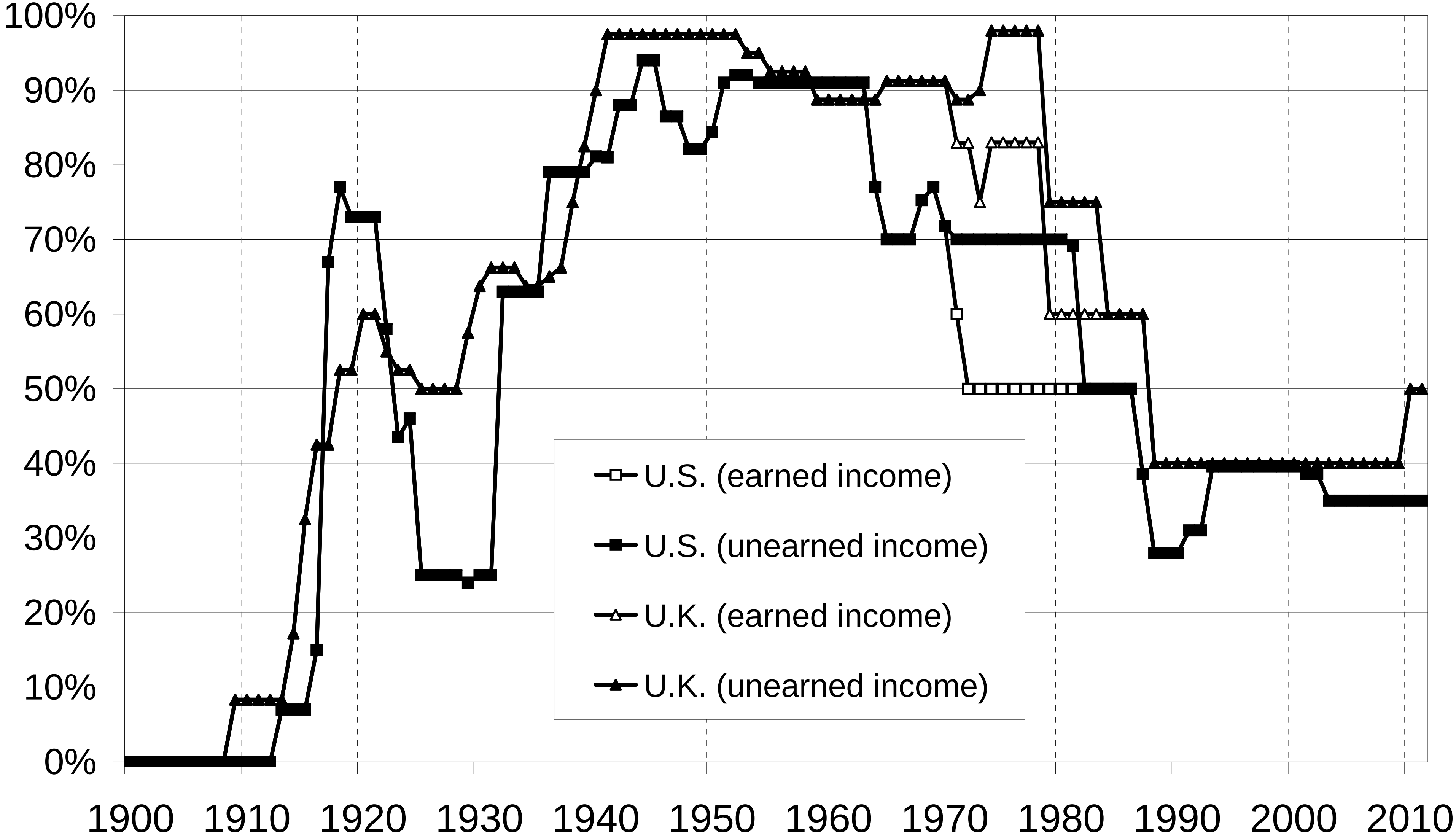
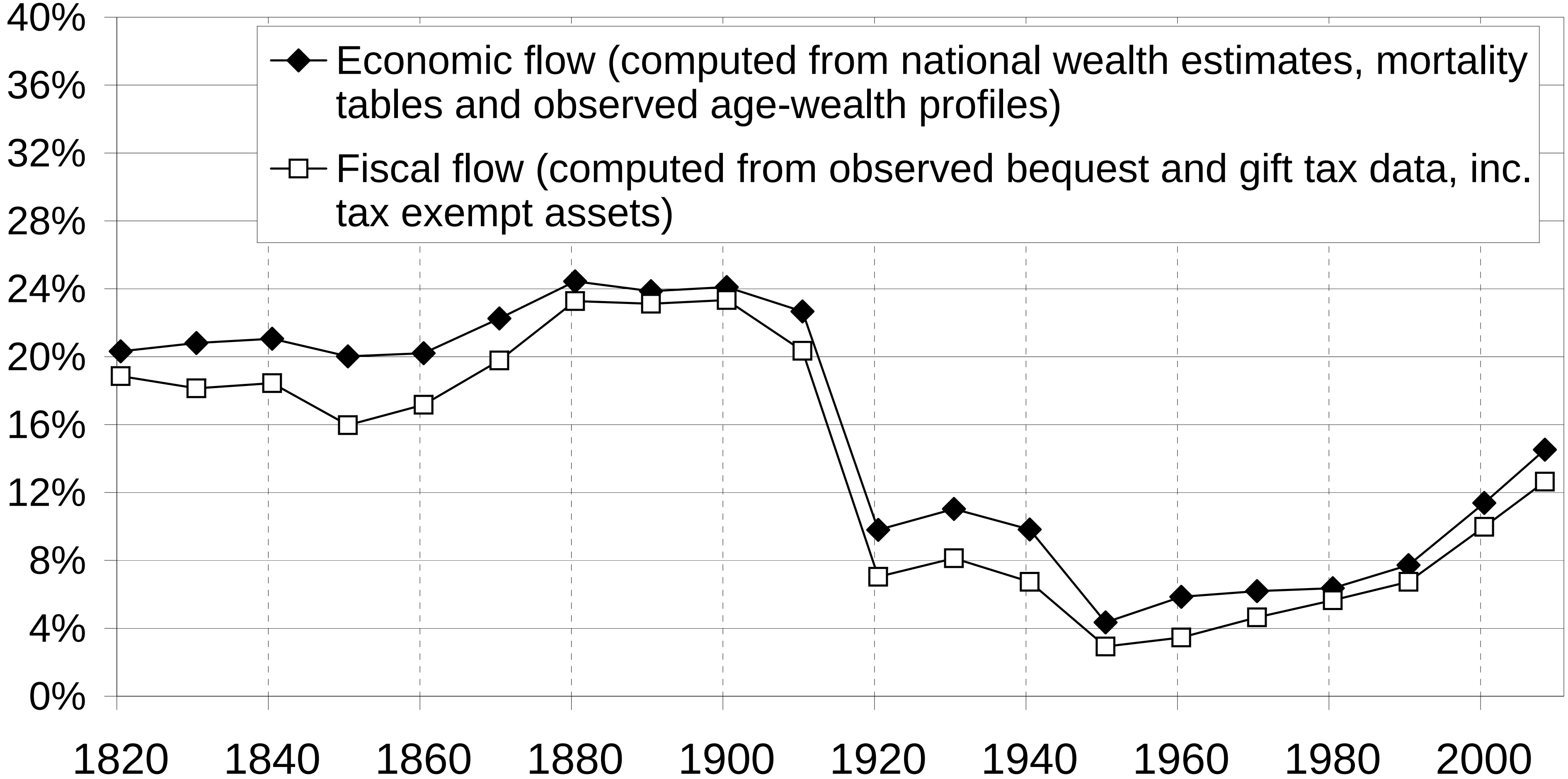
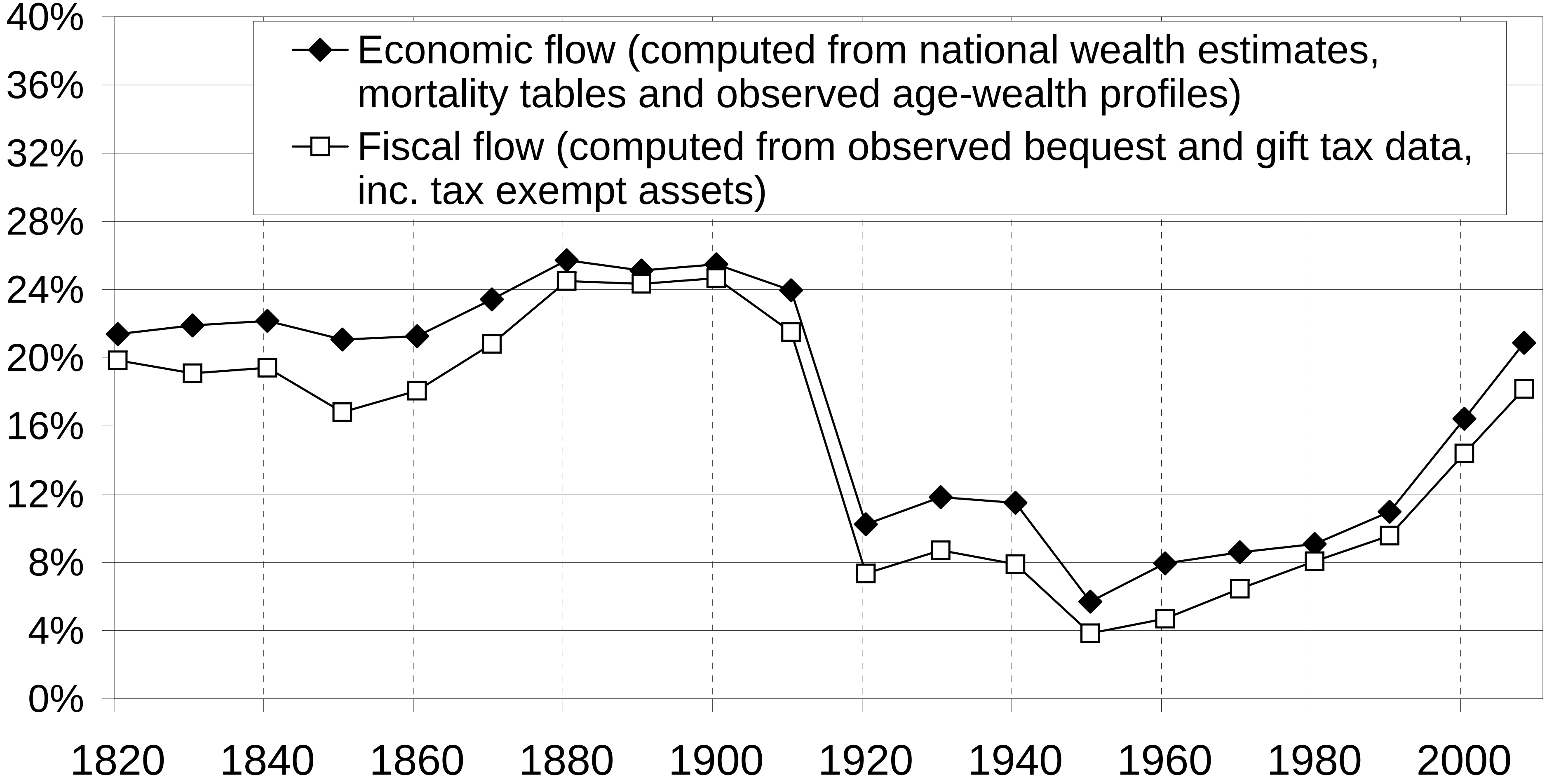


Figure 4: Annual inheritance flow as a fraction of national income, France 1820-2008



Source: T. Piketty, "On the long-run evolution of inheritance", QJE 2011

Figure 5: Annual inheritance flow as a fraction of disposable income, France 1820-2008



Source: T. Piketty, "On the long-run evolution of inheritance", QJE 2011