

A Theory of Optimal Capital Taxation*

Thomas Piketty, Paris School of Economics

Emmanuel Saez, UC Berkeley and NBER

April 4, 2012

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 random tastes for bequests to children and for wealth per se. We derive formulas for optimal tax rates on capitalized inheritance expressed in terms of estimable parameters and social preferences. The long-run optimal 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 tax rate on capitalized inheritance should be as high as 50%-60%—or even higher for top wealth holders—if the government has meritocratic preferences (i.e., puts higher welfare weights on those receiving little inheritance) and if capital is highly concentrated (as it is in the real world). 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, positive capital 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 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 are grateful to seminar participants at the Paris School of Economics, the London School of Economics, the National Bureau of Economic Research (Public Finance Group), the University of California at Berkeley, the Massachusetts Institute of Technology, and Stanford University for their comments. We thank Tony Atkinson, Alan Auerbach, Peter Diamond, Emmanuel Farhi, Mikhail Golosov, Louis Kaplow, Wojciech Kopczuk, Matt Weinzierl and Ivan Werning for helpful and stimulating discussions. We acknowledge financial support from the Center for Equitable Growth at UC Berkeley.

1 Introduction

According to the profession’s most popular theoretical models, tax rates on capital should be equal to zero in the long run—including from the viewpoint of those individuals or dynasties who own no capital at all. Taken seriously, those results imply that all inheritance taxes, property taxes, corporate profits taxes, and individual taxes on capital income should be eliminated and that the resulting tax revenue loss should be recouped with higher labor income or consumption or lump-sum taxes. Strikingly, even individuals with no capital or inheritance would benefit from such a change. E.g. we should be explaining to propertyless individuals that it is in their interest to set property taxes to zero and replace them by poll taxes.

Few economists however seem to endorse such a radical policy agenda. Presumably this reflects a lack of faith in the standard models and the zero-capital tax results - which are indeed well known to rely upon strong assumptions.¹ As a matter of fact, all advanced economies impose substantial capital taxes. For example, the European Union currently raises 9% of GDP in capital taxes (out of a total of 39% of GDP in total tax revenues) and the US raises about 8% of GDP in capital taxes (out of a total of about 27% of GDP in total tax revenues).²

However, in the absence of an alternative tractable model, the zero capital tax result remains an important reference point in economics teaching and in policy discussions.³ For instance, a number of economists and policy-makers support tax competition as a way to impose zero optimal capital taxes to reluctant governments.⁴ We view the large gap between optimal capital tax theory and practice as 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—at least for some parameter values). By realistic, we also mean a theory offering such conclusions for reasons that are consistent with the reasons that are at play in the

¹In particular, Atkinson and Stiglitz (1976; 1980, pp. 442-451) themselves have repeatedly stressed that their famous zero capital tax result relies upon unplausibly strong assumptions (most notably the absence of inheritance and the separability of preferences), and has little relevance for practical policy discussions. See also Atkinson and Sandmo (1980) and Stiglitz (1985).

²See European Commission (2011), p.282 (total taxes) and p.336 (capital taxes), for GDP-weighted EU 27 averages, and OECD (2011) for the United States.

³Lucas (1990, p.313) celebrates the zero-capital-tax result of Chamley-Judd as “the largest genuinely free lunch I have seen in 25 years in this business.”

⁴See Cai and Treisman (2005) and Edwards and Mitchell (2008) for references.

real world which—we feel—are related to the large concentration of inherited capital ownership. 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 distributional parameters and behavioral elasticities and hence be robust to changes in the underlying primitives of the model.⁵

In our view, the two key ingredients for a proper theory of capital taxation are, first, the large aggregate magnitude and the high concentration of inheritance, and, next, the imperfection of capital markets. With no inheritance (as in the Atkinson-Stiglitz model where all wealth is due to life-cycle savings or as in Chamley-Judd where life is infinite) or egalitarian inheritance (representative agent model), and with perfect capital markets (i.e. if agents can transfer resources across periods at a fixed and riskless interest rate r), then we believe that the case for no capital taxation would indeed be very strong—as in the standard Atkinson-Stiglitz or Chamley-Judd models. Hence, our paper proceeds in two steps.

First, we develop a theory of optimal capital taxation with perfect capital markets. 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 tastes for bequests and parental resources. Importantly, top labor earners and top successors are never exactly the same people, implying a non-degenerate trade-off between the taxation of labor income and the taxation of capitalized inheritance. 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 tax rates τ_B on capitalized inheritance expressed in terms of estimable parameters and social preferences. The long run optimal tax rate τ_B increases with the aggregate steady-state flow of bequests to output b_y , decreases with the elasticity of bequests with respect to the net-of-tax rate e_B , and decreases with the strength of preferences

⁵Such an approach has yielded fruitful results in the analysis of optimal labor income taxation (see Piketty and Saez, 2012 for a recent survey).

for bequests s_{b_0} . For realistic parameters, the optimal linear tax rate on capitalized inheritance 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—half of the population receives close to zero bequest, our results are robust to reasonable changes in the social welfare objective. For example, the optimal tax policy from the viewpoint of those receiving zero bequest is very close to the welfare optimum for bottom 50% bequest receivers. Interestingly, the optimal tax rate τ_B imposed on top wealth holders 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 top bequest tax rates observed in most advanced economies during the past 100 years, especially in Anglo-Saxon countries from the 1930s to the 1980s (see Figure 1). To our knowledge, this is the first time that a model of optimal inheritance taxation delivers tractable and estimable formulas that can be used to analyze such real world tax policies.

Our model also illustrates the 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 magnitude of aggregate bequest flows. When bequest flows are small, (e.g., 5% of national income, as was the case in Continental Europe during the 1950s-1970s), then bequest taxes should be moderate. When they are large (e.g., 15% of national income as in France currently or over 20% as in the 19th century France), then bequest taxes should be large—so as to reduce the tax burden falling on labor earners.⁶

Second, we show that if we introduce capital market imperfections and uninsurable idiosyncratic shocks to rates of return into our setting, then we can study the optimal tax mix between one-off inheritance taxation and lifetime capital taxation. With perfect and riskless capital markets, bequest taxes and capital income taxes are equivalent in our framework. However, with heterogeneous rates of returns, capital income taxation can provide insurance against return risk more powerfully than inheritance taxation. If the uninsurable uncertainty about future returns is large, and the moral hazard responses of the rate of return to capital income tax rates are moderate, the resulting optimal lifetime capital tax rate τ_K can be very high—typically higher than the optimal bequest tax rate $\tilde{\tau}_B$, and labor tax rate τ_L . This can account for the fact that

⁶The historical evolution and theoretical determinants of the aggregate bequest flow b_y were recently studied by Piketty (2010, 2011). Figures 4-5 summarize his results. We extend his model to study normative issues and optimal tax policy.

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). To our knowledge this is the first time that a model of optimal capital taxation can explain why these various policy tools can indeed be complementary.

The paper is organized as follows. Section 2 relates our results to the existing literature. Section 3 presents our dynamic model and its steady-state properties. Section 4 presents our basic formula for the optimal tax rate on capitalized inheritance. Section 5 introduces informational and capital market imperfections to analyze the optimal mix between inheritance taxation and lifetime capital taxation. Section 6 extends our results in a number of directions, including elastic labor supply, homogenous tastes, consumption tax, closed economy, life-cycle saving, population growth, dynamic efficiency, and tax competition. Section 7 offers some concluding comments. Most proofs and complete details about extensions 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: Atkinson-Stiglitz and Chamley-Judd. We discuss each in turn and then discuss the more recent literature.

Atkinson-Stiglitz. Atkinson and Stiglitz (1976) show that there is no need to supplement the optimal non-linear labor income tax with a capital income tax in a life-cycle 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.⁷ 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 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. To see this, consider a model with inelastic and uniform

⁷Saez (2002) shows that this result extends to heterogeneous preferences 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. Golosov et al. (2011) calibrate a model where higher skills individuals have higher saving taste and show that the resulting optimal capital income tax rate depends significantly on the inter-temporal elasticity of substitution but that the implied welfare gains are relatively small in all cases.

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 but taxing inheritances is desirable for redistribution. This important point has been made by Cremer, Pestieau, and Rochet (2003) in a stylized partial equilibrium model with unobservable inherited wealth where taxing capital income becomes desirable. Our model allows the government to directly observe (and hence tax) inherited wealth.

Farhi and Werning (2010) consider a model from the perspective of the first generation of donors who do not start with any inheritance (so that inheritance and labor income inequality are perfectly correlated). In this context, bequests should actually be subsidized as they should be untaxed by Atkinson-Stiglitz (ignoring inheritors) and hence should be subsidized when taking into account inheritors.⁸ As we shall see, this result is not robust, in the following sense. In our model, where people both receive and leave bequests, bequest subsidies can also be socially optimal, but this will arise only for specific—and unrealistic—parameters (e.g. if there is very little inequality of inheritance or social welfare weights are concentrated on large inheritors). For plausible parameter values, however, optimal bequest rates will be positive and large.

Chamley-Judd. Chamley (1986) and Judd (1985) 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 criterion 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 heterogeneous across the population, this does not seem like a valid social welfare objective as children of parents with no tastes for bequests would not be counted in the social welfare function. We will adopt instead a definition of social welfare based on long-run equilibrium steady-state utility.⁹ We show in appendix C how the within generation and across generation

⁸Kaplow (2001) made similar points informally. Farhi and Werning (2010) also extend their model to many periods and connect their results to the new dynamic public finance literature (see below).

⁹In models with dynamic 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 criterion. Importantly, Farhi and Werning (2007) show that considering instead the long-run steady-state equilibrium as we do in this paper eliminates the immiseration results.

redistribution problems can be disconnected using public debt so that there is essentially no loss 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 standard infinite horizon dynastic model with constant discount rate.¹⁰ The textbook model predicts enormous responses of aggregate capital accumulation to changes in capital tax rates, which just do not seem to be there in historical data. Capital-output ratios are relatively stable in the long run, in spite of large variations in tax rates (see e.g. Piketty, 2010, p.52). Our theory leaves this key elasticity as a free parameter to be estimated empirically. Our model naturally recovers the zero capital tax result of Chamley-Judd when the elasticity is infinite.

New Dynamic Public Finance. The recent and fast growing literature on new dynamic public finance shows that dynamic labor productivity risk leads to non-zero capital income taxes (see Golosov, Tsyvinski, Werning, 2006 and Kocherlakota 2010 for recent comprehensive surveys). The underlying logic is the following. When leisure is a normal good, more savings, *ceteris paribus*, will tend to reduce work later on. Thus, discouraging savings through capital income taxation enhances the ability to provide insurance against future poor labor market possibilities. Quantitatively however, the welfare gains from distorting savings optimally are very small in general equilibrium (Farhi and Werning, 2011).¹¹ Our model does not include future earnings uncertainty because individuals care only about the bequests they leave, independently of the labor income ability of their children. This simplification is justified in the case of bequest decisions as empirical analysis shows that bequests respond only very weakly to children earnings opportunities (see e.g., Wilhelm, 1996). In contrast to the new dynamic public finance, we find quantitatively large welfare gains from capital taxation in our model. Hence, our contribution is independent and complementary to the new dynamic public finance.

Methodologically, the new dynamic public finance solves for the fully optimal mechanism and hence obtains optimal tax systems that can be complex and history dependent, in contrast to actual practice. We instead limit ourselves to very simple (and more realistic) tax structures. This allows us to consider richer heterogeneity in preferences which we believe is important in

¹⁰This follows from the fact that the net-of-tax rate of return needs to be equal to the (modified) discount rate in steady-state.

¹¹Golosov, Troshkin, and Tsyvinski (2011) also calibrate such a model and show that the size of the optimal implicit capital income tax wedge is quantitatively fairly modest on average (Figure 2, p. 25).

the case of bequests.¹² Therefore, we also view our methodological approach as complementary to this literature (Diamond and Saez, 2011 for a longer discussion of this methodological debate).

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 desirable 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 savings¹³ but do not attempt to compute numerical values for optimal capital tax rates. Farhi and Werning (2011) (cited above) also propose a quantitative calibration of an infinite horizon model with borrowing constraints but they find small welfare gains from capital taxation. In contrast, Conesa, Kitao, and Krueger (2009) calibrate an optimal tax OLG life-cycle 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. While this is an interesting mechanism, we do not believe that this is the most important explanation for $\tau_K > 0$. There are other more direct ways to address the issue of taxing the young vs. the old (e.g. age-varying income taxes; some policies, e.g. pension schemes, do depend on age).¹⁴ In contrast, the theory of capital taxation offered in the present paper is centered upon the interaction between inheritance and capital market imperfections.¹⁵

Government time-inconsistency and lack of commitment. Yet another way to explain real-world, positive capital taxes is to assume time inconsistency and lack of commitment.¹⁶ Zero capital tax results are always long run results. In the short run, capital is on the table, and it is always tempting for short-sighted governments to have $\tau_K > 0$, even though the optimal long run τ_K is equal to 0%. More generally, if governments cannot commit to long run policies,

¹²As mentioned above, Farhi and Werning (2010) do combine inheritance with new dynamic public finance. They consider more general tax structures than we do but impose more structure on preferences.

¹³This 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 in some cases.

¹⁴On age-dependent taxes, see Weinzierl (2011).

¹⁵Cagetti and DeNardi (2009) provide very interesting simulations of estate taxation in a model with borrowing constraints and show that shifting part of the labor tax to the estate tax benefits low income workers. They do not try however to derive optimal tax formulas as we do here.

¹⁶See e.g. Farhi, Sleet, Werning and Yeltekin (2011) for a recent model along these lines.

they will always be tempted to renege on their past commitments and to implement high capital tax rates, even though this is detrimental to long run welfare.

We doubt that this is the main reason explaining why we observe positive capital taxes in the real world. Governments and public opinions seem to view positive and substantial inheritance tax rates (such as those implemented over the past 100 years in advanced economies, see Figure 1 above) as part of a fair and efficient permanent tax system—not as a consequence of short-sightedness and lack of commitment. Naturally political actors are not always long-sighted but they often find ways to commit to long run policies, e.g. by appealing to moral principles—such as equal opportunity and meritocratic values—that apply to all generations and not only to the current electorate, or by writing down their favored policies in party platforms. Governments could also find ways to implement the zero-tax long run optimum by delegating capital tax decisions to an independent authority with a zero-tax mandate (in the same way as the zero-inflation mandate of independent central banks), or by promoting international tax competition and bank secrecy laws. In models where positive capital taxes arise solely because of lack of commitment, such institutional arrangements would indeed be optimal.¹⁷

In contrast, we choose in this paper to assume away time inconsistency issues. Hence, we analyze solely the true long run optimal tax policies—assuming full commitment—and we take up the most difficult task of explaining positive capital tax rates in such environments.

3 The Model

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 $0, 1, \dots, t, \dots$. Each generation has measure one, lives one period (which can be interpreted as H -year-long, where $H =$ generation length, realistically around 30 years), then dies and is replaced by the next generation. Total population is stationary and equal to $N_t = 1$, so aggregate variables Y_t, K_t, L_t, B_t , and per capita variables y_t, k_t, l_t, b_t , are identical (we use the latter).

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

¹⁷In the real world, believers in zero capital tax policies do support tax competition for this very reason. See e.g. Edwards and Mitchell (2008). We return to the issue of tax competition in conclusion.

for a “generational” rate of return $1 + R = e^{rH}$. 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. This condensed timing greatly simplifies the notations and exposition of the model but is unnecessary for our results.¹⁸

Individual i in generation t maximizes utility:

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

With: $\tilde{y}_{ti} = (1 - \tau_B)b_{ti}e^{rH} + (1 - \tau_L)y_{Lti}$ = total after-tax lifetime income combining after-tax capitalized bequest $(1 - \tau_B)b_{ti}e^{rH}$ and after-tax labor income $(1 - \tau_L)y_{Lti}$

$$b_{ti}e^{rH} = b_{ti}(1 + R) = \text{capitalized bequest received} = \text{raw bequest } b_{ti} + \text{return } Rb_{ti}$$

$$c_{ti} = \text{consumption}$$

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

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

$$\tau_B \geq 0 \text{ is the tax rate on capitalized bequest, } \tau_L \geq 0 \text{ is the tax rate on labor income}$$

V_{ti} is the utility function assumed to be homogeneous of degree one to allow for balanced growth (and possibly heterogeneous across individuals).

In order to fix ideas, consider the special Cobb-Douglas (or log-log) 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. The parameters s_{wi} and s_{bi} measure the tastes for wealth per se and for bequest (more on this below).

In the general case with $V_i(c, w, \bar{b})$ homogeneous of degree one, the fraction s_i of lifetime resources saved depends on $(1 - \tau_B)e^{rH}$, i.e., the relative price of bequests. Using the first order condition of the individual $V_{ic} = V_{iw} + (1 - \tau_B)e^{rH}V_{i\bar{b}}$, we can then define $s_{bi} = s_i \cdot (1 - \tau_B)e^{rH}V_{i\bar{b}}/V_{ic}$ and $s_{wi} = s_i \cdot V_{iw}/V_{ic}$. Hence, s_i , s_{wi} , and s_{bi} are functions of $(1 - \tau_B)e^{rH}$ instead of being constant as with Cobb-Douglas where income and substitution effects cancel out.

¹⁸All results and optimal tax formulas can be extended to a full-fledged, multi-period, continuous-time model with overlapping generations and life-cycle savings. See section 6 below.

We use a standard wealth accumulation model with exogenous growth. Per capita 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 per capita physical (non-human) capital input and l_t is the per capita human capital input (efficient labor supply). Though this is unnecessary for our results, we assume a Cobb-Douglas production function: $y_t = k_t^\alpha l_t^{1-\alpha}$ to simplify the notations .

Per capita human capital l_t 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: $l_t = h_t = h_0 e^{gHt}$.

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 $\beta = \frac{k_t}{y_t} = \frac{\alpha}{R}$ = domestic generational capital-output ratio).¹⁹ It is important to keep in mind that y_t is domestic output. In the open economy case we consider, y_t might differ from national income if the domestic capital stock k_t (used for domestic production) differs from the national wealth b_t .

It follows that output $y_t = \beta^{\frac{\alpha}{1-\alpha}} l_t = \beta^{\frac{\alpha}{1-\alpha}} h_0 e^{gHt}$ also grows at rate $1 + G = e^{gH}$ per generation. So does aggregate labor income $y_{Lt} = (1 - \alpha)y_t$. The aggregate economy is on a steady-state growth path where everything grows at rate $1 + G = e^{gH}$ per generation.

E.g. 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:

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

In our model, 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} , earned different labor income y_{Lti} , or had different tastes for savings $s_{ti} = s_{wti} + s_{bti}$.²⁰

Savings Tastes. Importantly, taste parameters vary across individuals and over time in our model. E.g. some individuals might have zero taste for wealth and bequest ($s_{wti} = s_{bti} = 0$), in

¹⁹The annual capital-output ratio is $\beta_a = H \cdot \beta = \alpha(H/R) = \alpha H / (e^{rH} - 1) \simeq \alpha/r$ if r is small.

²⁰A fourth important factor in the real world is the existence of idiosyncratic shocks to rates of return r_{ti} (see section 5). Pure demographic shocks (such as shocks to the age at parenthood, age at death of parents and children, number of children, rank of birth, etc.) also play an important role.

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 ($s_{wti} > 0, s_{bti} = 0$) (“wealth-lovers”), while others might have no direct taste for wealth but taste for bequest ($s_{wti} = 0, s_{bti} > 0$) (“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.

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 and over generations, just like other tastes.²¹ We allow for any exogenous distribution for taste parameters $g(s_{wi}, s_{bi})$. For notational simplicity, we assume that tastes are drawn i.i.d. at each generation from the distribution $g(s_{wi}, s_{bi})$. Hence they are independent across individuals within a generation and independent across generations within a dynasty. In the Cobb-Douglas case, the parameters s_{wi}, s_{bi} are fixed independently of τ_B . In the general homogeneous of degree one case, the parameters s_{wi}, s_{bi} depend upon $(1 - \tau_B)e^{rH}$ and hence are not strictly parameters. We adopt this slight abuse of notation for presentational simplicity.²²

Assumption 1 *Taste parameters (s_{wi}, s_{bi}) are drawn i.i.d. at each generation from an exogenous distribution $g(s_{wi}, s_{bi})$ defined over a set of possible tastes $S \subset \bar{S}$ (where \bar{S} is the set of all possible tastes: $\bar{S} = \{(s_{wi}, s_{bi}) \text{ s.t. } s_{wi}, s_{bi} \geq 0 \text{ and } s_i = s_{wi} + s_{bi} \leq 1\}$).*

S and $g(\cdot)$ can be discrete or continuous. We denote by $s_0 = \min \{s_i = s_{wi} + s_{bi} \in S\}$, $s_1 = \max \{s_i = s_{wi} + s_{bi} \in S\}$, with $0 \leq s_0 \leq s_1 \leq 1$, and $s = E(s_i)$ the average taste.

We assume that S includes zero saving tastes and at least one other taste: $s_0 = 0, s_1 > 0$.

Assumption 1 implies that in each generation there are “zero bequest receivers” (i.e. individuals who receive zero bequest, because their parents had zero taste for wealth and bequest).²³

Productivity Shocks. Labor productivity shocks are specified as follows. Individual i in generation t has a within-cohort normalized productivity parameter $\theta_{ti} = h_{ti}/h_t$. By definition, we have: $y_{Lti} = \theta_{ti}y_{Lt}$ (with $E(\theta_{ti}) = 1$). Productivity differentials θ_{ti} could come from innate

²¹Kopczuk and Lupton (2007) and Kopczuk (2009, 2012) present evidence on heterogeneity in bequest motives.

²²Rigorously, we would need to parametrize utility functions so that $s_{bi} = s_b(\sigma_{bi}, (1 - \tau_B)e^{rH}), s_{wi} = s_w(\sigma_{wi}, (1 - \tau_B)e^{rH})$ with $(\sigma_{bi}, \sigma_{wi})$ i.i.d parameters and $s_b(\cdot)$ and $s_w(\cdot)$ fixed functions.

²³This could result from other types of shocks (see example below).

abilities, acquired skills, individual occupational choices, or sheer luck—and most likely from a complex combination between the four. We assume that productivity shocks are drawn i.i.d. from the same distribution $h(\theta_i)$ at each generation and independently of savings tastes.

Assumption 2 *Productivity parameters θ_i are drawn i.i.d. at each generation from an exogenous distribution $h(\theta_i)$ over some productivity set $\Theta \subset [0, +\infty[$ independently of savings tastes. The set Θ and the distribution $h(\cdot)$ can be discrete or continuous. We note: $\theta_0 = \min \{\theta_i \in \Theta\}$ and $\theta_1 = \max \{\theta_i \in \Theta\}$, with $0 \leq \theta_0 \leq 1 \leq \theta_1 \leq +\infty$. By construction: $E(\theta_i) = 1$.*

All our results can readily be extended to a setting with some intergenerational persistence of savings tastes and productivities. In that case, to ensure the existence of a unique ergodic steady-state joint distribution of inherited wealth and productivities, one would simply need to assume that the random process for tastes satisfies a simple ergodicity property. Any individual has a positive probability of having any savings taste×productivity no matter what his or her parental savings taste×productivity were (see appendix A1).

Steady State Distributions. Under assumptions 1-2, the individual transition equation (1) can be aggregated into:

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

Let us denote the aggregate *capitalized* bequest flow-domestic output ratio by $b_{yt} = \frac{e^{rH}b_t}{y_t}$. 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} :

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

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

Assumption 3 $s \cdot e^{(r-g)H} < 1$

If assumption 3 is violated, the economy can accumulate infinite wealth relative to domestic output, and will cease to be a small open economy at some point so that the world rate of return will have to fall to restore assumption 3. If assumption 3 is satisfied, then, as $\tau_B \geq 0$, $b_{yt} \rightarrow b_y = \frac{s(1 - \tau_L)(1 - \alpha)e^{(r-g)H}}{1 - s(1 - \tau_B)e^{(r-g)H}}$ as $t \rightarrow +\infty$. I.e. the aggregate inheritance-output ratio converges towards a finite value, and in steady-state, bequests grow at the same rate as output.

Finally, we denote by $z_{ti} = b_{ti}/b_t$ the within-cohort normalized bequest, 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 $g(\cdot)$ and $h(\cdot)$ and the individual transition equation (1) entirely determine the law of motion for the distribution of inheritance $\phi_t(z)$ and the joint distribution of inheritance and labor productivity, which we denote by $\psi_t(z, \theta) = \phi_t(z) \cdot h(\theta)$.

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

(b) *We have: $b_y = \frac{s(1 - \tau_L)(1 - \alpha)e^{(r-g)H}}{1 - s(1 - \tau_B)e^{(r-g)H}}$.*

(c) *The joint inheritance-productivity distribution $\psi(z, \theta) = \phi(z) \cdot h(\theta)$ is two-dimensional. At any productivity level, the distribution involves zero-bequest receivers and is non-degenerate. I.e. $z_0 = \min \{z \text{ s.t. } \phi(z) > 0\} = 0 < z_1 = \max \{z \text{ s.t. } \phi(z) > 0\} \leq \infty$*

Proof. The result follows from standard ergodic convergence theorems (Appendix A1). **QED**

Two points are worth noting. First, the aggregate magnitude of inheritance flows relative to output b_y grows with $r - g$. 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. The multiplicative factor associated to intergenerational wealth transmission is large and leads to high inheritance flows. 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. As shown in Piketty (2011), 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 19th and early 20th centuries (20%-25% of national income in 1820-1910), so low in the mid-20th century (less than 5% around 1950-1960), and why they are becoming large again in the late 20th and early 21st centuries (about 15% in 2010 in France) (see Figures 4-5). 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.²⁴ Available evidence suggests that the

²⁴E.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 $b_y/(1 - \tau) = 19\%$. The b_y formula given above relates to the generational, (capitalized bequest)/output ratio, while the empirical estimates depicted on Figures 4-5 refer to the cross-sectional, non-capitalized ratio. But one can show that both ratios are very

French pattern also applies to Continental European countries that were hit by similar growth and capital shocks. The long-run U-shaped pattern of aggregate inheritance flows was possibly somewhat less pronounced in the United States or United Kingdom (Piketty, 2010, 2011).

Second, one important feature of our model—and of the real world—is that inequality is two-dimensional. In steady-state, the relative positions in the distributions of inheritance and labor productivity are never perfectly correlated. This is the key property that we need for our optimal tax problem to make sense and for our results to hold: Labor income is not a perfect predictor for inheritance. With i.i.d. taste and productivity shocks, we even get that the two distributions are independent ($\psi(z, \theta) = \phi(z) \cdot h(\theta)$). All our results would still hold if we introduce some intergenerational persistence of tastes and productivities, as long as persistence is not complete and the two dimensions of shocks are not perfectly correlated. As we shall see below, this two-dimensionality property is the key feature explaining why the Atkinson-Stiglitz result does not hold in our model, and why we need a two-dimensional tax policy tool (τ_B, τ_L).

3.3 An Example with Binomial Random Tastes

A simple example might be useful in order to better understand the logic of two-dimensional inequality and the role played by random tastes in our model. Assume that taste shocks take only two values: $s_i = s_0 = 0$ with probability $1 - p$, and $s_i = s_1 > 0$ with probability p . The aggregate saving rate is equal to $s = E(s_i) = ps_1$. Let $\mu = s(1 - \tau_B)e^{(r-g)H}$, $\mu_1 = s_1(1 - \tau_B)e^{(r-g)H} = \mu/p$. Assume $\mu < 1 < \mu/p$, and no productivity heterogeneity: $\Theta = \{1\}$. One can easily show that the steady-state inheritance distribution $\phi(z)$ is discrete and looks as follows:

$z = z_0 = 0$ with probability $1 - p$ (children with zero-wealth-taste parents).

$z = z_1 = \frac{1 - \mu}{p} > 0$ with probability $(1 - p) \cdot p$ (children with wealth-loving parents but zero-wealth-taste grand-parents).

...

$z = z_k = \frac{1 - \mu}{p} + \frac{\mu}{p} \cdot z_{k-1} = \frac{1 - \mu}{\mu - p} \cdot \left[\left(\frac{\mu}{p} \right)^k - 1 \right]$ with probability $(1 - p) \cdot p^{k+1}$ (children with wealth-loving ancestors during the past $k + 1$ generations, but zero-wealth-taste $k + 2$ -ancestors).

That is, the steady-state distribution $\phi(z)$ is unbounded above and has the standard Pareto asymptotic upper tail found in empirical data and in wealth accumulation models with random

close when inheritance tends to happen around mid-life (see section 6 below). Piketty (2010, 2011) presents detailed simulations using a full-fledged, out-of-steady-state version of this model, with life-cycle savings and full demographic and macroeconomic shocks.

multiplicative shocks (see Appendix A1 and Atkinson, Piketty and Saez (2011)). Inheritances are obviously uncorrelated with labor income (since there is no inequality of labor income).

Taste shocks could also be interpreted as shocks to rates of return (e.g., p is the probability that one gets a high return, and $1 - p$ is the probability that one goes bankrupt, thereby leaving zero estate) or as a demographic shocks (e.g., p is the probability that one dies at a “normal age” and with “normal” health costs, and $1 - p$ is the probability that one dies very old or after large health costs, thereby leaving zero estate; shocks on number of children or rank of birth could also do). As long as the shocks have a multiplicative structure, the steady-state distribution of inheritance will have a Pareto upper tail, with a Pareto coefficient reflecting the relative importance of the various effects (see Appendix A1). In practice all these types of shocks clearly exist and matter a lot. The key point is that there are many factors - other than productivity shocks - explaining the large inequality of inherited wealth that we observe in the real world. The main limitation of models of wealth accumulation based solely upon productivity shocks is that they massively under-predict wealth concentration.²⁵

If we introduce productivity shocks (say $\theta_{ti} = \theta_0 \geq 0$ with probability $1 - q$ and $\theta_{ti} = \theta_1 > 1 > \theta_0$ with probability q), the steady-state joint distribution $\psi(z, \theta)$ is simply the product of the two distributions, i.e. $\psi(z, \theta) = \phi(z) \cdot h(\theta)$. So the joint distribution again involves zero correlation between the two dimensions. If we further introduce some intergenerational persistence in the productivity process (say, $\theta_{t+1i} = \theta_1$ with probability q_0 if $\theta_{ti} = \theta_0$, and with probability $q_1 \geq q_0$ if $\theta_{ti} = \theta_1$), then the steady-state distribution $\psi(z, \theta)$ will involve some positive correlation between the two dimensions. But the correlation will always be less than one: the entire history of ancestors’ tastes s_{ti}, s_{t-1i} , etc. and productivity shocks $\theta_{ti}, \theta_{t-1i}$, etc. matters for the determination of the current inheritance position z_{t+1i} , while only parental productivity θ_{ti} matters for the current productivity position θ_{t+1i} .²⁶

3.4 The Optimal Tax Problem

We now formally define our optimal tax problem. We assume that the government faces an exogenous revenue requirement: per capita public good spending must satisfy $g_t = \tau y_t$ where $\tau \geq 0$ is taken as given and y_t is exogenous per capita domestic output. We first assume

²⁵See discussion on homogeneous tastes in Section 6 below and references given in Piketty (2011, section II.C).

²⁶Our results can also be extended to a model without random tastes, as long as productivity shocks include a zero lower bound (see Section 6).

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 capitalized inheritance at rate $\tau_B \geq 0$. We impose a period-by-period (i.e. generation-by-generation) budget constraint: the government must raise from labor income y_{Lt} and capitalized inheritance $b_t e^{rH}$ received by generation t an amount sufficient to cover government spending τy_t for generation t .²⁷ We again assume that everything takes place at the end of period: output is realized, taxes are paid, government spending and private consumption occur. Hence, the period t government budget constraint looks as follows:

$$\tau_L y_{Lt} + \tau_B b_t e^{rH} = \tau y_t \quad \text{i.e.} \quad \tau_L(1 - \alpha) + \tau_B b_y = \tau \quad (4)$$

We assume that $\tau < 1 - \alpha$, i.e. the public good spending requirement is not too large and could be covered by a labor tax alone (in case the government so wishes).

Assumption 4 $\tau < 1 - \alpha$

It is worth stressing that all taxes are paid at the end of the period, and that the tax τ_B is a tax on capitalized bequest $b_t e^{rH} = b_t(1 + R)$, not a tax on raw bequest b_t . One natural interpretation of this tax on capitalized bequest is that at the end of the period the government taxes both raw bequests b_t and capital income (returns to bequest) Rb_t at the same rate τ_B . So the tax τ_B should really be viewed as a broad based “capital tax” (falling on wealth transmission as well as as on the returns to wealth) rather than a narrow based bequest tax. Note that as long as capital markets are perfect and everybody gets the same rate of return (we relax this assumption in section 5 below), it really does not matter how the government chooses to split the capital tax burden between one-off inheritance taxation and lifetime capital taxation on the flow return. In particular, rather than taxing bequests b_t and the returns to bequest Rb_t at the same rate τ_B , it would also be equivalent not to tax bequest b_t and instead to have a larger, single tax on the returns to capital Rb_t at rate τ_K such that:²⁸

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

Example. Assume $r = 4\%$, $H = 30$, so that $e^{rH} = 1 + R = 3.32$, i.e. $R = 2.32$.

²⁷We introduce intergenerational redistribution in Section 6 (appendix C provides complete details).

²⁸Here 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 $\bar{b}_{t+1i} = [1 - \tilde{\tau}_B + (1 - \tau_K)R]b_{t+1i}$. If V_{ti} were defined over after-tax *non-capitalized* bequest $\bar{b}_{t+1i} = (1 - \tilde{\tau}_B)b_{t+1i}$, 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.

If $\tau_B = 20\%$ then $\tau_K = 29\%$. If $\tau_B = 40\%$ then $\tau_K = 57\%$. If $\tau_B = 60\%$ then $\tau_K = 86\%$.

Hence, it is equivalent to tax capitalized bequests at $\tau_B = 40\%$ or to tax capital income flows at $\tau_K = 57\%$ (or $\tau_K = 43\%$ if we take the equivalent instantaneous tax rate).²⁹ More generally, any intermediate combination will do. I.e. for any tax mix $(\tilde{\tau}_B, \tau_K)$, $\tilde{\tau}_B$ is a tax on raw bequest and τ_K is an extra tax on the return to bequest, one can define $\tau_B = \tilde{\tau}_B + \tau_K \frac{R}{1+R}$.³⁰ Intuitively, τ_B is the adjusted total tax rate on capitalized bequest. For now, we focus on the broad capital tax interpretation ($\tau_B = \tilde{\tau}_B$, i.e. no extra tax on return: $\tau_K = 0$). In section 5 we introduce capital market imperfections to analyze the optimal tax mix between $\tilde{\tau}_B$ and τ_K .

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 1-4, 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:³¹

$$SWF = \iint_{z \geq 0, \theta \geq 0} \omega_{p_z p_\theta} \frac{V_{z\theta}^{1-\Gamma}}{1-\Gamma} dz d\theta \quad (5)$$

With: $V_{z\theta} = E(V_i | z_i = z, \theta_i = \theta)$ = average steady-state utility level V_i attained by individuals i with normalized inheritance $z_i = z$ and productivity $\theta_i = \theta$.

$\omega_{p_z p_\theta}$ = social welfare weights as a function of the percentile ranks p_z, p_θ in the steady-state distribution of normalized inheritance z and productivity θ .³²

Γ = concavity of the social welfare function ($\Gamma \geq 0$).³³

²⁹In 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 annual capital income tax rate during the H -year period, in which case the equivalence equation would be: $1 - \tau_B = e^{-\tau_K r H}$, 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 annual $\tau_K = 19\%, 43\%, 76\%$ (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).

³⁰The tax on raw bequest $\tilde{\tau}_B b_t$ is paid at the end of the period, and the tax payment is assumed to be $\tilde{\tau}_B b_t (1+R)$, so in effect τ_K can be interpreted as an extra tax on the return to bequest.

³¹This steady-state maximization problem can also be formulated as the asymptotic solution of an intertemporal social welfare maximization problem. See Appendix C, Proposition C1.

³²Here we implicitly assume that the welfare weights ω_i are the same for all individuals i with the same ranks p_z, p_θ in the distribution of normalized inheritance and productivity. Our optimal tax formulas can easily be extended to the general case where social welfare weights ω_i also depend upon taste parameters s_{wi} and s_{bi} - which can be justified for utility normalization purposes. See the discussion in Appendix A2.

³³If $\Gamma = 1$, then $SWF = \iint_{z \geq 0, \theta \geq 0} \omega_{p_z p_\theta} \log(V_{z\theta}) d\Psi(z, \theta)$.

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)):

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

In general, one might expect $e_B > 0$: with a higher net-of-tax rate $1 - \tau_B$, agents may choose to devote a larger fraction of their resources to inheritance, in which case the aggregate, steady-state inheritance ratio will be 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.

Substituting $\tau_L(1 - \alpha) = \tau - \tau_B b_y$ into the steady-state formula for b_y , we obtain:

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

Recall that s does not depend on τ_B in the Cobb-Douglas case with i.i.d shocks. Therefore, b_y depends on τ but not on the tax mix τ_L, τ_B and $e_B = 0$ in that case. For general utility functions and/or random processes, s depends on τ_B and e_B could really take any value (> 0 or < 0). We view e_B as a free parameter to be estimated empirically. There is no reason to expect e_B to be infinitely large, unlike in the infinite-horizon dynastic model of Chamley-Judd.

4 Basic Optimal Capital Tax Formula

4.1 The Zero-Bequest-Receiver Social Optimum

Throughout this paper we are particularly interested in the zero-bequest-receiver 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. This corresponds to the case with a linear social welfare function ($\Gamma = 0$) and the following welfare weights: $\omega_{p_z p_\theta} = 1$ if $p_z = 0$ (i.e. $z = 0$) and $\omega_{p_z p_\theta} = 0$ if $p_z > 0$. Since the $V_i()$ are homogenous of degree one, $\Gamma = 0$ implies that the government does not want to redistribute income from high productivity to low productivity individuals—perhaps because individuals are viewed as (partly) responsible for their productivity parameter θ . In contrast, individuals cannot be responsible for their bequest parameter z . Therefore trying to reduce as much as possible the inequality of lifetime welfare opportunities along the inheritance dimension seems normatively appealing.³⁴ So we start by characterizing

³⁴Perhaps surprisingly, the normative literature on 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)

this zero-bequest-receiver optimum, which we call the “meritocratic Rawlsian optimum”:

Proposition 2 (zero-bequest-receiver optimum). *Under assumptions 1-4, linear social welfare ($\Gamma = 0$), and the welfare weights: $\omega_{p_z p_\theta} = 1$ if $p_z = 0$, and $\omega_{p_z p_\theta} = 0$ if $p_z > 0$, 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 $s_{b0} = E(s_{bi} | z_i = 0) =$ average bequest taste of zero bequest receivers (weighted by marginal utility \times labor income).

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:

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

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

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 $0 \leq s_i \leq 1$). Recall that $s_{bi} = s_i \cdot (1 - \tau_B)(1 + R)V_{\bar{b}_i}/V_{ci}$.³⁵

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:

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

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

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

focus on labor income taxation. See however the interesting discussion in Fleurbaey (2008, pp.146-148).

³⁵In 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 $1 + R$.

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:

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

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

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

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

Note 1. This proof works with any utility function that is homogenous of degree one (and not only in the Cobb-Douglas case) and with any ergodic random process for taste and productivity shocks (and not only with i.i.d. shocks). In the case with Cobb-Douglas utility functions, the proof can be further simplified. See Appendix A2.

Note 2. In the general case, s_{b0} is the average of bequest tastes s_{bi} over all zero-bequest-receivers, weighted by the product of their marginal utility V_{ci} and of their productivity θ_i . In case $s_{bi} \perp V_{ci} \theta_i$, then s_{b0} is the simple average of s_{bi} over all zero-bequest-receivers: $s_{b0} = E(s_{bi} \mid z_i = 0)$. In the case with i.i.d. shocks and adequate utility normalization, then s_{b0} is the same as the average bequest taste for the entire population: $s_{b0} = s_b = E(s_{bi})$. See Appendix A2.

Note 3. We also show in the appendix how to extend the optimal tax formula to the case $\Gamma > 0$. One simply needs to replace s_{b0} by: $s_{b0} = \frac{E(V_{ci} \theta_i V_i^{-\Gamma} s_{bi} \mid z_i = 0)}{E(V_{ci} \theta_i V_i^{-\Gamma} \mid z_i = 0)}$. 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).

When the social welfare function is infinitely concave ($\Gamma \rightarrow +\infty$), in effect the planner puts infinite weight on the least productive, zero-bequest receivers. This is equivalent to assuming welfare weights $\omega_{p_z p_\theta} = 1$ iff $p_z = 0$ and $p_\theta = 0$. Therefore s_{b0} is simply the average bequest taste within this group: $s_{b0} = E(s_{bi} \mid z_i = 0, \theta_i = \theta_0)$. This could be called the “radical Rawlsian optimum”. This might be too radical, however, because individuals are - partly - responsible for their productivity, e.g. through their choice of occupation. From an ethical perspective, the most appealing social welfare optimum probably lies in between the meritocratic and the radical Rawlsian optima, depending on how much one considers individuals are responsible for their productivity (i.e. how much productivity parameters reflect individual choices rather than innate abilities or sheer luck) - an issue which we do not model explicitly in the present paper.³⁶

³⁶See Piketty and Saez (2012) for a more elaborate normative discussion.

Note 4. Using formula (7) for b_y , we also have $\tau_B = \frac{1 + s_{b0} - (s_{b0}/s)e^{-(r-g)H}}{1 + s_{b0} + e_B}$. This alternative formula shows more directly how the optimal rate varies with primitives $s, s_{b0}, r - g$ but is more difficult to calibrate than our formula in Proposition 2 (since we typically have data on b_y).

4.2 Numerical calibrations

The optimal tax formula $\tau_B = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}}$ is simple, intuitive, and can easily be calibrated using empirical estimates.

The optimal tax rate τ_B decreases with the elasticity of bequests to the net-of-tax rate e_B , increases with the aggregate steady-state flow of inheritances to output b_y , and decreases with the strength of preferences for leaving bequests s_{b0} . A higher bequest elasticity e_B unsurprisingly implies a lower τ_B . As $e_B \rightarrow +\infty$, $\tau_B \rightarrow 0\%$. I.e. one should never tax an infinitely elastic tax base as in the dynastic model of Chamley-Judd.

More interestingly, a higher bequest flow ratio b_y implies a higher τ_B . This is a very large effect, as the example below illustrates.

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 improve the welfare of those receiving no inheritance.

In our theory there is really no general reason why capitalized inheritance should be taxed more or less than labor income. Any situation can be optimal, depending on parameters. With the low b_y ratios observed in the 1950s-1960s, it is probably optimal to tax inheritance less than labor. But with the high b_y ratios observed in the 1900s-1910s or the 2000s-2010s, it is probably

optimal to tax inheritance more than labor (see Figures 4-5).

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.

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 (non-infinite) values, then one can naturally obtain fairly large levels for socially optimal capital tax rates. That is, 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. With a realistic value $e_B = 0.2$, we find $\tau_B = 56\%$ and $\tau_L = 31\%$.³⁷ In practice, this bequest elasticity effect e_B is also mitigated by the existence of a positive labor supply elasticity effect e_L , which makes low labor taxation and therefore high bequest taxation even more desirable (see section 6).

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 $\tau_B \rightarrow 100\%$ as $e_B \rightarrow 0$. 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. Intuitively, if b_y is sufficiently small (e.g. if g is sufficiently large), then the benefits of taxing bequests - in terms of tax revenue - become smaller than the utility costs (as measured by s_{b0}), so that even those who receive no bequest do not want to tax bequests. For plausible parameter values, however, the optimal bequest tax rate τ_B from the viewpoint of zero receivers is positive (we discuss bequest subsidies in detail in Appendix A2).

³⁷We 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_B is relatively small (at most $e_B = 0.1 - 0.2$). Using U.S. time and cross-section variations, Kopczuk and Slemrod (2001) also find elasticities e_B around $0.1 - 0.2$.

4.3 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, depending upon the distributions of shocks). However because real world inheritance is highly concentrated (half of the population receives negligible bequests), our optimal 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 to compute the optimal tax rate from the viewpoint of those individuals belonging to the percentile p_z of the distribution of inheritance:

Proposition 3 (p_z -bequest-receiver optimum). *Under assumptions 1-4, linear social welfare ($\Gamma = 0$), and the following welfare weights: $\omega_{p_z p_\theta} = 1$ for a given $p_z \geq 0$, and $\omega_{p'_z p_\theta} = 0$ if $p'_z \neq p_z$ ($z =$ normalized inheritance of p_z -receivers), then:*

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

with $s_{bz} = E(s_{bi} | p_{z_i} = p_z) =$ average bequest taste of p_z -receivers, $\theta_z = E(\theta_i | p_{z_i} = p_z) =$ average productivity of p_z -receivers (weighted by marginal utility \times labor income), (with i.i.d shocks $\theta_z = 1$).

(b) *There exists $p_{z^*} \geq 0$ (i.e. $z^* > 0$) such that $\tau_B > 0$ iff $p_z < p_{z^*}$ (i.e. $z < z^*$).*

The cut-off z^ is below average inheritance: $z^* < 1$. That is, average-bequest receivers prefer bequest subsidies.*

In case $\phi(z)$ is fully egalitarian, then $p_{z^} \rightarrow 0$: nobody wants bequest taxation.*

In case $\phi(z)$ is infinitely concentrated, then $p_{z^} \rightarrow 1$: everybody wants bequest taxation.*

Proof and notes. The proof is essentially the same as for Proposition 2 - and works again with any utility function that is homogenous of degree one and any ergodic random process for shocks. With i.i.d. productivity shocks, then $\theta_z = 1$. The formula can again be extended to the case $\Gamma > 0$, and to any combination of welfare weights ($\omega_{p_z p_\theta}$): one simply needs to replace s_{bz} , z and θ_z by the properly weighted averages \bar{s}_b , \bar{z} , and $\bar{\theta}$. In case $\Gamma \rightarrow +\infty$, then for any combination of positive welfare weights ($\omega_{p_z p_\theta}$) (in particular for uniform utilitarian weights: $\omega_{p_z p_\theta} = 1$ for all p_z, p_θ), we have: $\bar{s}_b \rightarrow s_{b0} = E(s_{bi} | z_i = 0, \theta_i = \theta_0)$ and $\bar{z}/\bar{\theta} \rightarrow 0$, i.e. we are back to the radical Rawlsian optimum. See Appendix A3. **QED.**

Unsurprisingly, the optimal tax rate τ_B is a decreasing function of z . I.e. individuals who receive higher inheritance prefer lower bequest taxes. People above percentile p_{z^*} (i.e. above

normalized inheritance 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. Interestingly, $z^* < 1$, i.e. agents with average bequest prefer bequest subsidies (if $z = 1$, then $\tau_B < 0$).³⁸ The intuition is the following. In terms of after-tax total resources, agents receiving average bequest have nothing to gain by (linearly) taxing successors from their own cohort. So since taxing bequests reduces the utility from leaving wealth to the next generation, there is really no point having a positive τ_B .

This also implies that there is no room for bequest taxation in the representative-agent version of this model. I.e. with uniform tastes and productivities and a fully egalitarian inheritance distribution $\phi(z)$, the tax optimum always involves a bequest subsidy $\tau_B < 0$ (financed by a labor tax $\tau_L > 0$), so as to induce agents to internalize the joy-of-giving externality (as in Kaplow, 2001). With full wealth equality, there is no point in taxing bequests in our model. Conversely, with infinite wealth inequality (almost everybody has zero wealth, and a vanishingly small fraction has all of it), then $p_{z^*} \rightarrow 1$: almost everybody wants the same bequest tax rate as zero receivers. More generally, for a given social welfare objective, the more unequal the distribution of inherited wealth, the higher the optimal tax rate. E.g. if one cares only about the welfare of the median successor ($p_z = 0.5$), then the optimal tax rate is higher if the median-to-average inheritance ratio z is lower.

The exact cut-off values z^* and p_{z^*} depend not only on the inequality of the inheritance distribution $\phi(z)$, but also on the aggregate level of inheritance b_y (for a given degree of inequality, a higher b_y implies a higher τ_B , in the same way as for zero receivers), as well as on the correlation between z and θ_z . That is, if the ranks z and θ_z in the inheritance and productivity distributions are almost perfectly correlated, then there little point taxing bequests: this brings limited additional redistributive power than labor taxes, and extra disutility costs. The point, however, is that real-world inherited wealth is a lot more concentrated than labor income.

One simple—yet plausible—way to calibrate the formula is the following. Assume that we are trying to maximize the average welfare of bottom 50% bequest receivers ($p_z \leq 0.5$). In every country for which we have data, the bottom 50% share in aggregate inherited wealth is typically

³⁸Strictly speaking, if $z \geq \theta_z$ (e.g. if $z = 1$ and $\theta_z = 1$), then τ_B is no longer well defined (the government would want an infinite subsidy to bequest to generate more “free utility”, see discussion below), unless one constraints τ_L to be less than one.

about 5% or less (see Piketty, 2011, p.1076), which means that their average z is about 10%. The average labor productivity θ_z within this group 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, given that z/θ_z is very small anyway, this θ_z effect has a limited impact on optimal tax rates. I.e. in the benchmark case with $b_y = 15\%$, $e_B = 0.2$, $z = 10\%$, the optimal bequest tax rate is equal to $\tau_B = 49\%$ with $\theta_z = 70\%$, vs. $\tau_B = 46\%$ with $\theta_z = 50\%$, (vs. $\tau_B = 56\%$ if $z = 0\%$). That is, inheritance is so concentrated that bottom 50% bequest receivers and zero bequest receivers have welfare maximizing bequest tax rates which are in any case 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 percentile below p_{z^*} should prefer a positive bequest tax. In actual fact, the distribution is so skewed that less than 20% of the population has inherited wealth above average (i.e. the true p_{z^*} is typically above 0.8).³⁹ But to the extent that many more people believe to be above average, either in terms of received or left bequest, 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 (p_z -bequest-receiver optimum). Under assumptions 1-4, linear social welfare ($\Gamma = 0$), and the following welfare weights: $\omega_{p_z p_\theta} = 1$ for a given $p_z \geq 0$, and $\omega_{p'_z p_\theta} = 0$ if $p'_z \neq p_z$, then:

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

³⁹We leave to future research a detailed calibration using cross-country data. Here we refer to rough estimates using the French data sources on inheritance presented in Piketty (2010, 2011).

with $x_z = E(z_{t+1i}|z_{ti} = z) =$ average normalized bequest left by p_z -receivers

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

$z =$ normalized inheritance of p_z -receivers.

(b) If $x_z \rightarrow 0$ as $z \rightarrow 0$, then $\tau_B \rightarrow 1/(1 + e_B)$ as $z \rightarrow 0$ (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 A3. **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 mostly on aggregate data) and the bequest taste s_{bz} (a preference parameter that is relatively difficult to estimate).⁴⁰ The alternative formula is based almost entirely on distributional parameters which in principle can be estimated empirically - but require comprehensive microeconomic data (such as wealth data spanning over two generations).⁴¹ 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

⁴⁰Due 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).

⁴¹High 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). One can also set ν_z equal to one in order to get lower bounds for the optimal tax rate.

children of bottom 50% successors receive as little as what their parents received (relative to the average), then the optimal bequest tax rate is $\tau_B = 77\%$ 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 $\tau_B = 45\%$. Presumably the real world is in between, say around $x_z = 50\%$, in which case $\tau_B = 61\%$.

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 $\tau_L = 34\%$.

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. As discussed in the literature (e.g., Cremer and Pestieau, 2004, Diamond, 2006 and Kaplow, 2001 and 2008), 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 $z \geq \theta_z$, 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. Our formulas can help focusing the public debate and future empirical research upon the most important parameters.

Lumpsum Demogrants. Our basic model has ruled out the use of demogrants. If we assume that the inheritance tax funds a demogrant (and that τ_L is fixed), we obtain exactly the same formulas as in Propositions 2-3 and Corollary 1 with the only difference that θ_z has to be replaced by one (because the increase $d\tau_B$ funds an equal additional demogrant to all instead of a labor tax cut proportional to θ_i).

4.4 Nonlinear Bequest 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 lower as in France where a significant fraction of estates are taxable (typically 10%-20%).⁴² 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 government budget constraint becomes

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

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 $1 - \tau_B$

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

where \bar{e}_B is the average elasticity (weighted by bequest size) of individual bequests b_{ti} above b_t^* . Empirical studies can in principle estimate \bar{e}_B and a is directly observable from tabulated statistics by estate size (typically $a \simeq 1.5$ for empirical estate distributions).

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-receiver optimum*). *Under adapted assumptions 1-4, and the following welfare weights: $\omega_{p_z p_\theta} = 1$ if $p_z = 0$, and $\omega_{p_z p_\theta} = 0$ if $p_z > 0$, then:*

$$\tau_B = \frac{1 - (1 - \alpha - \tau) s_{b_0}^* / b_y^*}{1 + e_B^* + s_{b_0}^*} \quad \text{and} \quad \tau_L = \frac{\tau - \tau_B b_y^*}{1 - \alpha},$$

with $s_{b_0}^* = E[(s_{bi} / s_i)(b_{t+1i} - b_{t+1}^*)^+ | z_i = 0] / E(\tilde{y}_{ti} | z_i = 0) = \text{strength and likelihood that non-receivers will leave taxable bequests (weighted by marginal utility} \times \text{labor income)}$.

⁴²In 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: bottom 50% successors always receive barely 5% of aggregate inheritance (while the top 10% receives over 60% in Europe and over 70% in the U.S.), so there is little point taxing them. See e.g. Piketty (2011, p.1076).

Proof. The proof is similar to Proposition 2 and can be easily extended to the case of p_z -bequests-receivers. See Appendix A4. **QED.**

Four points are worth noting. First, if zero-receivers never accumulate a bequest large enough to be taxable, then $s_{b_0}^* = 0$, and the formula reverts to the revenue maximizing tax rate $\tau_B = 1/(1 + e_B^*) = 1/(1 + a \cdot \bar{e}_B)$.⁴³ More generally, if zero-receivers have a very small probability to leave a taxable bequest (say, if b^* is sufficiently large), then $s_{b_0}^*$ is close to 0, and τ_B is close the revenue maximizing tax rate. This can be easily generalized to small p_z -receivers (say, bottom 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, $\tau_B = 70\% - 80\%$).

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 the world wars capital shocks.⁴⁴

Second, as b^* grows, there are two options: either $s_{b_0}^*/b_y^*$ 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 $s_{b_0}^*/b_y^*$ declines with b^* as the fraction of self-made wealth likely declines with the size of wealth accumulated. If the elasticity \bar{e}_B 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 \bar{e}_B might actually grow with b^* .⁴⁵

⁴³The formula takes the same form as in standard optimal labor income tax theory (see Saez 2001).

⁴⁴The 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 in the German public debate was the need to favor reconstruction and new capital accumulation. See e.g. Beckert (2008). In contrast, according to the "war mobilization" theory (see Scheve and Stasavadge 2011), inheritance taxes should have increased at least as much in Germany and France as in the UK and the US.

⁴⁵This is easily seen in the model with binomial random tastes.

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 than $\tau_B = 0$ and $b^* = 0$. The trade-off is the following: increasing b^* reduces the tax base b_y^* and hence estate tax revenue so this is a negative. The positive is that it reduces $s_{b_0}^*$ (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 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 further exploration of full non linear optima to future research.

5 Inheritance Taxation vs. Lifetime Capital Taxation

So far we have focused upon optimal taxation of capitalized inheritance and derived optimal tax formulas that can justify relatively large tax rates when the aggregate inheritance flow is large. 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 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. As mentioned in introduction, actual tax revenues from capital taxes are even slightly higher, around 8-9% in the European Union and the United States. However only a small part comes from inheritance taxes—generally less than 1% of GDP as bequest tax rates are usually relatively small, except for very top (taxable) estates. Most revenue comes from "lifetime capital taxes", falling either on the capital stock (annual property and wealth taxes, typically about 1-2% of GDP) or on the capital income flow (annual taxes on corporate profits, rental income, interest, dividend and capital gains, typically about 4%-5% of GDP).⁴⁶ Why do we observe small inheritance taxes and large lifetime capital taxes? Our basic model cannot tackle this question, since all forms of capital taxes are equivalent (Section 3).

⁴⁶The simulations presented by Piketty (2011) also show that lifetime capital taxes have had a much larger historical impact than bequest taxes on the magnitude and evolution of aggregate inheritance flows.

Clearly the conclusion would be different in a full-fledged, multi-period model with life-cycle savings.⁴⁷ Positive capital income taxes $\tau_K > 0$ would then impose additional distortions on inter-temporal consumption decisions within a given lifetime. 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 $\tau_B > 0$. Naturally, if the inter-temporal elasticity of substitution is fairly small, then this extra distortion would also be small, and both tax policies would be relatively close to one another. In the real world however we do observe a 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. What can account for this?

In this section, we explore two 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 mix. We certainly do not pretend that these are the only important factors. For example, individuals may be subject to various forms of tax illusion whereby smaller annual capital taxes are less visible than one big bequest tax per generation.⁴⁸ Other forms of capital market imperfections, such as borrowing constraints, might also play an important role. For example, large inheritances taxes may force successors to quickly and inefficiently sell their property.⁴⁹

5.1 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. Any gap between the labor income tax rate τ_L and the capital income tax rate τ_K may induce tax avoidance. E.g., self-employed individuals can largely 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). There is extensive empirical evidence that income shifting is a significant issue, and accounts for a large fraction of observed

⁴⁷See Section 6 below for such an extension.

⁴⁸This could contribute to explain why 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.

⁴⁹Anecdotal evidence suggests that this is an important reason why people dislike inheritance taxes (see Graetz and Shapiro, 2005).

behavioral responses to tax changes.⁵⁰ 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, distinguishing the returns to capital services from the returns to labor services would be easily feasible.

For simplicity, we assume “full fuzziness”. Individuals can shift their labor income flows into capital income flows (and conversely) at no cost. Hence, both income flows are undistinguishable for the tax administration, and tax rates have to be the same: $\tau_L = \tau_K = \tau_Y$, where $\tau_Y \geq 0$ is the comprehensive income tax rate. Under this assumption, our basic optimal tax formula (Proposition 2) can be easily extended, and the new fiscal optimum is such that:

Proposition 5 (comprehensive income tax cum inheritance tax). *Under the full-fuzziness assumption, the zero-bequest-receivers optimum has a bequest tax $\tilde{\tau}_B$ and a comprehensive income tax $\tau_L = \tau_K = \tau_Y$ such that:*

$$\tilde{\tau}_B = \tau_B - \tau_K \frac{R}{1+R} \quad \text{and} \quad \tau_L = \tau_K = \tau_Y = \frac{\tau - \tau_B b_y}{1 - \alpha}, \quad \text{with} \quad \tau_B = \frac{1 - (1 - \alpha - \tau) s_{b0} / b_y}{1 + e_B + s_{b0}}.$$

Proof. The proof is the same as Proposition 2. The new government budget constraint is $\tau_L(1 - \alpha) + \tilde{\tau}_B b_y + \tau_K b_y \frac{R}{1+R} = \tau$. Define $\tau_B = \tilde{\tau}_B + \tau_K \frac{R}{1+R}$ the adjusted tax rate on capitalized bequest (including the tax on bequest and the extra tax on the return to bequest; see section 3) so that $\tau_L = \frac{\tau - \tau_B b_y}{1 - \alpha}$. We obtain the same formula for τ_B as in Proposition 2. The formula for $\tilde{\tau}_B$ then follows directly from the tax enforcement constraint $\tau_K = \tau_L$. **QED**

The optimal tax combines a comprehensive income tax and an inheritance tax, as in the standard Haig-Simons-Vickrey ideal tax system.⁵¹ Most importantly, our simple optimal tax formulas allow us to quantify the trade-offs involved with this combination.

Example 6. Assume $\tau = 30\%$, $\alpha = 30\%$, $s_{b0} = 10\%$, $e_B = 0$, and $r = 4\%$, $H = 30$, so that $e^{rH} = 1 + R = 3.32$

If $b_y = 20\%$, then $\tau_B = 73\%$, so that $\tau_L = \tau_K = \tau_Y = 22\%$ and $\tilde{\tau}_B = 58\%$

If $b_y = 15\%$, then $\tau_B = 67\%$, so that $\tau_L = \tau_K = \tau_Y = 29\%$ and $\tilde{\tau}_B = 47\%$

If $b_y = 10\%$, then $\tau_B = 55\%$, so that $\tau_L = \tau_K = \tau_Y = 35\%$ and $\tilde{\tau}_B = 31\%$

If $b_y = 5\%$, then $\tau_B = 18\%$, so that $\tau_L = \tau_K = \tau_Y = 42\%$ and $\tilde{\tau}_B = -11\%$

⁵⁰See the recent survey by Saez, Slemrod, and Giertz (2012) for US evidence and Pirtila and Selin (2011) for an analysis of the dual income tax system introduced in Finland in 1993.

⁵¹In our basic model, we tax all cumulated resources at the end of the lifetime. One way to implement this is to tax cumulated average income and inheritance flows, as advocated by Vickrey (1947) in his classic reformulation of the Haig-Simons comprehensive income tax proposal.

For large bequest flows $b_y \simeq 10 - 20\%$, a comprehensive income tax system only reduces slightly the need for inheritance taxation. In contrast, for bequest flows $b_y \simeq 5\%$, the reduction can be very large. 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.

In practice, only a fraction of the population can easily shift capital into labor income (and conversely). 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.⁵²

5.2 Uninsurable Idiosyncratic Shocks to Rates of Return

Let us assume away the fuzzy-frontier problem and consider 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} . The problem of a bequest tax is that it depends only on b_{ti} , not on the idiosyncratic variations in $e^{r_{ti}H}$. So it makes more sense to charge part of the tax burden via bequest taxation $\tilde{\tau}_B$, 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. In practice there is also a difference in timing. At the time of setting the bequest tax rate τ_B , the future rate of return $e^{r_{ti}H}$ on a given asset over one generation is unknown. Rates of return are notoriously difficult to predict, and they vary enormously over assets and across individuals. In that context, it is preferable to impose a moderate bequest tax at time of receipt combined with an annual capital income tax on the returns.⁵³

Formally, let us 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.

⁵²Alternatively 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. 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.

⁵³E.g. take someone who inherited a Paris apartment worth 100,000€ (in today euros) in 1972 when 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 is more efficient to charge a moderate bequest tax in 1972, and then tax the asset continuously between 1972 and 2012, via property and/or rental income taxes.

If 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, implying that taxes have a role to play in order to reduce uncertainty.⁵⁴

If R_{ti} depends in part on unobservable individual effort (such as looking for new investment opportunities, monitoring one's financial intermediaries, etc.), then taxing returns can potentially discourage effort and hence reduce rates of return. We present such a formal model in appendix A5 using a simple reduced form cost of individual effort. In that model, we derive optimal tax rates on capital τ_K and bequests $\tilde{\tau}_B$ as a function of our previous parameters and the elasticity e_R of aggregate return R with respect to the net-of-tax rate $1 - \tau_K$ that captures the moral hazard effect of capital income taxation on returns. Optimal tax rates have two key properties. First, if e_R is sufficiently small then $\tau_K > \tau_L$. Second, if e_R is large enough, then τ_K is zero and $\tilde{\tau}_B$ is given by our standard formula.

In the appendix we also provide examples with numerical values. These simulations rely on simplifying assumptions, and are only illustrative and exploratory. In particular, we know very little about the elasticity e_R of the aggregate rate of return R . 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 large variations in taxes—would tend to suggest relatively low elasticities e_R (perhaps around $0.1 - 0.2$).⁵⁵ This would 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 we obtain $\tau_K = 78\%$ and $\tau_L = 35\%$. However the 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 not only reduces the capital income tax base but also has a negative impact on the aggregate steady-state bequest flow b_y .⁵⁶

⁵⁴Gordon (1985) quantifies this argument in the context of the corporate tax and argues that the efficiency gains associated with the reduction in uncertainty offsets the losses due to the reduction in average return.

⁵⁵Conceivably, higher individual effort e_{ti} translates into higher individual return R_{ti} mostly at the expense of others (e.g., traders obtaining advance information about when to sell a given financial asset), 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 an optimal tax model based upon pure rent-seeking elasticities, see Piketty, Saez and Stantcheva (2011).

⁵⁶In addition, these simulations do not take into account the distortionary impact of τ_K on inter-temporal

Interestingly, countries which implemented high 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 meritocratic social preferences.

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 capital income. It has now become more common to have $\tau_K < \tau_L$, via special tax exemptions for various categories of capital income. But we feel that this mostly reflects a rising concern for international tax competition and tax evasion and the persistent lack of tax coordination,⁵⁷ rather than considerations about the global welfare optimum.

6 Extensions

In this section, we consider various extensions of the basic model. Those extensions are summarized here and presented in detail in appendices B and C.

Elastic Labor Supply. We can introduce elastic labor supply along the balanced growth path by considering utility functions of the form $U_i = V_i e^{-h_i(l)}$ or equivalently $U_i = \log V_i - h_i(l)$. In that case, the small budget neutral reform $d\tau_B, d\tau_L$ generates behavioral responses not only along the savings margin but also along the labor supply margin. Denoting by e_L the elasticity of aggregate earnings with respect to $1 - \tau_L$ (when τ_B adjusts to keep budget balance), we show that the optimal tax formula of Proposition 2 takes the form:

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

This formula is similar to the inelastic case except that e_L appears both in the numerator and denominator. τ_B increases with e_L if $\tau(1 + e_B) + s_{b0}(1 - \alpha) \geq b_y$ which is satisfied empirically.

consumption allocation along the life-cycle.

⁵⁷The 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.

Hence, a higher e_L implies a higher τ_B and a lower τ_L . Intuitively, a higher labor supply elasticity makes high labor taxation less desirable and tilts the optimal tax mix tilt more towards bequest taxes. Numerical examples presented in appendix show that, for realistic parameters, very large bequest elasticities and very small labor supply elasticities are needed to obtain $\tau_B < \tau_L$.

Closed Economy. Our optimal tax results can easily be extended to the closed economy case where the capital stock K_t is equal to domestic inheritance (i.e. $K_t = B_t$). The factor prices (wage rate and rate of return) are now endogenous and given by the marginal product of labor and capital. As in standard optimal tax theory (Diamond and Mirrlees, 1971), optimal tax formulas are independent of the production side and hence remain the same with endogenous factor prices. The important point is that the elasticity e_B (and e_L with elastic labor supply) entering the formula is the pure supply elasticity, i.e. keeping factor prices constant.

Population Growth. With exogenous population growth (at rate $1+N = e^{nH}$ per generation), all formulas carry over by simply replacing g by $g+n$. This affects b_y as high population growth reduces the bequest flow. The optimal tax formula from Proposition 2 is unchanged as the effect of n goes through b_y . High population growth countries should tax capital less, because capital accumulation is less inheritance-based and more labor-based and forward looking.

Dynamic Efficiency and Intergenerational Redistribution. Our basic model imposes a generation-by-generation government budget constraint. Hence, the government cannot accumulate assets nor liabilities. Hence, the government cannot directly affect the aggregate level of capital accumulation in the economy and hence cannot address “dynamic efficiency” issues. In Appendix C, we show that our results go through even when we relax these assumptions and allow the government to accumulate assets or liabilities. Therefore and importantly, there is decoupling of optimal capital accumulation vs. optimal labor/capital income tax mix.⁵⁸

More precisely, we prove the following. In the closed economy case, the government will accumulate sufficient assets or liabilities to ensure that the Modified Golden Rule holds whereby $r = r^* = \delta + \Gamma g$ with δ = social rate of time preference and Γ = concavity of social welfare function.⁵⁹ The government will then apply the same optimal bequest and labor tax rates as in the case with a period-by-period budget constraint with two minor modifications (appendix C,

⁵⁸The same decoupling results arise in the overlapping generation model with only life-cycle savings with linear Ramsey taxation and a representative agent per generation (King, 1980 and Atkinson and Sandmo, 1980).

⁵⁹In the small open economy case, unrestricted accumulation or borrowing by the government naturally leads to corner solutions, infinite accumulation if $r > r^*$ and maximum debt if $r < r^*$.

proposition C3). First, s_{b0} is replaced by $s_{b0}e^{\delta'H}$ in the optimal τ_B formula with $\delta' = \delta + (\Gamma - 1)g$. This correction appears because τ_{Bt} hurts bequests leavers from generation $t - 1$ while revenue accrues in generation t . Note that with no social discounting $\delta = 0$ and log-utility $\Gamma = 1$, there is no correction. Second, the formula for τ_L has to be adjusted for the interest receipt or payment term if the government has assets or debts at the optimum.

Consumption Taxes. Whether a consumption tax at rate τ_C can usefully supplement the labor and inheritances taxes τ_L, τ_B depends on which tax structures are allowed and how one models the impact of a consumption tax on private utility and government finances.

If it is completely impossible to enforce a capitalized bequest tax τ_B —so that we are constrained to have $\tau_B = 0$ —then it is in general optimal to have some positive level of consumption tax τ_C in addition to the labor income tax τ_L , since this is the only way to charge some of the tax burden to successors rather than to labor earners.⁶⁰ E.g. with no revenue requirement ($\tau = 0$), a positive consumption tax $\tau_C > 0$ allows to finance a labor subsidy $\tau_L < 0$ —and hence to transfer some resources from successors to workers. This is a rather indirect way to proceed, however, since the consumption tax is also imposed on workers.

If both τ_L and τ_B can be used, then, under simple assumptions, any tax mix (τ_C, τ_B, τ_L) is equivalent to a tax mix with zero consumption tax ($\bar{\tau}_C = 0, \bar{\tau}_B, \bar{\tau}_L$), with corrected tax rates $\bar{\tau}_B, \bar{\tau}_L$ given by: $1 - \bar{\tau}_B = (1 - \tau_C)(1 - \tau_B)$ and $1 - \bar{\tau}_L = (1 - \tau_C)(1 - \tau_L)$. Hence, consumption taxes do not expand the tax toolset and hence are not necessary to implement the optimum.

Homogenous Tastes. In contrast to existing models, our basic model assumed heterogeneity in savings tastes. If we assume homogeneity in savings tastes ($s_i \equiv s$ uniform) and i.i.d productivity shocks θ_{it} , then our results continue to apply but the distribution z_{it} of relative bequests will be more equal than the distribution of productivities (as relative bequests are just a weighted average of ancestors' productivities). Hence such a model cannot generate the very high concentration of wealth observed empirically and hence cannot be realistically calibrated.

If we further assume perfect correlation of productivity shocks across generations ($\theta_{it} = \theta_{i0}$ for all t), we lose our key ergodicity assumption. In the long run, the distribution of inheritance $\phi(z)$ would then be perfectly correlated with the distribution of labor productivity $h(\theta)$. Hence, the labor income tax τ_L and the bequest tax τ_B would have the same distributional impact. Since the latter imposes an extra utility cost - via the usual joy-of-giving externality -, there is

⁶⁰This simple point (i.e. with ill functioning capital taxes one can use consumption taxes to tax successors) was first made by Kaldor (1955). See Appendix B for a more detailed discussion.

no point having a positive τ_B .⁶¹ But as long as inequality is two-dimensional there is room for a two-dimensional tax policy tool.

Overlapping Generations and Life-cycle Savings. Our results and optimal tax formulas can also be extended to a full-fledged continuous time model with overlapping generations and life-cycle savings. We keep the same closed-form formulas for optimal inheritance tax rates. Regarding optimal lifetime capital taxation, we keep the same general, qualitative intuitions, but numerical methods are needed to compute the full optimum.

In that model, 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 (as plotted on Figures 4-5). If inheritances are received around mid-life (relative to earnings), then the cross-sectional macroeconomic ratio is close to the share of capitalized inheritance in total lifetime resources of the cohort inheriting at time t of our basic model (there is small correction factor in the b_y formula, see Appendix B). In any case, the optimal tax formulas of Proposition 2 continue to apply in this model.

For optimal lifetime capital taxation, life-cycle savings now generate an extra distortion. That is, positive tax rates on capital income $\tau_K > 0$ distort the intertemporal allocation of consumption 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 as σ is relatively small, the impact on our optimal capital tax results should be moderate.

7 Conclusion

This paper has developed a tractable normative theory of optimal capital taxation. Our results challenge the conventional zero capital tax results, which in our view rely on ad hoc assumptions which are often left implicit. If one assumes from the beginning that there is little or egalitarian inheritance, then it is perhaps 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 relaxes these assumptions, and shows that the

⁶¹With elastic labor supply, as shown by Kopczuk (2001), whether one wants to tax or subsidize bequests in the steady-state of a model with perfect correlation of abilities across generations and homogenous tastes actually hinges on the extent of the bequest externality (bequests received are a signal of ability so in some specifications one might want to tax them).

optimal tax mix between labor and capital depends on the various elasticities at play and on critical distributional parameters. We hope our results will contribute to the emergence of more pragmatic debates about capital taxation, based more upon relevant empirical parameters than abstract theoretical results relying on strong assumptions.

At a deeper level, one of our main conclusions is that the profession's emphasis on the rate of return $1 + r$ as a relative price is perhaps excessive. We do not deny that capital taxation can entail distortions in the inter-temporal allocation of consumption. But as long as the inter-temporal elasticity of substitution is moderate, this effect is likely to be second order relative to distributional issues. In our view, rates of return have two important properties. First, they tend to be large, i.e. the average rate of return r is typically much larger than the growth rate g , which implies that inheritance flows are large and that inherited wealth should be taxed at least as much as labor income otherwise society is dominated by rentiers. Next, rates of return are highly volatile and unpredictable, which implies that one should heavily use capital income taxes in order to implement the optimal tax on capitalized inheritance.

Four avenues for future research are worth noting. First, it would be useful to provide more realistic numerical simulations for more complex optimal tax structures such as nonlinear inheritance taxes and nonlinear labor taxes. Second, one could introduce credit constraints and endogenous growth in the model to generate interesting two-way interactions between growth and inheritance. The main difficulty would be the empirical calibration of such effects. Third, our model with idiosyncratic shocks to returns has assumed away aggregate uncertainty in returns that is large and pervasive in reality. With aggregate uncertainty, there is no longer a stable steady-state for the bequest to output ratio and we conjecture that the optimal inheritance tax should increase with the bequest to output ratio and the optimal capital income tax rate should increase with the aggregate return. Fourth, we have abstracted from tax competition and tax coordination across countries. Tax competition does put significant downward pressure on actual capital income taxes from a one country perspective. While such tax competition is desirable in a model where optimal capital income taxes are zero to discipline governments, it is harmful in our model with positive optimal capital income taxes. For example, for realistic parameters, bottom 50% successors lose around 20% of net income when capital taxes are constrained to be zero. With meritocratic welfare weights, the loss in aggregate social welfare has a similar magnitude. Hence, tax coordination is quantitatively very valuable in our model.

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** 1992. "On Efficient Distribution with Private Information," *Review of Economic Studies* 59, 427-453.
- Atkinson, Anthony** 1980. "Inheritance and the Redistribution of Wealth", in G.M. Heal and G.A. Hughes, eds., *Public Policy and the Tax System*, 36-66, London: Allen and Unwin.
- Atkinson, Anthony, Thomas Piketty and Emmanuel Saez** 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49(1): 3-71.
- Atkinson, Anthony, and Agnar Sandmo** 1980. "Welfare Implications of the Taxation of Savings", *The Economic Journal*, 90, 529-549.
- 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.
- Atkinson, A.B., and Joseph Stiglitz** 1980. *Lectures on Public Economics*, McGraw Hill.
- Beckert, Jens** 2008. *Inherited Wealth*, Princeton University Press, 382p.
- Cagetti, Marco, and Mariacristina De Nardi** 2009. "Estate Taxation, Entrepreneurship, and Wealth." *American Economic Review*, 99(1), 85-111.
- Cai, Hongbin and Daniel Treisman** 2005. "Does Competition for Capital Discipline Governments? Decentralization, Globalization, and Public Policy." *American Economic Review* 95(3), 817-830.
- 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, Juan Carlos, Sagiri Kitao, and Dirk 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, 2475-90.
- Diamond, Peter** 2006. "Optimal Tax Treatment of Private Contributions for Public Goods with and without Warm Glow Preferences," *Journal of Public Economics*, 90(4-5), 897-919.
- Diamond, Peter, and James A. Mirrlees** 1971. "Optimal Taxation and Public Production: Production Efficiency." *American Economic Review*, 61(1), 8-27.
- Diamond, Peter, and Emmanuel Saez** 2011. "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25(4), 165-190.
- Edwards, Chris and Daniel J. Mitchell** 2008. *Global Tax Revolution: The Rise of Tax Competition and the Battle to Defend It*, Cato Institute, 250p.
- European Commission** 2011. *Taxation trends in the European Union*, 2011 Edition, Eurostat.

- Farhi, Emmanuel, Christopher Sleet, Iván Werning and Sevin Yeltekin** 2011. “Non-linear Capital Taxation Without Commitment”, Harvard Working Paper.
- Farhi, Emmanuel and Iván Werning** 2007. “Inequality and Social Discounting”, *Journal of Political Economy*, 115(3), 365-402.
- Farhi, Emmanuel and Iván Werning** 2010. “Progressive Estate Taxation,” *Quarterly Journal of Economics*, 125(2), 635-673.
- Farhi, Emmanuel and Iván Werning** 2011. “Capital Taxation: Quantitative Explorations of the Inverse Euler Equation,” MIT Working Paper.
- Fleurbaey, Marc, and François Maniquet** 2006. “Fair Income Tax”, *Review of Economic Studies* 73, 55-83.
- Fleurbaey, Marc** 2008. *Fairness, Responsibility and Welfare*, Oxford University Press.
- Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski** 2011. “Optimal Dynamic Taxes.” Princeton Working Paper.
- Golosov, Mikhail, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl** 2011. “Preference Heterogeneity and Optimal Capital Income Taxation”, Yale Working Paper.
- Golosov, Michael, Tsyvinski, Aleh, and Ivan Werning** 2006. “New Dynamic Public Finance: A User’s Guide,” NBER Macroeconomics Annual.
- Gordon, Roger H.** 1985. “Taxation of Corporate Capital Income: Tax Revenues Versus Tax Distortions”, *Quarterly Journal of Economics*, 100, 1-27.
- Graetz, Michael J. and Ian Shapiro** 2005. *Death by a Thousand Cuts: The Fight over Taxing Inherited Wealth*, Princeton, NJ: Princeton University Press.
- Judd, Kenneth** 1985. “Redistributive Taxation in a Simple Perfect Foresight Model.” *Journal of Public Economics*, 28(1), 59–83.
- Judd, Kenneth** 1999. “Optimal taxation and spending in general competitive growth models,” *Journal of Public Economics*, 71, 1-26.
- Kaldor, Nicholas** 1955. *An Expenditure Tax*, Unwin University Books, 248p.
- Kaplow, Louis** 2001. “A Framework for Assessing Estate and Gift Taxation”, in Gale, William G., James R. Hines Jr., and Joel Slemrod (eds.) *Rethinking estate and gift taxation* Washington, D.C.: Brookings Institution Press.
- Kaplow, L.** 2008. *The Theory of Taxation and Public Economics*, Princeton University Press.
- King, Mervyn** 1980. “Savings and Taxation”, in G. Hughes and G. Heal, eds., *Public Policy and the Tax System* (London: George Allen Unwin), 1-36.
- Kocherlakota, Narayana** 2010. *The New Dynamic Public Finance* (Princeton: Princeton University Press).
- Kopczuk, Wojciech** 2001. “Optimal Estate Taxation in the Steady-State”, Working Paper, University of British Columbia.
- Kopczuk, Wojciech** 2009. “Economics of Estate Taxation: Review of Theory and Evidence”, *Tax Law Review* 63(1), 139–157.
- Kopczuk, Wojciech** 2012. “Taxation of Transfers and Wealth”, in preparation *Handbook of Public Economics*, Volume 5, Elsevier.

- Kopczuk, Wojciech and J. Lupton** 2007. “To Leave or Not to Leave: The Distribution of Bequest Motives”, *Review of Economic Studies*, 74(1), 207-235.
- Kopczuk, Wojciech and Joel Slemrod** 2001. “The Impact of the Estate Tax on Wealth Accumulation and Avoidance Behavior”, in *Rethinking Estate and Gift Taxation*, W. Gale, J.R. Hines and J. Slemrod eds., The Brookings Institution, 299–349.
- Lucas, Robert** 1990. “Supply-Side Economics: An Analytical Review”, *Oxford Economic Papers*, New Series, 42(2), 293–316.
- OECD** 2011. *Revenue Statistics, 1965-2010*, OECD, Paris, 2011 Edition.
- Piketty, Thomas** 2010. “On the Long-Run Evolution of Inheritance: France 1820-2050”, Working Paper, Paris School of Economics, 424p.
- Piketty, Thomas** 2011. “On the Long-Run Evolution of Inheritance: France 1820-2050”, *Quarterly Journal of Economics*, 126(3), 1071–1131.
- Piketty, Thomas, Gilles Postel-Vinay and Jean-Laurent Rosenthal** 2006. “Wealth Concentration in a Developing Economy: Paris and France, 1807-1994”, *American Economic Review*, 96(1), 236-256.
- Piketty, Thomas, Gilles Postel-Vinay and Jean-Laurent Rosenthal** 2011. “Inherited vs Self-Made Wealth—Theory and Evidence from a Rentier Society (Paris 1872-1937)”, Working Paper, Paris School of Economics.
- Piketty, Thomas, and Emmanuel Saez** 2012. “Optimal Labor Income Taxation”, in preparation *Handbook of Public Economics*, Volume 5, Elsevier.
- Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva** 2011. “Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities”, NBER Working Paper No. 17616.
- Pirttila, Jukka, and Hakan Selin** 2011. “Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993.” *Scandinavian Journal of Economics*, 113(1).
- Roemer, John et al.** 2003. “To What Extent Do Fiscal Systems Equalize Opportunities for Income Acquisition Among Citizens?”, *Journal of Public Economics*, 87, 539–565.
- Saez, Emmanuel** 2001. “Using Elasticities to Derive Optimal Tax Rates”, *Review of Economic Studies*, 68, 205-229.
- Saez, Emmanuel** 2002. “The Desirability of Commodity Taxation under Nonlinear Income Taxation and Heterogeneous Tastes,” *Journal of Public Economics*, 83, 217-230.
- Saez, Emmanuel, Joel Slemrod, and Seth Giertz** 2012. “The Elasticity of Taxable Income with respect to Marginal Tax Rates: A Critical Review”, *Journal of Economic Literature*, 50(1).
- Scheve, Kenneth and David Stasavadge**, “Democracy, War and Wealth: Lessons from Two Centuries of Inheritance Taxation”, Working Paper, NYU, 2011
- Stiglitz, Joseph E.** 1985. “Inequality and Capital Taxation”, Working Paper, Stanford.
- Vickrey, William** 1947. *Agenda for Progressive Taxation*, New York: Ronald, 496p.
- Weinzierl, Matthew** 2011. “The Surprising Power of Age-Dependent Taxes”, *Review of Economic Studies* 78(4), 1490-1518.
- Wilhelm, Mark O.** 1996. “Bequest Behavior and the Effect of Heirs’ Earnings: Testing the Altruistic Model of Bequests”, *American Economic Review* 86(4), 874-892.

Figure 1: Top Inheritance Tax Rates 1900-2011

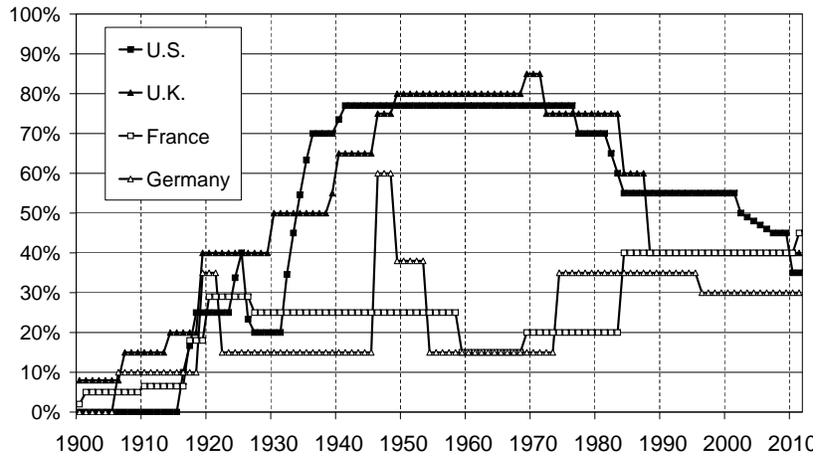


Figure 2: Top Income Tax Rates 1900-2011

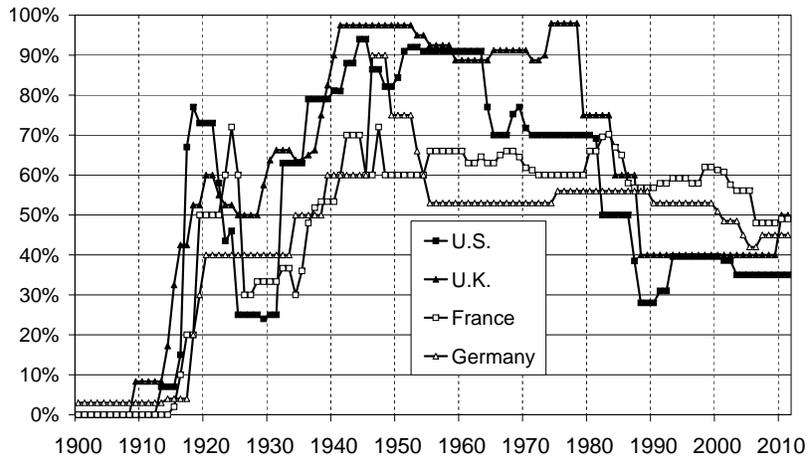


Figure 3: Top Income Tax Rates: Earned (Labor) vs Unearned (Capital)

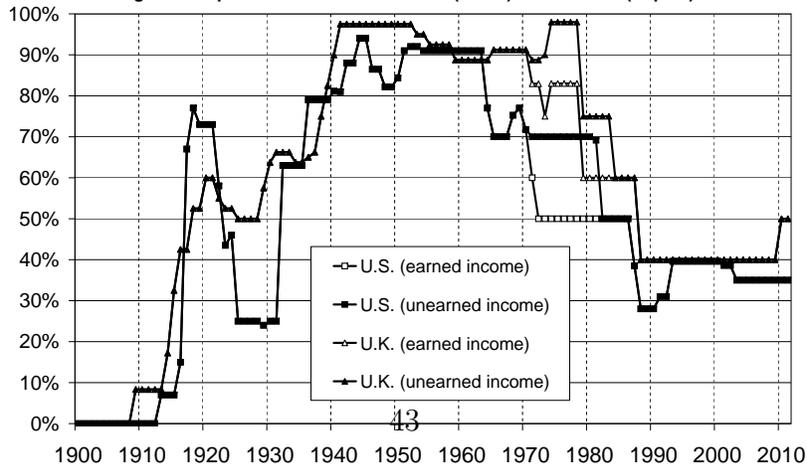
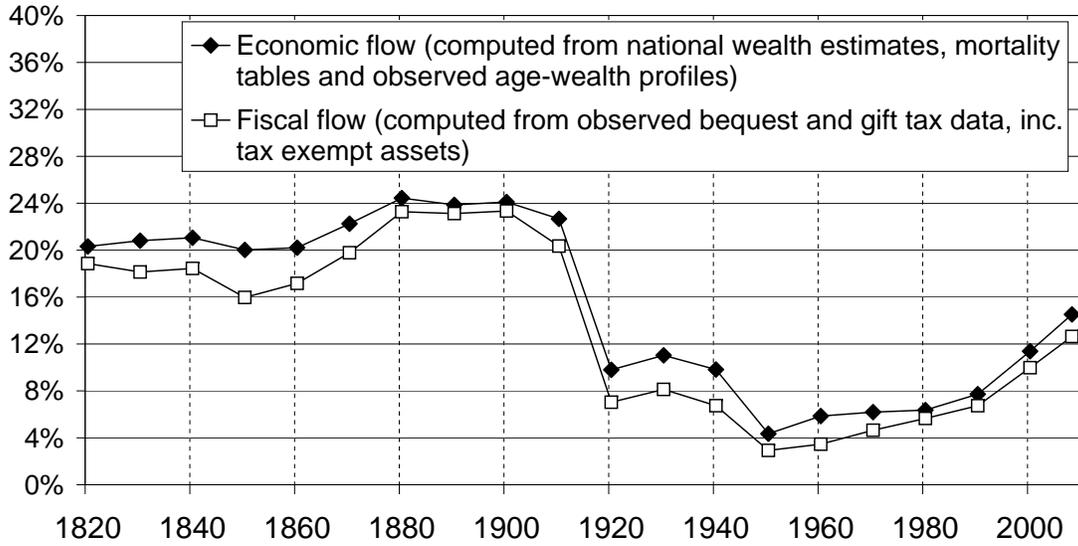
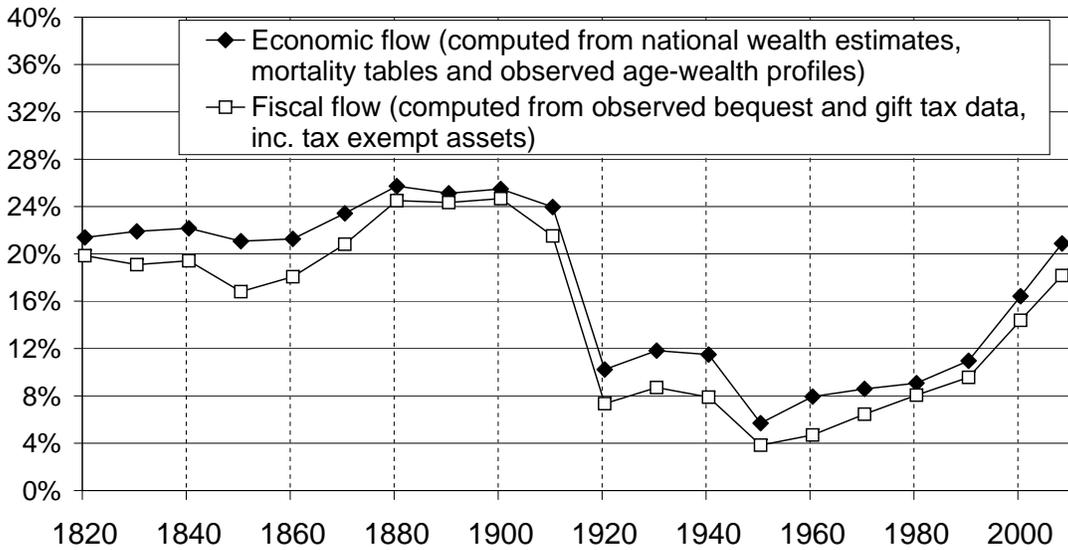


Figure 4: Annual inheritance flow as a fraction of national income, France 1820-2008



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

Appendix (not for publication)

This appendix is organized as follows. Appendix A presents all the proofs of the formal Propositions in the main text. Appendix B presents the extensions mentioned in the text in Sections 6 and Conclusion, to the exception of dynamic efficiency and inter-generational redistribution that is presented in the long and self-contained Appendix C.

A Proofs of Main Text Propositions

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

A.1.1 Main Proof

The four-dimensional, discrete-time stochastic process $X_{ti} = (z_{ti}, s_{wti}, s_{bti}, \theta_{ti})$ is a Markovian process with a state variable $b_{yt} = \frac{e^{rH} b_t}{y_t}$. In the special case with i.i.d. taste and productivity shocks, (s_{wti}, s_{bti}) and θ_{ti} are given by the stationary distributions $g(s_{wi}, s_{bi})$ and $h(\theta_i)$, and we can concentrate upon the convergence of the Markovian process for z_{ti} .

We have the following endogenous transition equation for inheritance b_{ti} :

$$b_{t+1i} = s_{ti}[(1 - \tau_L)y_{Lti} + (1 - \tau_B)b_{ti}e^{rH}] \quad (10)$$

This can be rewritten as a transition equation for normlized inheritance $z_{ti} = \frac{b_{ti}}{b_t}$:

$$z_{t+1i} = \frac{s_{ti}[(1 - \tau_L)(1 - \alpha)e^{(r-g)H}\theta_{ti} + (1 - \tau_B)b_{yt}e^{(r-g)H}z_{ti}]}{b_{yt+1}} \quad (11)$$

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

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

where $s = E(s_i)$ is the average saving taste. If we rule out explosive paths (assumption 3: $s \cdot e^{(r-g)H} < 1$), then whatever the initial conditions the state variable b_{yt} converges towards a unique given b_y as $t \rightarrow +\infty$, where b_y is given by:

$$b_y = \frac{s(1 - \tau_L)(1 - \alpha)e^{(r-g)H}}{1 - s(1 - \tau_B)e^{(r-g)H}}$$

As $t \rightarrow +\infty$, the transition equation for z_{ti} can therefore be rewritten as follows (by replacing b_{yt}, b_{yt+1} by b_y in the above transition equation, and by noting $\mu = s(1 - \tau_B)e^{(r-g)H} < 1$):

$$z_{t+1i} = \frac{s_{ti}}{s}[(1 - \mu) \cdot \theta_{ti} + \mu \cdot z_{ti}] \quad (12)$$

In the long-run, the minimal normalized inheritance level z_0 is given by: $z_0 = \frac{s_0(1 - \mu)\theta_0}{s - s_0\mu} < 1$. This is what an individual would get if his ancestors permanently receive the lowest possible

taste and productivity shocks (i.e. $s_{ti} = s_0$ and $\theta_{ti} = \theta_0$). In case $s_0 = 0$ (assumption 1), then $z_0 = 0$, i.e. there are zero bequest receivers in the long run.⁶²

In case $\frac{s_1}{s}\mu < 1$, then the long-run maximal normalized inheritance level z_1 is given by: $z_1 = \frac{s_1(1-\mu)\theta_1}{s-s_1\mu} > 1$. This is what an individual would get if his ancestors permanently receive the highest possible taste and productivity shocks (i.e. $s_{ti} = s_1$ and $\theta_{ti} = \theta_1$). In case $\frac{s_1}{s}\mu > 1$, then $z_1 = +\infty$, i.e. the long-run distribution of normalized inheritance is unbounded above (see example below).

In any case, thanks to assumptions 1 and 2, one can see that the Markovian process for z_{ti} verifies the following ‘‘concavity property’’ over the interval $[z_0, z_1]$: for any relative inheritance positions $z_0 \leq z < z' < z'' \leq z_1$, there exists $T \geq 1$ and $\varepsilon > 0$ such that $\text{proba}(z_{it+T} > z' \mid z_{it} = z) > \varepsilon$ and $\text{proba}(z_{it+T} < z' \mid z_{it} = z'') > \varepsilon$. (consider a sufficiently long sequence of good shocks in the first case, and of bad shocks in the second case). 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)). **QED.**

A.1.2 Extension to general random processes

For notational simplicity, we choose to concentrate throughout the paper upon the special case with i.i.d. taste and productivity shocks. Under additional assumptions, all our results and optimal tax formulas can be extended to the case with general random processes regarding taste and productivity shocks. E.g. assume exogenous transition functions $g(s_{wt+1i}, s_{bt+1i} \mid s_{wti}, s_{bti})$ and $h(\theta_{t+1i} \mid \theta_{ti})$.⁶³ In order to ensure global convergence of the Markovian process $X_{ti} = (z_{ti}, s_{wti}, s_{bti}, \theta_{ti})$, one must make ergodicity assumptions about these transition functions. That is, one must modify assumptions 1 and 2) and assume that the transition functions are monotonic (in the sense of first order stochastic dominance; see above) and have full support, in the sense that starting from any parental taste or productivity there is always a positive probability to attain any other taste or productivity:

$$\forall (s_{wti}, s_{bti}) \in S, (s_{wt+1i}, s_{bt+1i}) \in S, g(s_{wt+1i}, s_{bt+1i} \mid s_{wti}, s_{bti}) > 0$$

(with : $s_{wti}, s_{bti} =$ parental tastes, $s_{wt+1i}, s_{bt+1i} =$ children tastes)

And:

⁶²Note that the same conclusion $z_0 = 0$ would hold in case $s_0 > 0$ and $\theta_0 = 0$.

⁶³One could also assume more general forms (with taste or productivity memory over more than one generation, or with joint processes), as long as one makes adequate ergodicity assumptions.

$\forall \theta_{ti} \in \Theta, \theta_{t+1i} \in \Theta, h(\theta_{t+1i} | \theta_{ti}) > 0$
(with : θ_{ti} = parental productivity, θ_{t+1i} = children productivity)

Under these assumptions, standard ergodic convergence theorems ensure that for any initial distribution of tastes and productivities, the distributions $g_t(s_{wi}, s_{bi})$ and $h_t(\theta_i)$ converge towards unique stationary distributions $g(s_{wi}, s_{bi})$ and $h(\theta_i)$.

The law of motion for the state variable b_{yt} can now be written:

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

where $s = E(s_i)$ is the average saving taste and $s_{tz} = E(s_{ti}z_{ti})$ is the average saving taste weighted by normalized inheritance.

In the no-taste-memory special case (tastes are drawn i.i.d. at each generation), then $s_{ti} \perp z_{ti}$, so we have: $\forall t, s_{tz} = s$.

In the general case with taste memory, s_{ti} and z_{ti} might be correlated (they are both determined - partly - by parental tastes s_{t-1i}), so s_{tz} might differ from s . Typically children of high-saving-taste parents might have both higher saving taste and higher inheritance, so that $s_{tz} \geq s$.⁶⁴

Assume that the state variable b_{yt} converges towards a given b_y as $t \rightarrow +\infty$. This implies that s_{tz} has converged towards some given s_z (which can differ from s in the case with state memory) and that b_y satisfies:

$$b_y = \frac{s(1 - \tau_L)(1 - \alpha)e^{(r-g)H}}{1 - s_z(1 - \tau_B)e^{(r-g)H}}$$

As $t \rightarrow +\infty$, the transition equation for z_{ti} can therefore be rewritten as follows (again by replacing b_{yt}, b_{yt+1} by b_y in the above transition equation, and by noting $\mu = s(1 - \tau_B)e^{(r-g)H}$ and $\mu_z = s_z(1 - \tau_B)e^{(r-g)H}$):

$$z_{t+1i} = \frac{s_{ti}}{s}[(1 - \mu_z) \cdot \theta_{ti} + \mu \cdot z_{ti}] \tag{13}$$

The long run minimal and maximal normalized inheritance level z_0 and z_1 are now given by: $z_0 = \frac{s_0(1 - \mu_z)\theta_0}{s - s_0\mu}$ and $z_1 = \frac{s_1(1 - \mu_z)\theta_1}{s - s_1\mu}$ (assuming $\frac{s_1}{s}\mu < 1$; otherwise $z_1 = +\infty$, i.e. the long-run inheritance distribution is unbounded above).

In any case, one can see that the Markovian process for z_{ti} again verifies the concavity and monotonicity properties over the interval $[z_0, z_1]$. So we can again apply standard ergodic

⁶⁴Note that $s_{ti} \perp \theta_{ti}$ (whether or not there is productivity memory), so that: $\forall t, s_{t\theta} = E(s_{ti}\theta_{ti}) = s$. This is because we assumed the taste and productivity processes to be uncorrelated. This could easily be relaxed, providing that we make the appropriate full support assumption on the joint random process (so as to ensure ergodicity). One would then simply need to replace s by $s_{t\theta} = E(s_{ti}\theta_{ti})$ in the law of motion for b_{yt} .

convergence theorems to derive the existence of a unique stationary distribution $\phi(z)$ towards which $\phi_t(z)$ converges, independently of the initial distribution $\phi_0(z)$.

The only difference with the case with i.i.d. shocks is that there might now exist multiple steady-states for the b_{yt} process. First, in order to rule out explosive paths, we need to generalize assumption 3 and assume the following:

$$\bar{s} \cdot e^{(r-g)H} < 1, \text{ with: } \bar{s} = E(s_{t+1i} \mid s_{ti} = s_1)$$

This assumption ensures that for any initial condition, b_{yt} converges towards some finite b_y , and then - given b_y - the distributions $\phi_t(z)$ and $\Psi_t(z, \theta)$ converges towards unique stationary, ergodic distributions $\phi(z)$ and $\Psi(z, \theta)$. However this assumption is not sufficient to rule out the possibility of multiple steady-states. That is, one can construct examples where there are multiple steady-state pairs $(b_y, \Psi(z, \theta))$. Intuitively, a higher steady-state b_y (i.e. a higher steady-state s_z and μ_z) can be self-fulfilling because it implies higher steady-state inequality of inheritance, via the transition equation for z_{ti} : higher μ_z and $s_z = E(s_{ti}z_{ti})$ imply a smaller labor income term (i.e. a smaller equalizing effect) relatively to the multiplicative inheritance effect (i.e. relatively to the un-equalizing effect), which also tends to generate higher steady-state correlation between normalized inheritance z_i and saving taste s_i . This in turn can validate a higher steady-state s_z and b_y . Each steady-state is ergodic, but there is more inequality and less mobility in the high b_y steady-state. This ergodic steady-state multiplicity is similar to that studied by Piketty (1997) (but with b_y instead of r in the role of the state variable). This possibility can be ruled out in simple examples with binomial random tastes and taste memory (see below). But in order to rule it out in the general case, one would need stronger assumptions, e.g. one would need to assume that there is not too much taste persistence (in the sense that $s \rightarrow \bar{s}(s) = E(s_{t+1i} \mid s_{ti} = s)$ is not too steeply increasing over the interval $[s_0, s_1]$). In any case, note that this is a relatively secondary issue for our purposes in this paper. I.e. even if there were multiple steady-state values for b_y , then our optimal tax formulas would still be locally valid as long the tax change does not shift the economy towards another b_y steady-state.

A.1.3 Example with binomial random tastes

With i.i.d. binomial random taste shocks $s_{ti} = s_0 = 0$ with probability $1 - p$, and $s_{ti} = s_1 > 0$ with probability p , we have: $s = s_z = ps_1$, $\mu = \mu_z = s(1 - \tau_B)e^{(r-g)H}$. We assume: $\mu < 1 < \mu/p$. With no productivity heterogeneity, the transition equation for z_{ti} looks as follows:

$$z_{t+1i} = \frac{s_{ti}}{s} [(1 - \mu) + \mu \cdot z_{ti}]$$

That is:

$z_{t+1i} = 0$ with probability $1 - p$

$$z_{t+1i} = \frac{1 - \mu}{p} + \frac{\mu}{p} \cdot z_{ti}$$

It follows that the long-run distribution of normalized inheritance $\varphi(z)$ looks as follows:⁶⁵

⁶⁵In case $\mu/p < 1$, then $z_k = \frac{1 - \mu}{p - \mu} \cdot [1 - (\frac{\mu}{p})^k]$ has a finite upper bound $z_1 = \frac{1 - \mu}{p - \mu}$. Note that we do not

$z = z_k = \frac{1 - \mu}{\mu - p} \cdot [(\frac{\mu}{p})^k - 1]$ with probability $(1 - p) \cdot p^k$

As $k \rightarrow +\infty$, $z_k \approx \frac{1 - \mu}{\mu - p} \cdot (\frac{\mu}{p})^k$.

The cumulated distribution is given by: $1 - \Phi(z_k) = \text{proba}(z \geq z_k) = \sum_{k' \geq k} (1 - p) \cdot p^{k'} = p^k$.

It follows that as $z \rightarrow +\infty$, $\log[1 - \Phi(z)] \approx a \cdot [\log(\omega) - \log(z)]$, i.e. $1 - \Phi(z) \approx (\frac{\omega}{z})^a$,

with a constant term $\omega = \frac{1 - \mu}{\mu - p}$, a Pareto coefficient $a = \frac{\log(1/p)}{\log(\mu/p)} > 1$ and an inverted Pareto

coefficient $b = \frac{a}{a - 1} = \frac{\log(1/p)}{\log(1/\mu)} > 1$.

As $\mu \rightarrow 1^-$ (for given $p < \mu$), $a \rightarrow 1^+$ and $b \rightarrow +\infty$ (infinite inequality). Intuitively, the multiplicative inheritance effect becomes infinite as compared to the equalizing labor income effect. The same occurs as $p \rightarrow 0^+$ (for given $\mu > p$): an infinitely small group gets infinitely large random shocks.⁶⁶

Note 1. All theoretical wealth accumulation models with multiplicative random shocks give rise to distributions with Pareto upper tails, whether the shocks are binomial or multinomial, and whether they come from taste shocks or other kind of multiplicative shocks (such as shocks on rates of returns or demographic shocks on numbers of children, rank of birth or age at parenthood or age at death). Empirically, upper tails of wealth distribution follow Pareto laws, as well as upper tails of income distribution, in particular due to top capital incomes (labor incomes tend to dampen inequality and reduce inverted Pareto coefficient). Low income inequality typically corresponds to $b \simeq 1.5$; high income inequality to $b \simeq 2.5 - 3$. For wealth distributions, inverted Pareto coefficients often exceed $b \simeq 3 - 3.5$. For references to theoretical models and historical series on Pareto coefficients, see Atkinson, Piketty, Saez (2011, pp.13-14 and 50-58).

Note 2. One can easily introduce intergenerational taste persistence into this setting. E.g. assume a binomial random taste process with $s_0 = 0$, $s_1 > 0$ and with taste memory, say $s_{t+1i} = s_1$ with probability p_0 if $s_{ti} = 0$, and with probability $p_1 \geq p_0$ if $s_{ti} = s_1$. The steady-state taste distribution involves a fraction $1 - p$ of the population with zero wealth taste and p with positive wealth taste, with: $p \cdot (1 - p_1) = (1 - p) \cdot p_0$, i.e. $p = \frac{p_0}{1 + p_0 - p_1} \in [p_0, p_1]$. The average steady-state taste s is given by: $s = p \cdot s_1$. The steady-state distribution of normalized inheritance $\varphi(z)$ looks as follows:

$z = z_0 = 0$ with probability $1 - p$ (children with zero-wealth-taste parents)

$z = z_1 = \frac{1 - \mu z}{p} > 0$ with probability $(1 - p) \cdot p_0$ (children with wealth-loving parents but

need to specify the decomposition between wealth and bequest tastes ($s_1 = s_{1w} + s_{1b}$): this matters for welfare, but has no impact on the transition equations and steady-state distributions.

⁶⁶In the binomial model, one can directly compute the “empirical” inverted Pareto coefficient $b' = \frac{E(z | z \geq z_k)}{z_k} \rightarrow \frac{1 - p}{1 - \mu}$ as $k \rightarrow +\infty$. Note that $b' \simeq b$ if $p, \mu \simeq 1$ but that the two coefficients generally differ because the true distribution is discrete, while the Pareto law approximation is continuous.

zero-wealth-taste grand-parents)

...

$z = z_{k+1} = \frac{1 - \mu_z}{p} + \frac{\mu}{p} \cdot z_k > z_k$ with probability $(1 - p) \cdot p_0 \cdot p_1^k$ (children with wealth-loving

ancestors during the past $k + 1$ generations)

That is: $z_k = \frac{1 - \mu_z}{\mu - p} \cdot \left[\left(\frac{\mu}{p}\right)^k - 1\right]$ with probability $(1 - p) \cdot p_0 \cdot p_1^{k-1}$ if $k \geq 1$

One can easily see that: $s_z = E(s_i z_i) = \frac{p_1}{p} s = p_1 s_1 > s$. It follows that there again exists a unique steady-state b_y .

Note that it is less straightforward to guarantee steady-state uniqueness in the case of a binomial random taste process with $0 < s_0 < s_1$ and with taste memory. First note that with $s_0 > 0$ there exists a positive lower bound z_0 for the steady-state distribution $\varphi(z)$, with

$z_0 = \frac{\frac{s_0}{s}(1 - \mu_z)}{1 - \frac{s_0}{s}\mu} > 0$ (s is given by $s = (1 - p)s_0 + ps_1$). If $\frac{s_1}{s}\mu > 1$, there is no finite upper

bound (i.e. $z_1 = +\infty$), and the steady-state distribution is a continuous Pareto distribution over $[z_0; +\infty[$: $1 - \Phi(z) = \left(\frac{z_0}{z}\right)^a$, with $(1 - p) \cdot \mu_0^a + p \cdot \mu_1^a = 1$.⁶⁷ If $s_0 = s_1 = s$, then $z_0 = 1$ and $a = +\infty$ (perfect equality). Conversely as the variance rises, $z_0 \rightarrow 0$ and $a \rightarrow 1$ (infinite inequality). If we now introduce taste persistence $p_0 < p < p_1$, the steady-state distribution takes a more complicated form. We now have a declining fraction $p(z)$ of high-taste individuals as a function of $z \in [z_0; +\infty[$. The long run distribution $\varphi(z)$ has no reason to be Pareto any longer, because the distribution of the multiplicative shock is not the same for all z . One can construct numerical examples where p_0 is sufficiently low and p_1 sufficiently large so that there is steady-state multiplicity in b_y of the form described above.

A.2 Proof of Proposition 2 (basic optimal tax formula) (section 4)

The proof of Proposition 2 is given in the main text of the paper (section 4). Here we discuss and clarify the following points.

A.2.1 Simplified proof with Cobb-Douglas utility

The proof given in section 4 works with any utility function that is homogenous of degree one, and with any random process for tastes and productivity shocks. With Cobb-Douglas utility functions, there exists a simpler proof, since we have:

$$\begin{aligned} V_{ti} &= \max V_i(c_{ti}, w_{ti}, \bar{b}_{ti}) = c_{ti}^{1-s_i} w_{ti}^{s_i} \bar{b}_{ti}^{s_i} \quad \text{s.t.} \quad c_{ti} + w_{ti} \leq \tilde{y}_{ti} = (1 - \tau_B) z_{ti} b_t e^{rH} + (1 - \tau_L) \theta_{ti} y_{Lt} \\ &\rightarrow c_{ti} = (1 - s_i) \cdot \tilde{y}_{ti}, w_{ti} = s_i \cdot \tilde{y}_{ti}, \bar{b}_{ti} = (1 - \tau_B) e^{rH} \cdot s_i \cdot \tilde{y}_{ti} \end{aligned}$$

I.e. $V_{ti} = v_i \cdot \tilde{y}_{ti}$, with $v_i = (1 - s_i)^{1-s_i} s_i^{s_i} [(1 - \tau_B) e^{rH}]^{s_i}$, and $V_{ci} = v_i$

⁶⁷The formula works with any multinomial or continuous distribution of multiplicative shocks, not just with binomial shocks. See Nirei (2009, Proposition 1, p.9). See also Stiglitz (1969).

With $z_{ti} = 0$, we have $\tilde{y}_{ti} = (1 - \tau_L)\theta_i y_{Lt}$. So: $\max_{\tau_B, \tau_L} V_{ti} \iff \max_{\tau_B, \tau_L} (1 - \tau_B)^{s_{bi}}(1 - \tau_L)$. Since $1 - \tau_L = (1 - \alpha - \tau + \tau_B b_y)/(1 - \alpha)$ (from the government budget constraint), this is equivalent to:

$$\max_{\tau_B} (1 - \tau_B)^{s_{bi}}(1 - \alpha - \tau + \tau_B b_y)$$

In case $s_{bi} = 0$ (zero bequest taste), this is equivalent to bequest tax revenue maximization: $\max_{\tau_B} \tau_B b_y \iff \tau_B = \frac{1}{1 + e_B}$.

More generally, in case $s_{bi} \geq 0$, the first order condition in τ_B immediately leads to: $\tau_B = \frac{1 - (1 - \alpha - \tau)s_{bi}/b_y}{1 + e_B + s_{bi}}$

Now, assume that we maximize the social welfare function $SWF = E(V_{ti} \mid z_i = 0) = v \cdot (1 - \tau_L)y_{Lt}$, with $v = E(v_i \cdot \theta_i \mid z_i = 0)$.

The first-order condition in τ_B leads to: $\tau_B = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}}$, with $s_{b0} = \frac{E(v_i \cdot \theta_i \cdot s_{bi} \mid z_i = 0)}{E(v_i \cdot \theta_i \mid z_i = 0)}$.

Note that since $V_{ci} = v_i$ in the Cobb-Douglas utility case, this is equivalent to the definition of s_{b0} obtained in the general case.

With i.i.d. productivity and taste shocks, then $\theta_i \perp s_{bi} v_i$, so we have: $v = E(v_i)$ and $s_{b0} = \frac{E(v_i \cdot s_{bi})}{E(v_i)}$. That is, v and s_{b0} are entirely determined by the exogenous distribution of taste parameters $g(s_{wi}, s_{bi})$. Note however that v_i and s_{bi} are not orthogonal, so that in general $s_{b0} \neq s_b = E(s_{bi})$. This is due to the absence of utility normalization (see discussion below).

A.2.2 Utility normalization and social welfare

In section 3, we define social welfare by summing up heterogenous utility functions without imposing any utility normalization. That is, we define: $SWF = \iint_{z \geq 0, \theta_0 \leq \theta \leq \theta_1} \omega_{p_z p_\theta} \frac{V_{z\theta}^{1-\Gamma}}{1-\Gamma} d\Psi(z, \theta)$, with $V_{z\theta} = E(V_i \mid z_i = z, \theta_i = \theta)$ (see section 3). I.e. $V_{z\theta}$ is defined as average steady-state utility level V_i attained by individuals i with the same normalized inheritance $z_i = z$ and productivity $\theta_i = \theta$ (i.e. the same after-tax total income $\tilde{y}_{ti} = \tilde{y}_{tz\theta} = (1 - \tau_B)z b_t e^{rH} + (1 - \tau_L)\theta y_{Lt}$) but with different taste parameters s_{wi}, s_{bi} . In effect, we are implicitly assuming that the welfare weights ω_i are the same for all individuals i with the same ranks p_z, p_θ in the distribution of normalized inheritance and productivity, i.e. are the same for all taste parameters s_{wi} and s_{bi} . The absence of utility normalization implies that in effect we put more weight on agents with utility functions delivering higher marginal utility for consumption (which is relatively arbitrary from a normative viewpoint). So for instance in the case with Cobb-Douglas utility functions and i.i.d. shocks, we have: $V_{z\theta} = v \cdot \tilde{y}_{tz\theta}$, with $v = E(v_i)$; and $\tau_B = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}}$, with $s_{b0} = \frac{E(v_i \cdot s_{bi})}{E(v_i)}$. So in effect the average bequest taste s_{b0} that matters for the optimal tax policy is different from the raw average bequest taste $s_b = E(s_{bi})$, because we put more weight on individuals with higher marginal utility $V_{ci} = v_i = (1 - s_i)^{1-s_i} s_i^{s_i} [(1 - \tau_B)e^{rH}]^{s_{bi}}$ (which is

not particularly appealing).⁶⁸

E.g. in the binomial random taste example with $s_i = 0$ with probability $1 - p$, and $s_i = s_{w1} + s_{b1} = s_1 > 0$ with probability p , we have: $s_{b0} = \frac{p \cdot v_1 \cdot s_{b1}}{1 - p + p \cdot v_1}$ (with $v_1 = s_1^{s_1} (1 - s_1)^{1 - s_1} [(1 - \tau_B) e^{rH}]^{s_{b1}}$). That is, depending whether $v_1 > 1$ or $v_1 < 1$, then $s_{b0} > s_b$ or $s_{b0} < s_b$ (where $s_b = p \cdot s_{b1} = E(s_{bi})$).

All our results can easily be extended to allow social welfare weights ω_i to depend on taste parameters, for instance for utility normalization purposes. For instance assume we define $V_{z\theta} = E(\omega_i \cdot V_i \mid z_i = z, \theta_i = \theta)$, with $\omega_i = 1/v_i$ (so as to normalize marginal utilities), and the same *SWF* definition as before. The zero-receiver tax optimum would then be: $\tau_B = \frac{1 - (1 - \alpha - \tau) s_{b0}/b_y}{1 + e_B + s_{b0}}$ with $s_{b0} = \frac{E(\omega_i \cdot v_i \cdot s_{bi})}{E(\omega_i \cdot v_i)} = E(s_{bi}) = s_b$. Note however that the weights $\omega_i = 1/v_i$ would have to be endogenous, in the sense that they need to be defined at the level of the optimal τ_B (so that marginal utilities are normalized right at the optimum).⁶⁹

In the special Cobb-Douglas case, an alternative equivalent formulation would be the following log form: define $V_{z\theta} = \exp(E(\log(V_i) \mid z_i = z, \theta_i = \theta))$, again with the same *SWF* definition as before. We would then have: $V_{z\theta} = v' \cdot [(1 - \tau_B) e^{rH}]^{s_b} \cdot \tilde{y}_{tz\theta}$, with $s_b = E(s_{bi})$ and $v' = \exp[E(\log((1 - s_i)^{1 - s_i} s_i^{s_i}))]$. The first order condition in τ_B leads directly to: $\tau_B = \frac{1 - (1 - \alpha - \tau) s_{b0}/b_y}{1 + e_B + s_{b0}}$, with $s_{b0} = s_b$.

A.2.3 Conditions under which $\tau_B > 0$.

Finally, we discuss and clarify the conditions under which $\tau_B = \frac{1 - (1 - \alpha - \tau) s_{b0}/b_y}{1 + e_B + s_{b0}} > 0$.

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 ($\tau_B > 0$) if and only if the latter effect is larger than the former. Conversely, if $b_y < s_{b0}(1 - \alpha - \tau)$, then zero-receivers prefer bequest subsidies ($\tau_B < 0$). Although this is a theoretical possibility, this requires pretty extreme parameters.

E.g. consider the case with Cobb-Douglas utility, i.i.d. taste and productivity shocks, and adequate utility normalization, so that $s_{b0} = s_b = E(s_{bi})$. By substituting $b_y = \frac{s(1 - \tau - \alpha)e^{(r-g)H}}{1 - se^{(r-g)H}}$ into the τ_B formula, we obtain: $\tau_B = \frac{1 + s_b - (s_b/s)e^{-(r-g)H}}{1 + e_B + s_b}$. We get the following condition

⁶⁸For a given $[(1 - \tau_B) e^{rH}]^{s_{bi}}$ term, v_i is higher for more extreme preferences (i.e. individuals with s_i close to 0 or close to 1 generate higher utility than middle-of-the-road individuals). For a given $(1 - s_i)^{1 - s_i} s_i^{s_i}$ term, v_i is higher for bequest lovers iff $(1 - \tau_B) e^{rH} > 1$ (i.e. bequest lovers generate higher utility if τ_B small).

⁶⁹Saez and Stantcheva (2012) propose a new theory of optimal taxation using systematically such endogenous social welfare weights instead of standard social welfare maximization and show that they can be useful in a number of contexts.

for $\tau_B > 0$:

$$\tau_B > 0 \quad \text{if and only if} \quad \left(s + \frac{s}{s_b} \right) e^{(r-g)H} > 1$$

In the case where saving motives entirely come from utility for bequests (i.e. $s_b/s = 1$), the condition becomes $(1 + s)e^{(r-g)H} > 1$. In particular, if $r - g > 0$, as is generally the case in the real world, then we always have $\tau_B > 0$. In theory, in case g is sufficiently large as compared to r , then zero receivers would prefer a bequest subsidy. Intuitively, infinite growth corresponds to an infinitely small b_y , i.e. to an infinitely low benefit in terms of tax revenue. Note however that $r - g < 0$ would violate the transversality condition, i.e. an infinite horizon social planner (assuming such planners exist) should react by borrowing indefinitely against the resources of future generations. That is, with $r - g < 0$, our steady-state maximization problem could no longer be defined as the limit solution to an intertemporal maximization problem (see Appendix B).

A.3 Proof of Proposition 3 (alternative welfare weights) (section 4).

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

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

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

We again consider a budget balanced tax reform $d\tau_B > 0, d\tau_L < 0$, with: $d\tau_L = -\frac{b_y d\tau_B}{1 - \alpha} \left(1 - \frac{e_B \tau_B}{1 - \tau_B} \right)$.

The difference with the zero-receiver case is that the utility change dV_i created by the tax reform $d\tau_B, d\tau_L$ now includes a third 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 third 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 $db_{ti} = -e_B z_i b_t d\tau_B / (1 - \tau_B)$ (for a given normalized inheritance level z_i).

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

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

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:

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

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

$$dSWF = \frac{y_{Lt} d\tau_B \int_{z_i=z} V_{ci} \theta_i}{(1 - \tau_B)(1 - \alpha)} \left[(1 - (1 + e_B)\tau_B) b_y - (1 - \alpha - \tau + \tau_B b_y) s_{bz} - \frac{(1 + e_B + s_{bz})(1 - \tau_B) z b_y}{\theta_z} \right]$$

$$\text{with } s_{bz} = \frac{E(V_{ci} \theta_i s_{bi} | z_i = z)}{E(V_{ci} \theta_i | z_i = z)}, \text{ and } \theta_z = \frac{E(V_{ci} (1 + e_B + s_{bi}) \theta_i | z_i = z)}{E(V_{ci} (1 + e_B + s_{bi}) | z_i = z)}$$

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 case) and with any ergodic random process for taste and productivity shocks (and not only with i.i.d. shocks). In the case with Cobb-Douglas utility functions, there exists a simpler proof. See Appendix A2.

Note 2. In the general case, s_{bz} is the average of s_{bi} over all p_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 p_z -bequest receivers, weighted by the product of their marginal utility V_{ci} and of their bequest taste s_{bi} . In case $s_{bi} \perp V_{ci} \theta_i$, then s_{bz} is the simple average of s_{bi} over all p_z -bequest-receivers, and θ_z is the simple average of θ_i over all p_z -bequest receivers : $s_{bz} = E(s_{bi} | z_i = z)$ and $\theta_z = E(\theta_i | z_i = z)$. In the case with i.i.d. productivity shocks, then $\theta_z = 1$. In the case with i.i.d. productivity and taste shocks and adequate utility normalization (see Appendix A2), then s_{bz} is the same as the average bequest taste for the entire population: $s_{bz} = s_b = E(s_{bi})$.

Note 3. If $\theta_z = 1$ and $s_{bz} = s_b$, the by substituting $b_y = \frac{s(1 - \tau - \alpha)e^{(r-g)H}}{1 - se^{(r-g)H}}$ into the τ_B formula, we obtain: $\tau_B = \frac{1 - e^{-(r-g)H}(1 - se^{(r-g)H})s_b/s - (1 + e_B + s_b)z}{(1 + e_B + s_b)(1 - z)}$. It follows that $\tau_B > 0$ iff $z < z^*$ with:

$$z^* = \frac{1 - e^{-(r-g)H}(1 - se^{(r-g)H})s_b/s}{1 + e_B + s_b} < 1$$

Note 4. The derivation of τ_B presented above implicitly neglects the fact that the normalized inheritance z of p_z -receivers might change as a consequence of the marginal tax change, because of the induced changes in the steady-state distribution $\Psi(z, \theta)$. E.g. assume $p_z = 0.1$, i.e. we are trying to maximize the steady-state welfare of individuals standing at the 10th percentile of

the inheritance distribution. It could be that a marginal tax change $d\tau_B > 0, d\tau_L < 0$ induces a marginal reduction in the inequality of inheritance, so that the normalized inheritance z at the 10th percentile rises by $dz > 0$. Define e_{p_z} the elasticity of the p_z percentile normalized inheritance z with respect to τ_B (along a budget balanced path): $e_{p_z} = [dz/z]/[d\tau_B/(1 - \tau_B)]$. Note that because z averages to one in the population, the (z -weighted) average of e_{p_z} in the population is zero.

It is reasonable to expect e_{p_z} to be positive for low p_z and negative for high p_z : higher bequest taxes (and hence lower labor taxes) are likely to reduce steady-state inequality, i.e. to raise z for low p_z and reduce z for high p_z . One can see that this dz effect introduces an extra term in $dSWF$. Namely one needs to add $V_{cz}(1 - \tau_B)(1 + R)b_t dz = e_{p_z} V_{cz} z b_y y_t d\tau_B$ in the above equation for $dSWF$. In effect, one simply needs to replace $(1 + e_B + s_{bz})$ by $(1 + e_B - e_{p_z} + s_{bz})$ in the third term in the square bracket above. So the corrected optimal tax formula is the following:

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

If $e_{p_z} > 0$, then this dz effect raises τ_B (and conversely if $e_{p_z} < 0$), which is intuitive: a positive dz effect makes bequest taxation even more desirable (and conversely).

In practice, this dz effect seems unlikely to be large. First, note that $e_{p_z} = 0$ for $p_z = 0$. That is, there is always a positive density of zero receivers (thanks to assumption 1), so for $p_z = 0$ we always have $z = 0$, independently of the tax policy. So this dz effect can be entirely ignored when we are interested in the zero-receiver optimum (Proposition 2). More generally, empirical evidence suggests that endogenous distribution effects are relatively small - 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.

Take for instance the model with binomial random taste. The steady-state distribution $\varphi(z)$ looks as follows:

$$z = z_k = \frac{1 - \mu}{\mu - p} \cdot \left[\left(\frac{\mu}{p} \right)^k - 1 \right] \text{ with probability } (1 - p) \cdot p^k \text{ (with } \mu = s(1 - \tau_B)e^{(r-g)H})$$

So if $p_z \leq 1 - p$, then $e_{p_z} = 0$.

E.g. if $1 - p = 0.5$ (i.e. the bottom 50% successors always receive zero bequests), then as long as we care only about the bottom 50% the dz effect can be ignored.

If $1 - p < p_z \leq (1 - p)(1 + p)$, then $e_{p_z} = \frac{dz_1/z_1}{d\tau_B/(1 - \tau_B)} = \frac{\mu}{1 - \mu} > 0$. I.e. the dz effect raises the optimal tax rate.

$$\text{If } (1 - p)(1 + .. + p^{k-1}) < p_z \leq (1 - p)(1 + ... + p^k), \text{ then } e_{p_z} = \frac{dz_k/z_k}{d\tau_B/(1 - \tau_B)}$$

One can easily see that $e_{p_z} > 0$ for k small enough (p_z small enough) and $e_{p_z} < 0$ for k large enough (p_z large enough). I.e. the dz effect raises the optimal tax rate as long as we care about bottom receivers, and reduces it if we care about top receivers.

Note 5. The optimal tax formula can be extended to the case $\Gamma > 0$, and to any welfare weights combination $(\omega_{p_z p_\theta})$. I.e. summing up dV_i over the entire distribution $\Psi(z, \theta)$, we have:

$$dSWF = \frac{y_{Lt} d\tau_B \int V_{ci} \theta_i V_i^{-\Gamma}}{(1 - \tau_B)(1 - \alpha)} \left[(1 - (1 + e_B)\tau_B) b_y - (1 - \alpha - \tau + \tau_B b_y) \bar{s}_b - \frac{(1 + e_B + \bar{s}_b)(1 - \tau_B) \bar{z} b_y}{\bar{\theta}} \right]$$

with:

$$\bar{s}_b = \frac{E(\omega_{p_z p_\theta} V_{ci} \theta_i s_{bi} V_i^{-\Gamma} | z_i \geq 0, \theta_0 \leq \theta_i \leq \theta_1)}{E(\omega_{p_z p_\theta} V_{ci} \theta_i V_i^{-\Gamma} | z_i \geq 0, \theta_0 \leq \theta_i \leq \theta_1)},$$

$$\bar{\theta} = \frac{E(\omega_{p_z p_\theta} V_{ci} \theta_i (1 + e_B + s_{bi}) V_i^{-\Gamma} | z_i \geq 0, \theta_0 \leq \theta_i \leq \theta_1)}{E(\omega_{p_z p_\theta} V_{ci} (1 + e_B + s_{bi}) V_i^{-\Gamma} | z_i \geq 0, \theta_0 \leq \theta_i \leq \theta_1)},$$

$$\bar{z} = \frac{E(\omega_{p_z p_\theta} V_{ci} z_i (1 + e_B + s_{bi}) V_i^{-\Gamma} | z_i \geq 0, \theta_0 \leq \theta_i \leq \theta_1)}{E(\omega_{p_z p_\theta} V_{ci} (1 + e_B + s_{bi}) V_i^{-\Gamma} | z_i \geq 0, \theta_0 \leq \theta_i \leq \theta_1)}.$$

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

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

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

A.4 Proof of Corollary 1 (Section 4)

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:

$$(1 - \alpha - \tau) s_{bz} / b_y = e^{-(r-g)H} \nu_z x_z / \theta_z - s_{bz} [\tau_B + (1 - \tau_B) z / \theta_z]$$

with:

$$s_{bz} = \frac{E(V_{ci} \theta_i s_{bi} | z_i = z)}{E(V_{ci} \theta_i | z_i = z)}, \quad \theta_z = \frac{E(V_{ci} s_{bi} \theta_i | z_i = z)}{E(V_{ci} s_{bi} | z_i = z)},$$

$$\nu_z x_z = \frac{E(V_{ci} \nu_i x_i | z_i = z)}{E(V_{ci} \theta_i | z_i = z)} \cdot \frac{E(V_{ci} \theta_i s_{bi} | z_i = z)}{E(V_{ci} s_{bi} | z_i = z)}$$

(ii) Alternatively, one can return to the equation $dV_i = -V_{ci}y_{Lti}d\tau_L - V_{bi}(1+R)b_{t+1i}d\tau_B - V_{ci}(1+R)b_{ti}(1+e_B)d\tau_B$. By substituting $b_{t+1i} = x_i b_{t+1} = x_i e^{gH} b_t$ and $y_{Lti}d\tau_L = -b_t e^{rH}(1 - \frac{e_B \tau_B}{1 - \tau_B})d\tau_B$, 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 $s_{bz} = s \cdot x_z \cdot \nu_z / \pi_z$.⁷⁰ 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).⁷¹ 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.

A.5 Proof of Proposition 4 (non-linear taxes) (section 4).

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:

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

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

⁷⁰With: $\pi_z = E(\tilde{y}_i | z_i = z) / \tilde{y}_t =$ 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).

⁷¹ $s = b_{t+1} / \tilde{y}_t =$ aggregate steady-state saving rate (bequests/lifetime resources). In the no-taste-memory special case, $\pi_z = E(\pi_i | z_i = z)$ (with $\pi_i = \tilde{y}_i / \tilde{y}_t$) = 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 $V_{ci} \theta_i$.

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:

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

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

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

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

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 $V_{ti} = V(c_{ti}, w_{ti}, (1 + R)(b_{t+1i} - \tau_B(b_{t+1i} - b_{t+1}^*)^+))$. Using the envelope theorem, a change in $d\tau_B$ keeping \tilde{y}_{ti} constant leads to a utility loss equal to $-(1 + R)V_{\tilde{b}_i}(b_{t+1i} - b_{t+1}^*)^+ d\tau_B$. 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+1}^* , the first-order condition is $V_{ci} = V_{wi} + (1 - \tau_B)(1 + R)V_{\tilde{b}_i}$, and one can again define $s_i = b_{t+1i}/\tilde{y}_{ti}$ 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 $s_{bi} = s_i(1 - \tau_B)(1 + R)V_{\tilde{b}_i}/V_{ci}$. Hence, we have:

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

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

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

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

Introducing

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

We have:

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

Setting $dSWF = 0$, we get:

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

Note. With a nonlinear estate tax, there is no closed form solution for b_{t+1i} as a function of lifetime resources and s_i . In particular, s_{b0}^* is no longer a weighted average of the individual s_{bi} . Numerical simulations would be required to provide a calibration in that context that we leave for future research.

A.6 Idiosyncratic Returns with Moral Hazard (Section 5.3)

In order to make the problem non trivial (and more realistic), we introduce moral hazard in the model with idiosyncratic returns, 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.

These efforts should be viewed as informal financial services that are directly supplied and consumed by households. Unlike the formal financial services supplied by financial corporations, these informal financial services are ignored by national accounts - which implies that pure capital income tends to be over-estimated.⁷²

The parameter ξ measures the extent to which rates of return are responsive to such efforts. 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 $C(e_{ti}) = (1 - \tilde{\tau}_B)b_{ti}c(e_{ti})$, where $(1 - \tilde{\tau}_B)b_{ti}$ is portfolio size (net-of-tax bequest) and $c(e_{ti})$ is a convex, increasing function of effort.⁷³

⁷²In order to compute the value of these services (and deduct it from conventionally measured capital income), one could try to estimate the amount of time that households spend in portfolio management and put a price on this time. Note that this is unlikely to reduce drastically the conventionally measured capital share (say, 25%-30% of national income, see e.g. Piketty (2010, Table A8)). For instance, the share of the formal financial sector has been fluctuating around 5%-7% of national income historically (see e.g. Philippon (2011, figure 1)). It is hard to imagine how unmeasured, informal financial services supplied by households represent more than a fraction of the formal finance industry - say 2%-3% of national income at the very most. Even if the overall volume of such services is limited, it could be however that they respond a lot to incentives, i.e. that the elasticity is significant. See discussion below.

⁷³It 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; e.g. very large capital endowments held by universities tend to generate higher net returns than smaller endowments). We leave this important issue to future research.

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:

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

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 $1 - \tau_K$.⁷⁴ 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.⁷⁵

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 all ($\tau_K = 0\%$). Unfortunately, there exists no simple closed-form formula for the intermediate case, so one needs to use numerical solutions methods in order to calibrate the optimal tax rate.

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 $\tilde{\tau}_B$, a capital income tax τ_K and a labor income tax τ_L such that:*

- (a) If $e_R \rightarrow 0$, then $\tau_K \rightarrow 100\%$, $\tilde{\tau}_B \rightarrow \tilde{\tau}_{B0} = \tau_B - \frac{R}{1 + R} < \tau_B$ and $\tau_L \rightarrow \frac{\tau - \tau_B b_y}{1 - \alpha}$ (with $\tau_B = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}}$)
- (b) If e_R is sufficiently small, then $\tau_K > \tau_L$; if e_R is sufficiently large, then $\tau_K < \tau_L$
- (c) There exists $\bar{e}_R > 0$ s.t. if $e_R \rightarrow \bar{e}_R$, then $\tau_K \rightarrow 0\%$, $\tilde{\tau}_B \rightarrow \tau_B$ and $\tau_L \rightarrow \frac{\tau - \tau_B b_y}{1 - \alpha} > \tau_K$

Proof. 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

⁷⁴Alternatively, 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). This complicates the analysis and brings little additional insight.

⁷⁵The 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 $R = R_0 + \xi^{1+1/\eta}(1 - \tau_K)^{1/\eta}$.

applies to $e_R = \bar{e}_R$ and $\tau_K = 0\%$. Note that \bar{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

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

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

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

Using the definition of $\nu_i = V_b^i (1 - \tau_B) (1 + (R - \varepsilon_0)(1 - \tau_K) - c(R)) / V_c^i$, we have

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

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 :

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

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

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

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

Example. 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 $\tau_L = 38\%$.

B Extensions

B.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.

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

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

where l is labor supply and $h_i(\cdot)$ 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 $v_t = v_0 e^{gHt}$ is the average wage rate of generation t .⁷⁶ 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:

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

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:

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

Hence (uncompensated) labor supply $l_i = l(v_{ti}, \bar{b}_{ti})$ is a function of the net-wage and non-labor income and is 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:

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

which can be rewritten as:

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

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-receiver optimum with elastic labor supply). *Under adapted assumptions 1-4, and welfare weights: $\omega_{p_z p_\theta} = 1$ if $p_z = 0$, and $\omega_{p_z p_\theta} = 0$ if $p_z > 0$:*

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

⁷⁶As discussed above $v_t = F_L = v_0(1 + G)^t$ grows at rate $1 + G$ per generation.

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

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

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

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

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

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

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

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.⁷⁷

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

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

We have:

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

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

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

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

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

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

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

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

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

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

⁷⁷With inelastic labor supply, we could use actual domestic output Y_t which was independent of taxes.

For zero receivers, we have $b_{ti} = 0$, $\tilde{y}_{ti} = y_{Lti}(1 - \tau_L)$ and hence:

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

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

$$s_{b0} = \frac{E[(V_c^i/V^i)y_{Lti}s_{bi} \mid z_i = 0]}{E[(V_c^i/V^i)y_{Lti} \mid z_i = 0]},$$

we obtain:

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

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

This formula is similar to the inelastic case except that e_L appears 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 $\tau_B = 1/(1 + e_B)$, we have $d\tau_L = 0$ for any small $d\tau_B$. Hence, the labor supply response becomes irrelevant.⁷⁸

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. If labor is more elastic than bequests, then behavioral responses reinforce the case for taxing labor income less than bequests. With $b_y = 15\%$ (current French level), for reasonable elasticity values $\tau_L < \tau_B$. Very large bequest elasticities—above one—and very small labor supply elasticities—close to zero—are needed to reverse this conclusion.

Example 7. 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 $\tau_L = 29\%$.

If $e_B = 1$ and $e_L = 0$, then $\tau_B = 35\%$ and $\tau_L = 35\%$.

If $e_B = 1$ and $e_L = 0.2$, then $\tau_B = 37\%$ and $\tau_L = 34\%$.

If $e_B = 1$ and $e_L = 1$, then $\tau_B = 43\%$ and $\tau_L = 33\%$.

⁷⁸This is analogous to the fact that $1/(1 + e_L)$ is the revenue maximizing rate in optimal linear labor income taxation even if there are income effects.

B.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 $1 + R = e^{rH}$). Our optimal tax results can easily be extended 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 $\beta_t = \frac{K_t}{Y_t} = b_{yt} e^{-r_t H}$ = domestic capital-output ratio.

This can be rewritten: $\frac{R_t}{1 + R_t} = \frac{\alpha}{b_{yt}}$. 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 1-4, then for any given tax policy (τ_B, τ_L) , we again have a unique long run steady-state: $b_{yt} \rightarrow b_y$, $R_t \rightarrow R$, $\Psi_t \rightarrow \Psi$ (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: $\frac{R}{1 + R} = \frac{\alpha}{b_y}$.

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.

B.3 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 $1 + N$ 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 $(1 + G)(1 + N) = e^{(g+n)H}$ 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 $Y_t = \beta^{\frac{\alpha}{1-\alpha}} L_t = \beta^{\frac{\alpha}{1-\alpha}} N_0 h_0 e^{(g+n)Ht}$ also grows at rate $(1+G)(1+N) = e^{(g+n)H}$ 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 $1+G = e^{gH}$.

The transition equation for $b_{yt} = \frac{e^{rH} B_t}{Y_t}$ (where $B_t = N_t \cdot b_t$ is the aggregate bequest flow received by generation t) becomes:

$$b_{yt+1} = e^{(r-g-n)H} [s(1-\tau_L)(1-\alpha) + s(1-\tau_B)b_{yt}] \quad (14)$$

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

Therefore, one simply needs to replace the productivity growth rate g by the sum of population and productivity growth rates $g+n$. In societies with infinitely large population growth (i.e. where individuals have an infinite number of children), inheritance becomes negligible. 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 $N < 0$).

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.

B.4 Consumption Taxes

So far we ignored we possibility of using a consumption tax τ_C in addition to the labor income tax τ_L and the capitalized bequest tax τ_B . Whether τ_C has a useful role to play in our model depends on which tax structures are allowed and how one models the impact of a consumption tax on private utility and government finances.

First of all, it is worth recalling that one of the main motivations behind Kaldor (1955)'s famous consumption tax proposal was to raise the share of the tax burden paid by wealthy successors. That is, Kaldor repeatedly stresses that there are many ways to avoid paying taxes on inheritance and especially on capital income (e.g. via trust funds and capital gains). He is very much concerned with the fact that the highly progressive income taxes applied in the U.K. in the 1950s might hurt high labor income earners (typically, civil servants and university professors such as himself) much more heavily than wealthy successors and rentiers. Kaldor therefore advocates for a steeply progressive tax on large consumption levels, with τ_C up to 75% for consumption levels over 5,000£ (i.e. living standards over about 10 times the average

income per tax unit of the time), which he views as easier to enforce administratively than a tax on large capital incomes or large wealth holdings.⁷⁹

One simple way to capture Kaldor's intuition in the context of our model is the following.⁸⁰ Assume that it is completely impossible to enforce a capitalized bequest tax τ_B , so that we are constrained to have $\tau_B = 0$. Then it is clearly optimal to have some positive level of consumption tax τ_C in addition to the labor income tax τ_L , since this is the only way to charge some of the tax burden to successors rather than to labor earners. E.g. in case there is no public revenue requirement ($\tau = 0$), then the only way to redistribute from successors to labor earners is to have a consumption tax $\tau_C > 0$ (taxing the consumption of both successors and labor earners), the proceeds of which are used to finance a wage subsidy $\tau_L < 0$.

In order to fully solve the model with a consumption tax, we first need to specify how τ_C enters in private utility for wealth and bequest. The most natural specification is to assume that agents care about the consumption value (purchasing power) of wealth and bequest, so that a consumption tax reduces proportionally the utility for wealth and bequest.⁸¹ The individual maximization programme in presence of a consumption tax τ_C can then be written as follows:

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

With: $\tilde{y}_{ti} = (1 - \tau_B)b_{ti}e^{rH} + (1 - \tau_L)y_{Lti} =$ total after-tax lifetime income

$c_{ti} =$ consumption

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

$\bar{w}_{ti} = (1 - \tau_C)w_{ti} =$ purchasing power of end-of-life wealth

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

$\tau_B =$ capitalized bequest tax rate, $\tau_L =$ labor income tax rate, $\tau_C =$ consumption tax rate

⁷⁹Kaldor formulates his consumption tax as $(1 + \tau_C)c_{ti}$ (rather than $c_{ti}/(1 - \tau_C)$), so his proposed top tax rate is actually $\tau_C = 300\%$ (rather than $\tau_C = 75\%$). Note that Kaldor is very much influenced by Morgenthau's (failed) attempt to introduce a progressive consumption tax in the U.S. in 1942, and views this new progressive tax as a complement to existing progressive inheritance and income taxes (not as a substitute). See in particular Kaldor (1955, pp.11-17, p.224-242). The argument according to which consumption taxes can play a useful role when capital taxes do not work very well (e.g. when capital income is badly measured) can also be found in Meade (1978) and King (1980).

⁸⁰Unfortunately Kaldor does not use a formal model, so it is hard to know the exact consumption tax theory that he has in mind. In addition to his tax enforcement rationale, another reason why he favors consumption taxes over income taxes seems to be the view that there is insufficient aggregate savings and that savings ought to be encouraged over consumption (this argument is never made fully explicit, however). We discussed this argument in the section on dynamic efficiency.

⁸¹In case agents care about the nominal value of wealth, irrespective of its consumption value (say because they care about the pure prestige or social status value of wealth), then the consumption tax does not reduce at all the utility for wealth and bequest, and this is obviously a pretty good policy tool (it raises revenue without reducing utility). This is the same tax illusion issue as that discussed for the capital income tax (see above).

One can see immediately that from the individual utility viewpoint, any tax mix with consumption tax (τ_C, τ_B, τ_L) is equivalent to a tax mix with zero consumption tax $(\bar{\tau}_C = 0, \bar{\tau}_B, \bar{\tau}_L)$, where the corrected tax rates $\bar{\tau}_B, \bar{\tau}_L$ are given by: $1 - \bar{\tau}_B = (1 - \tau_C)(1 - \tau_B)$ and $1 - \bar{\tau}_L = (1 - \tau_C)(1 - \tau_L)$. It is also equivalent to a tax mix with zero bequest taxes $(\bar{\bar{\tau}}_C, \bar{\bar{\tau}}_B = 0, \bar{\bar{\tau}}_L)$, where the corrected tax rates $\bar{\bar{\tau}}_C, \bar{\bar{\tau}}_L$ are given by: $1 - \bar{\bar{\tau}}_C = (1 - \tau_C)(1 - \tau_B)$ and $1 - \bar{\bar{\tau}}_L = (1 - \tau_L)/(1 - \tau_B)$.

From the government budget constraint viewpoint, these various tax arrangements are equivalent only if the consumption tax is pre-paid by donors, in the sense that it is paid on their total after-tax income, whether or not they consume it right away or transmit it to the next generation for future consumption. In this formulation, which we call “broad consumption tax” and note τ_C , the budget constraint is the following:

$$\tau_L y_{Lt} + \tau_B b_t e^{rH} + \tau_C [(1 - \tau_L) y_{Lt} + (1 - \tau_B) b_t e^{rH}] = \tau y_t$$

In case the consumption tax is paid only on current consumption, which we call “restricted consumption tax” and note $\hat{\tau}_C$, then the consequences for individual welfare are the same as with the broad tax, but the government budget constraint now looks as follows:

$$\tau_L y_{Lt} + \tau_B b_t e^{rH} + \hat{\tau}_C (1 - s) [(1 - \tau_L) y_{Lt} + (1 - \tau_B) b_t e^{rH}] = \tau y_t$$

In the broad consumption tax formulation τ_C , we have a full equivalence result.

E.g. assume $\tau = 30\%$, $\alpha = 30\%$, $s_{b0} = s = 10\%$, $e_B = 0$, $b_y = 15\%$, so that the zero-bequest-receiver optimum involves $\tau_B = 67\%$ and $\tau_L = 29\%$ (see example 1, section 4). An equivalent tax mix would be $\bar{\tau}_B = 0$, $\bar{\tau}_C = 67\%$ and $\bar{\tau}_L = -115\%$. I.e. instead of taxing bequests at 67% and labor income at 29%, then it is equivalent to tax all consumption expenditures at 67% (including those originating from labor income) and to subsidize labor income at -115% (so that labor earners end up with exactly the same after tax resources). This is fully equivalent, both from the viewpoint of individual welfare and government finances.

This is a fairly indirect way to implement the social optimum, however. In effect, a lot of money is being taxed and redistributed to the same people. So unless there exists a very strong tax enforcement argument in favor of consumption taxes over capital taxes (which we do not find very compelling, especially if one needs to implement progressive consumption taxes), a direct implementation via capital taxes seems more desirable.⁸²

⁸²Capital taxes do require substantial information, and are to some extent more complex to implement than proportional consumption taxes. But in order to implement a progressive consumption tax, one would need to measure individual consumption levels, which requires information on annual wealth holdings and/or capitalized inheritance, in which case it is easier to directly tax wealth or capitalized inheritance. In our view, the “tax enforcement” argument in favor of consumption taxes is often closer to a claim about “political feasibility” (wealth holders certainly do not like capital taxes and tend to strongly resist them) than to a claim about “administrative feasibility”. E.g. the recent Mirrlees Review (2011) argues that larger inheritance taxes might be desirable, but would generate substantial political opposition, and therefore chooses not to explore the issue any further.

In the restricted consumption tax formulation $\widehat{\tau}_C$, then in order to maintain budget balance one needs higher consumption tax rates. E.g. in the numerical example above, then with $s = 10\%$ one needs $\widehat{\tau}_C = \tau_B/(1 - s) = 74\%$. In terms of individual welfare, however, this is clearly inferior to the broad tax formulation. This is because the restricted tax imposes the same utility costs as the broad tax, but raises less revenue. In the same way, the difference between a tax τ_B on received capitalized bequest $b_t e^{rH}$ and a tax $\widehat{\tau}_B$ on left bequests b_{t+1} is that the former imposes the same utility costs on zero receivers but raises more tax revenue, so is superior in terms of welfare.⁸³

To summarize: unless one makes fairly ad hoc assumptions, consumption taxes are not very useful in the context of our model. In a world with two-dimensional heterogeneity -capitalized bequest vs labor income- the appropriate tax policy tools are a capitalized bequest tax τ_B and a labor income tax τ_L , not a consumption tax τ_C .

B.5 Homogenous Tastes

So far we assumed heterogenous random tastes (see section 3, Assumption 1). But strictly speaking, random tastes - or other types of multiplicative shocks - are not necessary for our results. I.e. if we assume no-taste-heterogeneity ($s_0 = s_1 = s$) and non-degenerate productivity heterogeneity ($\theta_0 < 1 < \theta_1$), then one can easily see that the steady-state distribution $\psi(z, \theta)$ involves only partial correlation between the two dimensions (the entire history $\theta_{ti}, \theta_{t-1i}$, etc. matters for z_{t+1i} , while θ_{ti} matters for θ_{t+1i}). All our results and tax formulas would go through, with two caveats. First, in order to ensure the existence of zero-bequest receivers one would need to assume zero minimal productivity ($\theta_0 = 0$), so that an infinitely long sequence of low productivity shocks leads to zero bequest.

E.g. assume uniform tastes $s_{it} = s$ and i.i.d. productivity shocks θ_{ti} . We again note $\mu = s(1 - \tau_B)e^{(r-g)H}$, and assume $\mu < 1$. The transition equation is: $z_{t+1i} = (1 - \mu)\theta_{ti} + \mu z_{ti}$.⁸⁴ This implies that in steady-state, z_{ti} is a (geometric) average of all past labor shocks: $z_{ti} = \sum_{s=0}^{+\infty} (1 - \mu)\mu^s \theta_{t-1-si}$. Hence, the steady-state inheritance distribution $\phi(z)$ is a continuous distribution over the interval $[\theta_0, \theta_1]$.

Next, as one can see from this simple example, one limitation of the pure productivity-shocks model is that it has little flexibility (the parameters for the distribution of inherited wealth are entirely determined by the parameters for the distribution of labor income) and tends to generate too little wealth inequality. The advantage of the model with random tastes (or other multiplicative shocks) is that it is more realistic and flexible. In particular, it can generate the right level of wealth concentration that one observes in the data (very low bottom shares, very

⁸³Assuming $r > g$, then in steady-state we have $b_t e^{rH} > b_{t+1} = b_t e^{gH}$, i.e. τ_B raises higher tax revenues than $\widehat{\tau}_B$. If one nevertheless decides to use the second form of tax, then the obvious conclusion is that the corresponding optimal tax rate should be lower, i.e. if $r > g$ then $\widehat{\tau}_B < \tau_B$.

⁸⁴This can be seen using the individual transition $b_{t+1} z_{t+1i} = s \cdot [z_{ti}(1 - \tau_B)b_t e^{rH} + \theta_{ti}(1 - \tau_L)y_{Lt}]$ and the macro equation $b_t e^{rH} = s e^{(r-g)H} (1 - \tau_L)y_{Lt} / (1 - s(1 - \tau_B)e^{(r-g)H})$.

high top shares) without assuming extreme values for the inequality of labor income. Given that our primary purpose is to obtain optimal tax formulas that can be calibrated to actual data, the random tastes model is clearly superior. But strictly speaking our results also apply to the pure productivity-shocks, uniform-taste model.

The only case where our results would cease to apply is if one assumes uniform taste and perfect correlation of productivity shocks across generations: i.e some dynasties i have for ever a low productivity shock ($\forall t, \theta_{ti} = \theta_0$), while some other dynasties j have for ever a high productivity shock ($\forall t, \theta_{tj} = \theta_1$), thereby violating the ergodicity assumption 2. In the long run, the distribution of inheritance $\phi(z)$ would then be perfectly correlated with the distribution of labor productivity $h(\theta)$. As a consequence, the labor income tax τ_L and the bequest tax τ_B would have the same distributional impact. Since the latter imposes an extra utility cost - via the usual joy-of-giving externality -, there is no point having a positive τ_B .⁸⁵ But as long as inequality is two-dimensional there is room for a two-dimensional tax policy tool.

B.6 Overlapping Generations and Life-cycle Savings

B.6.1 Model and Key Results

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 life-cycle 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 life-cycle 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 to compute the full optimum.

We assume the following deterministic, stationary, continuous-time OLG demographic structure.⁸⁶ 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 $I = 50$.

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

⁸⁵As shown by Kopczuk (2001) in the case with elastic labor supply, whether one wants to tax or subsidize bequests in the steady-state of a model with perfect correlation of abilities across generations and homogenous tastes actually hinges on the extent of the bequest externality (bequests received are a signal of ability so in some specifications one might want to tax them). See also Brunner and Pech (2011a, 2011b).

⁸⁶To 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; 2011, section 5).

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:

$$\max V_{ti} = V(U_{ti}, w_{tiD}, \bar{b}_{t+Hi}) \quad \text{subject to} \quad \tilde{c}_{ti} + w_{tiD} \leq \tilde{y}_{ti} = (1 - \tau_B)\tilde{b}_{ti} + (1 - \tau_L)\tilde{y}_{Lti}$$

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

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 $U_{ti} = [\int_{a=A}^{a=D} e^{-\delta(a-A)} c_{tia}^{\frac{1-\gamma}{1-\gamma}}]^{-\frac{1}{1-\gamma}}$, as well as for less standard (but maybe more realistic) utility specifications involving for instance consumption habit formation (see technical details subsection below). 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 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.⁸⁷ 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:

$$\tilde{y}_{Lti} = \theta_i \lambda (D - A) y_{Lt} e^{rH} \quad \text{with} \quad \lambda = \frac{e^{(r-g)(I-A)} - e^{-(r-g)(D-I)}}{(r-g)(D-A)} \quad (15)$$

⁸⁷In 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 accrue 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 (say, if $A = 20, H = 30, D = 80$, so that $I = D - H = 50$).⁸⁸

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

$$b_{t+Hi} = s_i[(1 - \tau_L)\tilde{y}_{Lti} + (1 - \tau_B)b_{ti}e^{rH}] \quad (16)$$

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

$$b_{t+H} = s[(1 - \tau_L)\lambda(D - A)y_{Lt}e^{rH} + (1 - \tau_B)b_t e^{rH}] \quad (17)$$

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

$$b_{yt+H} = e^{(r-g)H}[s(1 - \tau_L)\lambda(1 - \alpha) + s(1 - \tau_B)b_{yt}] \quad (18)$$

In case assumption 3 is satisfied, then $b_{yt} \rightarrow b_y = \frac{s(1 - \tau_L)\lambda(1 - \alpha)e^{(r-g)H}}{1 - s(1 - \tau_B)e^{(r-g)H}}$ as $t \rightarrow +\infty$. Hence, 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 the cross-sectional ratio plotted on Figures 4-5. The interesting point is that if $\lambda \simeq 1$, then the cross-sectional macroeconomic ratio is very close to the share of capitalized inheritance in total lifetime resources of the cohort inheriting at time t .

We impose a cross-sectional government budget constraint:

$$\tau_L Y_{Lt} + \tau_B B_t = \tau Y_t \quad \text{i.e. :} \quad \tau_L(1 - \alpha) + \tau_B b_y = \tau \quad (19)$$

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

$$b_y = \frac{s\lambda(1 - \tau - \alpha)e^{(r-g)H}}{1 - s[1 + (\lambda - 1)\tau_B]e^{(r-g)H}} \quad (20)$$

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 stimulates

⁸⁸For detailed empirical calibrations and theoretical extensions of the λ formula, see Piketty (2010, sections 5-7, and appendix E, tables E5-E10).

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, as in the discrete-time model.

Next, we obtain the same optimal bequest tax formula for the continuous-time model with overlapping generations and life-cycle savings as in the simplified discrete-time model where each generation leaves only one period. The proof is exactly the same as for Proposition 2, except that the time subscript t now denotes the time at which cohort \tilde{N}_t inherits.

Regarding optimal lifetime capital taxation, the key difference is that with life-cycle savings we now have an extra distortion. That is, positive tax rates on capital income $\tau_K > 0$ distort the intertemporal allocation of consumption $(c_{tia})_{A \leq a \leq D}$ 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 as σ is relatively small, the impact on our optimal capital tax results should be moderate. Unfortunately there does not 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 such a setting, one might also want to 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.

B.6.2 Technical Details on overlapping generations, continuous-time model

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} :

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

with $\delta =$ rate of time preference.

$\gamma =$ elasticity of marginal utility of consumption (=coefficient of relative risk aversion)

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

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

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 U of this consumption path is given by:

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

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

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:

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

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

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

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

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

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

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

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 τ_B and τ_L , so in effect it cancels out from the first-order condition for optimal tax rates.⁸⁹

⁸⁹ μ depends on r and hence would depend on the annual capital income tax rate τ_K when $\tau_K > 0$ making the analysis of τ_K more complex.

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 1 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 Carroll 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 can be written: $U = \mu \tilde{c}_{ti}$, with $\mu = \frac{\mu_c}{\tilde{\mu}}$ 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.7 Uninsurable Aggregate Shocks to Rates of Return

So far we assumed no aggregate uncertainty. It would be interesting to extend our results to a setting with aggregate, uninsurable uncertainty about the future rate of return (by definition, uncertainty at the level of the world rate of return 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 $b_{y2} > b_{y1}$. 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 the 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.

B.8 Endogenous Growth and Credit Constraints

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).⁹⁰ 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. From a theoretical perspective, anything could happen: depending on how one models endogenous growth and the accumulation of the growth-inducing production factor, various tax structures putting different emphasis on labor vs capital vs consumption taxes could be optimal.⁹¹ However basic cross-country evidence does not seem to bring much support to the view according to which tax policies entail systematic

⁹⁰See Piketty (1997) for a similar steady-state multiplicity.

⁹¹See e.g. Milesi-Ferretti and Roubini (1998).

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 more realistic way to proceed would be to keep growth exogenous, and to introduce the impact of borrowing constraints and inheritance on investment, output and income levels. We leave this to future research.

B.9 Tax Competition

So far 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 countries (i.e. they cannot change their residence), and that each country is able to enforce the residence principle of taxation (i.e. if they move their assets 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 entails no welfare cost: welfare maximizing governments should want to remove capital taxes anyway. Tax competition can even be a way to force inefficient governments to implement the optimal policy. 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.

For instance, in our baseline estimates with $\tau = 30\%$, $\alpha = 30\%$, $s_{b0} = 10\%$, and $b_y = 15\%$, the social optimum from the viewpoint of the bottom 50% typically involves a tax rate $\tau_B \simeq 60\%$ and $\tau_L \simeq 30\%$.⁹² With full capital mobility and tax competition, all capital taxes would be driven to $\tau_B = \tau_K = 0\%$, so labor taxes would have to be $\tau_L = \frac{\tau}{1 - \alpha} = 43\%$. So the net-of-tax-income of zero bequest receivers would fall by about 20%. Taking into account the utility gain from the zero bequest tax, and including labor supply and bequest elasticities e_L

⁹²See example 4 with $x_z = 50\%$, $e_B = 0.2$; note that with $e_L > 0$, τ_L would be even smaller and τ_B even larger; see example 7.

and e_B into the computations, we find total welfare losses for bottom 50% successors around 15%-25% - depending on parameters. These calibrations need to be refined. But they illustrate that the costs of tax competition in terms of social welfare can be substantial. This stands in sharp contrast to models where positive capital taxes come solely from lack of government commitment, in which case tax competition can only bring welfare gains.

C Dynamic Efficiency and Intergenerational Redistribution

Our optimal tax results can be extended in order to analyze the interaction between optimal capital taxation and the so-called dynamic efficiency issue (i.e. the issue of optimal aggregate capital accumulation). The main results and conclusions arising from these extensions are summarized in the main text of the paper (see section 6). Here we provide the formal statements and proofs.

Our basic model imposed a period-by-period (i.e. generation-by-generation) government budget constraint. That is, we assumed that each cohort pays in taxes exactly what they receive in public spending, so that the government cannot accumulate assets nor liabilities. This implies in particular that the government cannot directly affect the aggregate level of capital accumulation in the economy, and hence cannot address the so-called “dynamic efficiency” issue. In this Appendix C, we show that our results go through even when we relax these assumptions and allow the government to accumulate assets or liabilities. That is, we show 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 $\delta =$ social rate of time preference and $\Gamma =$ concavity of social welfare function),⁹³ then the government should accumulate infinite assets in order to have zero taxes or maximal subsidies 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 with two minor modifications (see proposition C3 below). First, s_{b0} is replaced by $s_{b0}e^{\delta'H}$ in the optimal τ_B formula with $\delta' = \delta + (\Gamma - 1)g$. This correction appears because τ_{Bt} hurts bequests leavers from generation $t - 1$ while revenue accrues in generation t . With no social discounting $\delta = 0$ and log-utility $\Gamma = 1$, there is no

⁹³With 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 below).

correction. Second, the formula for τ_L has to be adjusted for the interest receipt or payment term if the government has assets or debts at the optimum.

The decoupling of optimal capital accumulation vs. optimal labor/capital income tax mix is important, because both issues have sometimes 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.⁹⁴ 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 in a definite situation of excessive or insufficient aggregate capital accumulation, there would exist other and more efficient policy tools to address the problem than the capital vs. labor tax mix. Namely, the government should accumulate assets or liabilities (depending on whether there is too little or too much capital accumulation to start with), with no effect on optimal capital vs. labor tax formulas.⁹⁵

C.1 Intertemporal social welfare function

We first need to properly specify the intertemporal social welfare function. Throughout the paper, we study a steady-state social welfare maximization problem. That is, we assume that the government attempts to maximize the following, steady-state social welfare function (see section 3):

$$SWF = \int \int_{z \geq 0, \theta \geq 0} \omega_{p_z p_\theta} \frac{V_{z\theta}^{1-\Gamma}}{1-\Gamma} dz d\theta \quad (21)$$

With:

$V_{z\theta} = E(V_i \mid z_i = z, \theta_i = \theta)$ = average steady-state utility level V_i attained by individuals i with normalized inheritance $z_i = z$ and productivity $\theta_i = \theta$

$\omega_{p_z p_\theta}$ = social welfare weights as a function of the percentile ranks p_z, p_θ in the steady-state distribution of normalized inheritance z and productivity θ

Γ = concavity of social welfare function ($\Gamma \geq 0$)⁹⁶

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

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

⁹⁴See the discussion above about Kaldor (1955).

⁹⁵Those issues have been addressed by King (1980) in the standard OLG model. On the equivalence between the steady-state tax optimum and the full dynamic tax optimum with inter-temporal maximization, see also Atkinson and Sandmo (1980). In a different context, Stiglitz (1978) also stresses the idea that the government can use other policy tools (such as debt policy) in order to undo the potentially negative impact of estate taxation on aggregate capital accumulation. In contrast, Bourguignon (1981) ties in the issues of optimal wealth distribution and optimal aggregate capital accumulation. But with additional policy tools both issues can be untied.

⁹⁶If $\Gamma = 1$, then $SWF = \int \int_{z \geq 0, \theta \geq 0} \omega_{z\theta} \log(V_{z\theta}) d\Psi(z, \theta)$.

With: $1 + \Delta = e^{\delta H}$ = social rate of time preference (social discount rate)⁹⁷

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

$$\tilde{V}_t = \int \int_{z \geq 0, \theta \geq 0} \omega_{tp_z p_\theta} \frac{V_{tz\theta}^{1-\Gamma}}{1-\Gamma} dz d\theta$$

With: $V_{tz\theta} = E(V_{ti} | z_{ti} = z, \theta_{ti} = \theta)$ = average utility level attained at time t by individuals i with normalized inheritance $z_{ti} = z$ and productivity $\theta_{ti} = \theta$

And: $V_{ti} = \max V_i(c_{ti}, w_{ti}, \bar{b}_{ti})$ s.t. $c_{ti} + w_{ti} \leq \tilde{y}_{ti} = (1 - \tau_{Bt})z_{ti}b_t e^{rH} + (1 - \tau_{Lt})\theta_{ti}y_{Lt}$

To keep notations tractable, we focus upon the simple case with Cobb-Douglas utility functions and i.i.d. taste and productivity shocks (so that $e_B = 0$). All results can be extended to the general case with any family of utility functions that are homogenous of degree one, and with any ergodic random process for taste and productivity shocks (so that e_B can take any value).⁹⁸

C.2 Convergence of the intertemporal social welfare function

This intertemporal social welfare function might not be well defined, i.e. the intertemporal sum $SWF = \sum_{t \geq 0} \tilde{V}_t e^{-\delta H t}$ might be infinite. In order to ensure that the sum converges, we need to put constraints on parameters.

First, note that for any z, θ , the average utility level $V_{tz\theta}$ grows at the same rate as per capita output y_t as $t \rightarrow +\infty$, i.e. at generational rate $1 + G = e^{gH}$. Namely, with Cobb-Douglas utility $V_i(c, w, \bar{b}) = c^{1-s_i} w^{s_{wi}} \bar{b}^{s_{bi}}$, and with i.i.d. taste and productivity shocks, we have:

$$V_{tz\theta} = v_t \cdot \tilde{y}_{tz\theta}$$

With: $v_t = E(v_{ti})$, $v_{ti} = (1 - s_i)^{1-s_i} s_i^{s_i} [(1 - \tau_{Bt})e^{rH}]^{s_{bi}}$, and $\tilde{y}_{tz\theta} = (1 - \tau_B)z b_t e^{rH} + (1 - \tau_L)\theta y_{Lt}$.⁹⁹

As $t \rightarrow +\infty$, under assumptions 1-3, and assuming that the tax policy sequence τ_{Bt}, τ_{Lt} converges towards some τ_B, τ_L , then $v_t \rightarrow v = E(v_i)$, and $b_{yt} = b_t e^{rH} / y_t \rightarrow b_y = \frac{s(1 - \tau_L)(1 - \alpha)e^{(r-g-n)H}}{1 - s(1 - \tau_B)e^{(r-g-n)H}}$.

It follows that after-tax income $\tilde{y}_{tz\theta} \rightarrow q_{z\theta} \cdot y_t$, with $q_{z\theta} = (1 - \tau_B)b_y \cdot z + (1 - \tau_L)(1 - \alpha) \cdot \theta$. I.e. for any z, θ , after-tax income $\tilde{y}_{tz\theta}$ grows proportionally to per capita output $y_t = Y_t / N_t = y_0 e^{gH t}$.

It also follows that $V_{tz\theta} \rightarrow v \cdot q_{z\theta} \cdot y_t$ grows at instantaneous rate g (i.e. at generational rate $1 + G = e^{gH}$) in the long-run.

⁹⁷In the same way as for productivity 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.

⁹⁸All results can also be easily extended to the case with utility normalization. See Appendix A2.

⁹⁹See Appendix A2.

Since utilities are proportional to incomes, the parameter $\Gamma \geq 0$ can be viewed as a parameter measuring the concavity of the social planner's preferences with respect to income (it is also equal to the constant coefficient of relative risk aversion).

In case $\Gamma = 0$, the social planner does not care at all about inequality (linear social welfare), so redistribution is useless.

In case $\Gamma = 1$, the social planner has a moderate concern for inequality (logarithmic social welfare, i.e. unitary coefficient of relative risk aversion).¹⁰⁰

In case $\Gamma > 1$, the social planner has a large concern for inequality. With $\Gamma > 1$, social welfare is bounded above, i.e. even infinitely rich agents in a given cohort or infinitely rich future cohorts generate finite social welfare.

As $\Gamma \rightarrow +\infty$, the social planner becomes infinitely inequality averse, both in the cross-section (as long as the poor are poorer than the rich, transferring one unit of income from the latter to the former hugely raises total social welfare - even if a large fraction of the one unit is lost in the process) and in the long-run (as long as today's generations are poorer than future generations, transferring one unit of income from the latter to the former hugely raises total social welfare - even if a large fraction of one unit is lost in the process). This corresponds to Rawlsian (or maximin) social welfare.

Next, a natural constraint to put on welfare weights $\omega_{tpzp\theta}$ is that their sum grows at rate $(1 - \Gamma')n$, where n is the instantaneous, exogenous population growth rate (i.e. $N_t = N_0 e^{nHt}$, with n possibly equal to 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.¹⁰¹ That is, we assume that $\omega_{tpzp\theta} = \omega_t \cdot \omega_{pzp\theta}$, with:

$$\int \int_{z \geq 0, \theta \geq 0} \omega_{pzp\theta} dz d\theta = 1 \text{ and } \omega_t = N_t^{1-\Gamma'} = N_0^{1-\Gamma'} e^{(1-\Gamma')nHt}$$

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. Therefore larger cohorts matter more in terms of social impact. This is sometime called the "Benthamite" case in the normative economics 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. The planner cares only about average welfare of each cohort

¹⁰⁰One limitation of this standard formulation of intertemporal social welfare is that the social planner is bound to have the same concern for cross-sectional and intertemporal inequality. See discussion below.

¹⁰¹The reason we introduce population growth here is because it plays an important role in the analysis of dynamic efficiency and socially optimal capital accumulation. Of course everything also holds in the special case with $n = 0$ and population normalized to 1 (i.e. $N_t = N_0 = 1$, so that aggregate and per capita variables are the same: $Y_t = N_t \cdot y_t = y_t$, $B_t = N_t \cdot b_t = b_t$, etc.). For simplicity, population growth is assumed to be exogenous and to be neutral with respect to saving behavior, i.e. utility for bequest does not depend on the actual number of children (see the extension introduced in section 6).

(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. Both approaches have some merit - and so does the intermediate formulation with $\Gamma' \in [0, 1]$.¹⁰² 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 $\Gamma \geq 0$.

Our point here is simply that as $t \rightarrow +\infty$, the social welfare of generation t grows at instantaneous rate $(1 - \Gamma)g + (1 - \Gamma')n$:

$$\tilde{V}_t \rightarrow \frac{\tilde{v}^{1-\Gamma}}{1-\Gamma} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'} = \frac{\tilde{v}^{1-\Gamma}}{1-\Gamma} \cdot y_0^{1-\Gamma} \cdot N_0^{1-\Gamma'} \cdot e^{(1-\Gamma)gHt + (1-\Gamma')nHt}$$

With: $\tilde{v} = \left[\int \int_{z \geq 0, \theta_0 \leq \theta \leq \theta_1} \omega_{p_z p_\theta} \cdot (v \cdot q_{z\theta})^{1-\Gamma} \cdot dz d\theta \right]^{1/(1-\Gamma)}$

It follows that the intertemporal sum $SWF = \sum_{t \geq 0} \tilde{V}_t e^{-\delta Ht}$ is well defined (non-infinite) iff $\delta > (1 - \Gamma)g + (1 - \Gamma')n$, i.e. if and only if the following condition is satisfied:

Assumption 5 $\delta' = \delta - (1 - \Gamma)g - (1 - \Gamma')n > 0$

In what follows, we constantly assume that assumption 5 is satisfied (otherwise the intertemporal social welfare function would not be well defined). Intuitively, δ' can be viewed as the “modified” social discounted rate, i.e. the difference between the “raw” social discount rate δ and the growth rate of social welfare $(1 - \Gamma)g + (1 - \Gamma')n$. In case $\Gamma = \Gamma' = 0$ (linear social welfare function), then social welfare grows at rate $g + n$, so that $\delta' = \delta - g - n$, i.e. the intertemporal welfare sum is well defined iff $\delta > g + n$. In case $\Gamma = \Gamma' = 1$ (logarithmic social welfare function), then the sum is well-defined for any $\delta > 0$. In case $\Gamma > 1$ (bounded above social welfare) and $\Gamma' = 1$, then the sum is well defined even with $\delta = 0$.

C.3 Period-by-period government budget constraint

Throughout the paper, we take as given a fixed, exogenous public good requirement $G_t = \tau Y_t$ each period (with $\tau \geq 0$), and we assume the following period-by-period (i.e. generation-by-generation) budget constraint:

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

In Proposition 2 (and subsequent propositions), we solve for the stationary tax policy ($\tau_{Lt} = \tau_L, \tau_{Bt} = \tau_B$) $_{t \geq 0}$ maximizing steady-state social welfare.

¹⁰²On Benthamite vs. non-Benthamite social welfare functions ($\Gamma' = 0$ vs $\Gamma' = 1$), see e.g. Blanchard and Fischer (1989, Chapter 2, pp. 39-45, notes 4 and 13). One could also extend the normative framework by allowing for any $\Gamma' \geq 0$ (including $\Gamma' > 1$) by assuming $\omega_t = \frac{N_t^{1-\Gamma'}}{1-\Gamma'}$ (and $\omega_t = \log(N_t)$ in case $\Gamma' = 1$). Of course if n is close to zero, the choice of Γ' makes little difference.

We now assume that the social planner seeks to maximize the intertemporal sum $SWF = \sum_{t \geq 0} \tilde{V}_t e^{-\delta H t}$. We look for the tax policy sequence $(\tau_{Lt}, \tau_{Bt})_{t \geq 0}$ maximizing this intertemporal social welfare function, and we are particularly interested in the limit outcomes as $t \rightarrow +\infty$.

For simplicity we focus upon the zero-bequest receiver optimum ($\omega_{p_z p_\theta} = 1$ if $p_z = 0$, and $\omega_{p_z p_\theta} = 0$ if $p_z > 0$) (Proposition 2), but all results can be extended to arbitrary welfare weights (Proposition 3).

If we maintain the period-by-period budget constraint (i.e. at any time t , $\tau_{Lt} = \frac{\tau - \tau_{Bt} b_{yt}}{1 - \alpha}$), we have the following results.

First, if we allow for non-stationary tax policy sequence $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$, 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 is on the table and can be taxed at no efficiency cost, so the socially optimal tax policy sequence always involves $\tau_{B0} = 1$ and $\tau_{L0} = \frac{\tau - b_{y0}}{1 - \alpha}$. I.e. in the short run it is always tempting for zero receivers to impose confiscatory bequest taxes so as to subsidize labor income as much as possible (or to have labor taxes that are as moderate as possible, in case bequest tax revenues are insufficient to cover public spending, i.e. in case $b_{y0} < \tau$). In order to avoid time-inconsistency problems, we solve for the full-commitment optimum, i.e. we assume that the planner can commit to a sequence $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$ and stick to it for ever. The optimal sequence τ_{Bt}, τ_{Lt} always involves confiscatory (or quasi-confiscatory) bequest tax rates during the first time periods, and converges towards some long run, stationary tax policy τ_B, τ_L as $t \rightarrow +\infty$.

Next, and more interestingly, these asymptotic tax rates converge towards the steady-state welfare optimum as the modified social discount rate goes to zero.

Proposition 8 C1 (zero-bequest-receiver optimum, period-by-period budget constraint).

Under assumptions 1-5, with a period-by-period government budget constraint, the tax policy sequence $(\tau_{Lt}, \tau_{Bt})_{t \geq 0}$ maximizing intertemporal social welfare converges towards the steady-state welfare optimum as the corrected social discount factor goes to zero:

- (1) As $t \rightarrow +\infty$, $\tau_{Bt} \rightarrow \tau_B(\delta') = \frac{1 - (1 - \alpha - \tau) s_{b0} e^{\delta' H} / b_y}{1 + s_{b0} e^{\delta' H}}$ and $\tau_{Lt} \rightarrow \tau_L(\delta') = \frac{\tau - \tau_B(\delta') b_y}{1 - \alpha}$
with: $\delta' = \delta - (1 - \Gamma)g - (1 - \Gamma')n =$ modified social discount factor
- (2) As $\delta' \rightarrow 0$, $\tau_B(\delta') \rightarrow \tau_B^* = \frac{1 - (1 - \alpha - \tau) s_{b0} / b_y}{1 + s_{b0}}$ and $\tau_{Lt} \rightarrow \tau_L^* = \frac{\tau - \tau_B b_y}{1 - \alpha}$

Proof. The proof is essentially the same as Propositions 2 and is given in section C8 below.

The intuition behind this result is straightforward. As $\delta' \rightarrow 0$, the social planner puts approximately the same weight on the welfare of all future generations, so in effect the planner cares almost exclusively about the long run. Therefore the asymptotic, intertemporal welfare op-

timum becomes arbitrarily close to the the steady-state welfare optimum. E.g. this corresponds to the case of an infinitely patient, logarithmic planner, i.e. $\Gamma = \Gamma' = 1$ and $\delta \rightarrow 0$.

Conversely, if $\delta' \rightarrow +\infty$, then $\tau_B(\delta') \rightarrow -\frac{1-\alpha-\tau}{b_y}$ and $\tau_L(\delta') \rightarrow 1$. I.e. if the social planner puts infinite weight on the current generation, then the asymptotic, intertemporal welfare optimum involves a maximal bequest subsidy financed by a 100% labor income tax rate. Intuitively, in the extreme case with $\delta' = +\infty$, i.e. where the planner cares only about generation $t = 0$ and does not care at all about generation $t = 1$, then he/she will choose to move directly from $\tau_{B0} = 1, \tau_{L0} = \frac{\tau - b_{y0}}{1 - \alpha}$ to $\tau_{B1} = -\frac{1 - \alpha - \tau}{b_y}, \tau_{L1} = 1$. That is, from the viewpoint of the zero-bequest receivers of generation $t = 0$, it is optimal to tax received bequests at 100%, but to subsidize left bequests as much as possible. Note that $\delta' \rightarrow +\infty$ can arise either because $\delta \rightarrow +\infty$ (the planner is infinitely impatient) or because $\Gamma \rightarrow +\infty$ (the planner is infinitely concave with respect to income growth, i.e. he/she views future zero-bequest receivers as infinitely rich as compared to current zero-bequest receivers, which in effect makes him/her infinitely impatient).

C.4 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, generational world rate of return $1 + R = e^{rH}$. We again assume an exogenous public good requirement $G_t = \tau Y_t$ each period (with $\tau \geq 0$). With no loss in generality, we assume zero initial government assets ($A_0 = 0$). The intertemporal government budget constraint can be written as follows:

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

Denoting by $\bar{\tau}_t = \tau_{Lt}(1 - \alpha) + \tau_{Bt} b_{yt}$ the aggregate tax rate imposed on generation t , this can be rewritten as follows:

$$\sum_{t \geq 0} \bar{\tau}_t Y_t e^{-rHt} = \sum_{t \geq 0} \tau Y_t e^{-rHt}$$

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 6 $r > g + n$

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 (i.e. at some point the domestic economy would borrow so much that it 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: $A_{t+1} = (1 + R)A_t + (\bar{\tau}_t - \tau)Y_t$. The ratio between net government assets and domestic output $a_t = A_t/Y_t$ can be written as follows:

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

Take any tax policy sequence $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$ satisfying the intertemporal budget constraint and converging towards some asymptotic tax policy (τ_B, τ_L) as $t \rightarrow +\infty$. Under assumptions 1-6, $b_{yt} \rightarrow b_y$, and $\bar{\tau}_t \rightarrow \bar{\tau} = \tau_L(1 - \alpha) + \tau_B b_y$. Then two cases can happen:¹⁰³

(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 $t \rightarrow +\infty$, $a_t \rightarrow a \geq 0$. I.e. the government has a positive asset position in the long run.

(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 $t \rightarrow +\infty$, $a_t \rightarrow a \leq 0$. I.e. the government has a negative asset position in the long run.

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

$$\begin{aligned} \tau_L(1 - \alpha) + \tau_B b_y + \bar{R}a &= \bar{\tau} + \bar{R}a = \tau \\ \text{i.e.} \quad a &= \frac{\tau - \bar{\tau}}{\bar{R}} \end{aligned}$$

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 so as to make sure that assets keep up with economic growth (or equivalently the rate at which the government should reimburse its debt so as to avoid exploding debt).¹⁰⁴

Finally, note that in the long run private agents accumulate more private wealth when taxes are lower (i.e. when the government has accumulated higher public wealth) - and conversely. That is, with Cobb-Douglas utility and i.i.d. taste and productivity shocks, the aggregate transition equation looks as follows:

$$\begin{aligned} b_{yt+1} &= s(1 - \tau_{Lt})(1 - \alpha)e^{(r-g-n)H} + s(1 - \tau_{Bt})e^{(r-g-n)H}b_{yt} \\ \text{i.e.} \quad b_{yt+1} &= s(1 - \alpha - \bar{\tau}_t)e^{(r-g-n)H} + s \cdot e^{(r-g-n)H}b_{yt} \end{aligned}$$

¹⁰³Here we neglect exploding asset accumulation paths ($a_t \rightarrow +\infty$ or $a_t \rightarrow -\infty$), which in effect are ruled out by the assumptions $\bar{\tau}_t \geq 0$ and $\bar{\tau}_t \leq 1$ (see below).

¹⁰⁴In a continuous time model, this rate would simply be $\bar{r} = r - g - n$.

Therefore as $t \rightarrow +\infty$, $b_{yt} \rightarrow b_y = \frac{s(1 - \tau_L)(1 - \alpha)e^{(r-g-n)H}}{1 - s(1 - \tau_B)e^{(r-g-n)H}} = \frac{s(1 - \alpha - \bar{\tau})e^{(r-g-n)H}}{1 - s \cdot e^{(r-g-n)H}}$. I.e. b_y is a decreasing function of long run tax rates τ_B and τ_L (and of the long run aggregate tax rate $\bar{\tau}$). In the case with a period-by-period government budget constraint, the tax rate $\bar{\tau}_t$ was constrained to be equal to τ , so that b_y was fixed.

C.5 Open economy

The key question is the following: in the long run, will the government choose to accumulate positive assets or debt ($a > 0$ or $a < 0$), and how does this decision interact with the choice of an optimal tax mix τ_B, τ_L ?

In the open economy case, the answer depends entirely on whether the world rate of return r is smaller or larger than the so-called modified Golden rule rate of return $r^* = \delta + \Gamma'n + \Gamma g$.

Proposition 9 C2 (zero-bequest receiver intertemporal optimum, open economy). *Under assumptions 1-6, with an intertemporal government budget constraint and an open economy, the asset and tax policy sequence $(a_t, \tau_{Lt}, \tau_{Bt})_{t \geq 0}$ maximizing intertemporal social welfare depends on whether the world rate of return r is smaller or larger than the modified Golden rule rate of return $r^* = \delta + \Gamma'n + \Gamma g$:*

(1) *If $r < r^*$, then the social planner chooses to postpone tax payments to the long run (future generations) and to accumulate maximal public debt compatible with the financing of public good provision. That is, as $t \rightarrow +\infty$, $\tau_{Lt} \rightarrow 1$, $\tau_{Bt} \rightarrow 1$, $b_{yt} \rightarrow 0$, $\bar{\tau}_t \rightarrow 1 - \alpha$, and $a_t \rightarrow \underline{a} = -\frac{1 - \alpha - \tau}{\bar{R}} < 0$.*

(2) *If $r > r^*$, then the social planner chooses to have all tax payments in the short run (current or nearby generations) and to accumulate maximal public assets to finance public good provision. That is, as $t \rightarrow +\infty$, $\tau_{Lt} \rightarrow \underline{\tau}_L$, $\tau_{Bt} \rightarrow \underline{\tau}_B$, $b_{yt} \rightarrow \bar{b}_y$, $\bar{\tau}_t \rightarrow \underline{\tau} \leq 0$, and $a_t \rightarrow \bar{a} = \frac{\tau - \underline{\tau}}{\bar{R}} > 0$.*

(3) *In the knife-hedge case $r = r^*$, 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 $a \geq 0$ or $a \leq 0$, then the tax policy sequence $(\tau_{Lt}, \tau_{Bt})_{t \geq 0}$ maximizing intertemporal social welfare converges towards the steady-state welfare optimum as the corrected social discount factor goes to zero. That is:*

As $t \rightarrow +\infty$, $\tau_{Bt} \rightarrow \tau_B(\delta') = \frac{1 - (1 - \alpha - \tau)s_{b0}e^{\delta'H}/b_y}{1 + s_{b0}e^{\delta'H}}$ and $\tau_{Lt} \rightarrow \tau_L(\delta', a) = \frac{\tau - \tau_B(\delta')b_y - \bar{R} \cdot a}{1 - \alpha}$
(with: $\delta' = \delta - (1 - \Gamma)g - (1 - \Gamma')n =$ modified social discount factor).

As $\delta' \rightarrow 0$, $\tau_B(\delta') \rightarrow \tau_B^ = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + s_{b0}}$ and $\tau_{Lt} \rightarrow \tau_L^* = \frac{\tau - \tau_B^*b_y - \bar{R} \cdot a}{1 - \alpha}$*

Proof. The proof is given in section C9 below.

The intuition behind this result is the following.

In case r is sufficiently low, then it is worth borrowing in order to consume more now. More precisely, in case $r < r^*$, then the social planner can always raise intertemporal social welfare by shifting additional resources from future generations to current generations. So he/she will choose to postpone tax payments indefinitely, by having zero or negative taxes in the short run and by issuing public debt on international financial markets in order to finance public expenditures. In the long run, tax rates τ_{Lt} , τ_{Bt} will converge towards revenue-maximizing levels - which, in the simple model with zero labor supply and bequest elasticities, are simply equal to $\tau_L = 1$, $\tau_B = 1$. As a consequence there is no private wealth accumulation in the long run ($b_{yt} \rightarrow 0$, i.e. the domestic capital stock is entirely owned by foreigners, just like the public debt), and the aggregate tax rate converges towards $\bar{\tau} = 1 - \alpha$ (i.e. the labor share is taxed at 100%). By assumption 4, this is sufficient to cover public spending (i.e. $\bar{\tau} = 1 - \alpha > \tau$), and extra tax revenue $\bar{\tau} - \tau$ allows the government to finance its debt service and stabilize its (negative) asset position at $\underline{a} = -\frac{1 - \alpha - \tau}{\bar{R}} < 0$.

Conversely, in case r is sufficiently large, then it is worth investing in order to consume more later. More precisely, in case $r > r^*$, then the social planner can always raise intertemporal social welfare by shifting additional resources from current generations to future generations. So he/she will choose to have all tax payments in the short run and to accumulate sufficient public assets so as to finance public good provision in the long run. Tax rates τ_{Lt} , τ_{Bt} will converge towards their minimal levels $\underline{\tau}_L, \underline{\tau}_B$. If we put some exogenous constraints on these minimal levels, say $\underline{\tau}_L = \underline{\tau}_B = 0$ (no labor or bequest subsidy) or $\underline{\tau}_L = \underline{\tau}_B = -1$ (subsidy rates cannot be larger than 100%), then this determines the long run level of private wealth accumulation \bar{b}_y and aggregate tax rate $\underline{\tau}$. This in turn determines the long run positive asset position $\bar{a} = \frac{\tau - \underline{\tau}}{\bar{R}}$. So for instance if $\underline{\tau}_L = \underline{\tau}_B = \underline{\tau} = 0$, then $a = \frac{\tau}{\bar{R}}$.

Note that the only way we can get finite asset accumulation in the case $r > r^*$ is by putting some exogenous minimal constraints $\underline{\tau}_L, \underline{\tau}_B$. With no such constraint, the planner would accumulate infinite assets ($a_t \rightarrow +\infty$) so as to be able to distribute infinite subsidies ($\tau_{Lt} \rightarrow -\infty$, $\tau_{Bt} \rightarrow -\infty$). Private wealth accumulation would also follow an exploding path. That is, as $t \rightarrow +\infty$, $s(1 - \tau_{Bt})e^{(r-g-n)H} > 1$, and therefore $b_{yt} \rightarrow +\infty$. In effect, the economy accumulates both infinite public assets and infinite private assets, and would soon cease to be a small open economy any more.

To summarize: if $r < r^*$, then under the guidance of the social planner our small open economy will attempt to accumulate as much debt as possible; if $r > r^*$, then it will attempt to accumulate as much assets as possible. It is only in the knife-edge case $r = r^*$ (which is very unlikely to happen in the open economy case where r is exogenous) that we have a balanced social optimum with an interior asymptotic tax mix (τ_B, τ_L) . Note that the optimal tax mix that we obtain in this knife-edge case is exactly the same as in Proposition C1 - except of course for the $-\bar{R} \cdot a$ term now entering into the τ_L formula.¹⁰⁵

¹⁰⁵I.e. if $a > 0$ then τ_L is smaller than before (positive government assets allow for lower taxes in the long

C.6 Closed economy

We now turn to the most interesting case, namely the closed economy intertemporal optimum.

In the closed economy case, the domestic capital stock K_t is equal to the sum of private and government assets, i.e. $K_t = B_t + A_t$. At every period $t \geq 0$, the generational rate of return $1 + R_t = e^{r_t H}$ is equal to the marginal product of capital: $R_t = F_K$. With a Cobb-Douglas production function $F(K_t, L_t) = K_t^\alpha L_t^{1-\alpha}$, we have:

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

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

It is straightforward to show that in the closed economy intertemporal optimum, the social planner will accumulate assets until the point where the modified Golden Rule condition is satisfied: as $t \rightarrow +\infty$, $r_t \rightarrow r^* = \delta + \Gamma'n + \Gamma g$. (or, equivalently, $1 + R_t = e^{r_t H} \rightarrow 1 + R^* = e^{r^* H} = (1 + \Delta)(1 + N)^{\Gamma'}(1 + G)^{\Gamma}$). That is, the government will accumulate assets until the point where $\beta_t \rightarrow \beta^* = \frac{\alpha}{R^*}$.

To see why, note first that the long run rate of return cannot be below r^* . In case $r_t \rightarrow r < r^*$, then from Proposition C2 we know that the social planner will choose to accumulate maximal public debt ($a_t \rightarrow \underline{a} < 0$) and there will be no long run private wealth accumulation ($b_{yt} \rightarrow b_y = 0$). I.e. the long run domestic capital-output ratio is scheduled to be negative ($\beta_t \rightarrow \beta = \underline{a} < 0$), which is impossible: at some point β_t will be infinitely small, i.e. r_t will be infinitely large, thereby contradicting the assumption $r_t \rightarrow r < r^*$.

Conversely, in case $r_t \rightarrow r > r^*$, then from Proposition C2 we know that the social planner will choose to accumulate maximal public assets ($a_t \rightarrow \bar{a} > 0$) and private assets ($b_{yt} \rightarrow \bar{b}_y$). With no minimal constraints on τ_L, τ_B , then we obtain infinite capital accumulation ($\beta_t \rightarrow +\infty$), i.e. $r_t \rightarrow 0$, which contradicts $r_t \rightarrow r > r^*$. More generally, with exogenous minimal constraints on τ_L, τ_B , one simply needs to assume that the corresponding capital accumulation level $\bar{\beta} = \bar{b}_y e^{-r^* H} + \bar{a}$ is larger than β^* :

Assumption 7 $\bar{\beta} > \beta^*$

Under this assumption, the intertemporal social optimum necessarily involves $r_t \rightarrow r^*$ and $\beta_t \rightarrow \beta^*$. Following part (3) of Proposition C2, we then know that the optimal tax policy sequence converges towards the steady-state welfare optimum goes to zero. Therefore we have the following characterization of the full intertemporal optimum:

Proposition 10 C3 (zero-bequest-receiver intertemporal optimum, closed economy).

Under assumptions 1-7, with an intertemporal government budget constraint and a closed economy, the asset and tax policy sequence $(a_t, \tau_{Lt}, \tau_{Bt})_{t \geq 0}$ maximizing intertemporal social welfare can be characterized as follows:

run); if $a < 0$ then τ_L is larger than before (negative government assets, i.e. public debt, require higher long run taxes).

(1) First, the optimal government asset and debt policy is chosen so as to satisfy the modified Golden rule: $r_t \rightarrow r^* = \delta + \Gamma'n + \Gamma g$. The capital-output ratio converges towards the corresponding level: $\beta_t = b_{yt}e^{-r_t H} + a_t \rightarrow \beta^* = \frac{\alpha}{R^*} = \frac{\alpha}{e^{r^* H} - 1}$.

(ii) Next, the optimal tax policy sequence $(\tau_{Lt}, \tau_{Bt})_{t \geq 0}$ converges towards the steady-state welfare optimum as the corrected social discount factor goes to zero. That is:

$$\text{As } t \rightarrow +\infty, \tau_{Bt} \rightarrow \tau_B(\delta') = \frac{1 - (1 - \alpha - \tau)s_{b0}e^{\delta' H}/b_y}{1 + s_{b0}e^{\delta' H}} \text{ and } \tau_{Lt} \rightarrow \tau_L(\delta', a) = \frac{\tau - \tau_B(\delta')b_y - \bar{R} \cdot a}{1 - \alpha}$$

(with: $\delta' = \delta - (1 - \Gamma)g - (1 - \Gamma')n = \text{modified social discount factor}$).

$$\text{As } \delta' \rightarrow 0, \tau_B(\delta') \rightarrow \tau_B^* = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + s_{b0}} \text{ and } \tau_{Lt} \rightarrow \tau_L^* = \frac{\tau - \tau_B b_y - \bar{R} \cdot a}{1 - \alpha}$$

Proof. The proposition follows directly from the above observations and from part (3) of Proposition C2. **Q.E.D.**

C.7 Discussion

Does the intertemporal social optimum involve the accumulation of positive government assets ($a_t \rightarrow a > 0$) or the accumulation of public debt ($a_t \rightarrow a < 0$)? Both cases can happen, depending on parameters. The socially optimal level of capital accumulation and the market equilibrium level of capital accumulation depend on largely independent parameters, so this can really go both ways.

On the one hand, the socially optimal level $\beta^* = \frac{\alpha}{R^*} = \frac{\alpha}{e^{r^* H} - 1}$ (with $r^* = \delta + \Gamma'n + \Gamma g$) depends on the capital share α and on the parameters of the social welfare function δ, Γ', Γ . Typically, a more patient planner ($\delta \rightarrow 0$) will accumulate more capital, while a more concave planner ($\Gamma \rightarrow +\infty$) will accumulate less capital. To take an extreme case, an infinitely concave planner will feel that there is no need to leave any capital to future generations ($\beta^* \rightarrow 0$ as $\Gamma \rightarrow +\infty$, as long as $g > 0$), since they will be richer than us anyway.¹⁰⁶ Conversely, in case

¹⁰⁶Of course, one problem with this reasoning is that if we leave no capital, then productivity growth itself might decline, or even disappear (here we assume g to be exogenous, so this problem does not arise). Another major shortcoming of the standard theoretical framework that we are using is that the same parameter Γ determines the preference for intra-generational and inter-generational redistribution, which sometimes leads to surprising disputes. E.g. 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; both sides take the long run n to be negligible), 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

$\Gamma = \Gamma' = 0$ (or $g = n = 0$), then an infinitely patient planner wants to accumulate infinite quantities of capital: $\beta^* \rightarrow +\infty$ as $\delta \rightarrow 0$. Note that the socially optimal capital-output ratio β^* does not depend at all on the parameters of private preferences, and in particular does not depend on the average saving taste $s = E(s_i)$.

On the other hand, the market equilibrium level of capital accumulation depends a lot on the parameters of private preferences (and not at all on the parameters of the social welfare function). That is, for a given long run tax policy τ_L, τ_B , we have: $b_{yt} \rightarrow b_y = \frac{s(1 - \tau_L)(1 - \alpha)e^{(r-g-n)H}}{1 - s(1 - \tau_B)e^{(r-g-n)H}}$.

Assume $a_t \rightarrow a = 0$. Then $\beta_t = b_{yt}e^{-r_t H} \rightarrow \hat{\beta}$ and $r_t \rightarrow \hat{r}$ such that $\hat{\beta} = \frac{s(1 - \alpha - \tau)e^{-(g+n)H}}{1 - s \cdot e^{(\hat{r}-g-n)H}}$ and $\hat{R} = e^{\hat{r}H} - 1 = \frac{\alpha}{\hat{\beta}}$. By substituting $e^{\hat{r}H} = 1 + \frac{\alpha}{\hat{\beta}}$ into the first equation we obtain:

$$\hat{\beta} = \frac{s(1 - \tau)}{e^{(g+n)H} - s} = \frac{s(1 - \tau)}{1 + G + N + GN - s}$$

This is the capital-output ratio that would be attained by private accumulation alone, in case the government has a zero asset position in the long run. In the same way as in the standard Harrod-Domar-Solow formula, private capital accumulation $\hat{\beta}$ depends positively on the saving taste s and negatively on the growth rate $g + n$. In case $s \rightarrow 1$ and $g + n \rightarrow 0$, then $\hat{\beta} \rightarrow +\infty$. Conversely, in case $s \rightarrow 0$ then unsurprisingly $\hat{\beta} \rightarrow 0$.

Note that the formula for $\hat{\beta}$ can also be rewritten in the standard Harrod-Domar-Solow form, i.e. $\hat{\beta} = \frac{\tilde{s}}{G + N + GN}$, where $\tilde{s} = s(1 - \tau + \hat{\beta}) - \hat{\beta}$ is the conventional saving rate as defined in the macroeconomic literature, i.e. \tilde{s} is equal to new saving as a fraction of new output (as opposed to s , which includes savings out of bequest received).

In case $\hat{\beta} > \beta^*$, e.g. if the average saving taste s is large enough as compared to $g + n$, then private agents tend to accumulate too much capital, so in order to satisfy the modified Golden rule the social planner will need to accumulate public debt: $a_t \rightarrow a < 0$.

Conversely, in case $\hat{\beta} < \beta^*$, e.g. if the average saving taste s is small enough, then private agents tend to accumulate too little capital, so in order to satisfy the modified Golden rule the social planner will need to accumulate public assets: $a_t \rightarrow a > 0$.

In a full fledged life cycle model, pure demographic parameters - and not only saving tastes - would also matter, following the Modigliani triangle formula. One could again end up with too large or too small capital accumulation, depending on the specific parameters. The general point is that there is no reason in general to expect the market equilibrium to deliver more or less capital accumulation than the social optimum: it can really go both ways.

In practice, the only case where one can be pretty confident that there is excessive private capital accumulation is in the case where s and $\hat{\beta}$ are so large than $\hat{r} < g + n$. That is, if one observes that in the absence of government intervention the rate of return to wealth is less than the economy's growth rate, then one can be sure that this is collectively inefficient. An infinite

the long run relative price of the environment in a two-good growth model (see Guesnerie 2004).

horizon planner could raise intertemporal welfare by borrowing against future resources and forcing agents to consume more today. Technically, the net present value of future resources in this case would be infinite, so the planner budget constraint would not even be well defined (i.e. assumption 6 above would be violated). Intuitively, as long as $r < g + n$, then a planner can improve everybody's welfare by taking some of today's private savings (with market return r) and put them into a pay-as-you-go pension system (whose internal return is equal to $g + n$) - which is equivalent to issuing public debt, so as to reduce aggregate capital accumulation.¹⁰⁷

However available evidence shows that the aggregate rate of return to wealth is generally much larger than the growth rate, which suggests that real world economies are not in this situation of extreme dynamic inefficiency.¹⁰⁸ Yet another way to see this is to note that $r < g + n$ is equivalent to $\alpha < \tilde{s}$ (one simply needs to multiply both sides by β). That is, in steady-state the rate of return is less than the economy's growth rate if and only if the capital share is less than the saving rate (defined in the conventional sense). This clearly corresponds to a situation of excessive capital accumulation: capital brings less extra output than what we need to save in order to keep capital-output constant. This was the theoretical point made by Allais, Phelps and other authors in their original derivation of the (non-modified) Golden rule $r^* = g + n$: along a Golden rule path, a society should save for future generations exactly as much as the product share coming from the capital stock accumulated by past generations (these authors were implicitly assuming $\delta = 0$ and $\Gamma' = \Gamma = 1$).¹⁰⁹ Empirically, one alternative way to make

¹⁰⁷The central point made by Diamond (1965) was exactly this: in a general OLG model, one could very well get so much over-accumulation of capital that $r < g + n$, in which case pay-as-you-go pension systems can be an efficient way to reduce aggregate capital accumulation.

¹⁰⁸In practice the rate of return varies enormously over assets, from very low levels for money and government bonds (typically less than 1-2%) to intermediate levels for real estate (say 3-5%) to high levels for equity and other risky financial assets (say 7-8%). One way to proceed is to compute the average macroeconomic rate of return $r = \alpha/\beta$ by dividing the capital share by the capital-output ratio. If we do this, we typically find an average r around 4%-5%, much larger than $g + n$. E.g. in France the average macroeconomic r has been above $g + n$ during each single decade of the 1820-2010 period. See Piketty (2010, 2011).

¹⁰⁹The application of the term "Golden rule" to the optimal capital accumulation problem is generally attributed to Phelps (1961, 1965), who also proposed the simple "optimal savings rate" derivation described below. Allais (1947, 1962) first stated explicitly the idea that the return to capital should be equal to the growth rate of output in the "capitalistic optimum", but the modeling used by Allais was much less transparent. See also von Neumann (1945) and Malinvaud (1953). See Nobel Committee (2006, pp.17-22) for full references. Phelps' derivation works as follows. Assume $g = 0$ and $n > 0$. In steady-state, the aggregate capital stock K_t grows at the same rate as population. Per capital capital stock $k_t = K_t/N_t$ and output $y_t = f(k_t)$ are stationary. Phelps asks the following question: what is the saving rate s maximizing steady-state per capita consumption $c = (1 - s) \cdot f(k)$? In steady-state, we have $s \cdot f(k) = n \cdot k$, so per capita consumption can be rewritten $c = f(k) - n \cdot k$, the maximization of which leads to $r^* = f'(k^*) = n$. In effect, Phelps is maximizing a social welfare objective with $\delta = 0$ (this is self-evident to Phelps and other authors: the whole point of the Golden-rule literature was to study whether the basic moral principle "do unto others as you would have them do unto you" could be applied intergenerationally inside the Solow growth model to arrive at some form of social optimum; so it would have been strange to put less weight on future generations) and $\Gamma' = 1$ (Phelps cares about maximizing per capita consumption, i.e. uses a non-Benthamite welfare objective; with

sure that we are not in a situation of extreme dynamic inefficiency (i.e. $r < g + n$) is simply to check that capital shares are indeed larger than saving rates.¹¹⁰

Aside from this extreme case, i.e. as long as $r > g + n$, it is relatively difficult to decide whether $r > r^*$ or $r < r^*$ - this really depends on the choice of normative parameters δ, Γ', Γ , which as noted above are relatively controversial, and on which we do not take a stand here.¹¹¹

In any case, whether there is too much or too little aggregate capital accumulation in the real world, the key point here is that this dynamic efficiency issue is essentially orthogonal to the issue of optimal tax mix between capital and labor. That is, the optimal long run tax rate on capitalized bequest $\tau_B(\delta') = \frac{1 - (1 - \alpha - \tau)s_{b0}e^{\delta'H}/b_y}{1 + s_{b0}e^{\delta'H}}$ does not depend at all on whether $a > 0$ or $a < 0$.

C.8 Proof of Proposition C1

C.8.1 Main Proof

Consider a tax policy sequence $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$, and assume that this is the intertemporal welfare optimum.

Period-by-period budget balance implies: $\forall t \geq 0, \tau_{Lt} = \frac{\tau - \tau_{Bt}b_{yt}}{1 - \alpha}$.

The social welfare \tilde{V}_t of generation t -zero bequest receivers is given by:

$$\tilde{V}_t = \frac{\tilde{v}_t^{1-\Gamma}}{1-\Gamma} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'} = \frac{\tilde{v}_t^{1-\Gamma}}{1-\Gamma} \cdot y_0^{1-\Gamma} \cdot N_0^{1-\Gamma'} \cdot e^{(1-\Gamma)gHt + (1-\Gamma')nHt}$$

With: $\tilde{v}_t = v_t \cdot (1 - \tau_{Lt}) \cdot (1 - \alpha) \cdot \tilde{\theta} = v_t \cdot (1 - \alpha - \tau + \tau_{Bt}b_{yt}) \cdot \tilde{\theta}$

$$v_t = [E(v_{ti}^{1-\Gamma})]^{1/(1-\Gamma)},$$

$$v_{ti} = (1 - s_i)^{1-s_i} s_i^{s_i} [(1 - \tau_{Bt+1})e^{rH}]^{s_{bi}}$$

$$\tilde{\theta} = [E(\theta_i^{1-\Gamma})]^{1/(1-\Gamma)}$$

Total intertemporal social welfare is given by: $SWF = \sum_{t \geq 0} \tilde{V}_t e^{-\delta Ht}$

$\Gamma' = 0$, the optimum would always involve infinite capital accumulation: as long as $r > 0$, one can always raise total welfare by postponing consumption). With $g > 0$, the derivation is the same except that one nows tries to maximize the level of the multiplicative term in the steady-state per capita consumption growth path $c_t = (1 - s) \cdot f(\tilde{k}) \cdot e^{gHt} = [f(\tilde{k}) - (n + g) \cdot \tilde{k}] \cdot e^{gHt}$, with $\tilde{k}_t = K_t/L_t =$ capital per efficiency labor unit, which yields $r^* = f'(\tilde{k}^*) = n + g$. In effect this is assuming a social welfare function with a logarithmic form ($\Gamma = 1$).

¹¹⁰This is the approach followed by Abel et al. (1989). They compare gross profit rates and investment rates in the non-financial corporate sector in the U.S., the U.K., France, Germany, Italy, Canada and Japan over the 1950-1985 period and find that the former is larger than the latter in every year and in every country (usually by at least 10 points of GDP: say 20-25% of GDP for gross corporate profits and 10-15% of GDP for gross corporate investment).

¹¹¹Note that if the modified social discount rate becomes infinitely small, then in effect the modified Golden rule becomes infinitely close to the non-modified Golden rule. That is, if $\delta' = \delta - (1 - \Gamma)g - (1 - \Gamma')n \rightarrow 0$, then $r^* = \delta + \Gamma g + \Gamma' n \rightarrow g + n$. However this normative framework suffers from a number of limitations (see above).

It is straightforward that the maximization of SWF leads to $\tau_{B0} = 1$ and $\tau_{L0} = \frac{\tau - b_{y0}}{1 - \alpha}$. This follows from the fact that τ_{B0} only enters -positively- into generation-0 social welfare \tilde{V}_0 . I.e. at time $t = 0$, capital is on the table and should be taxed as much as possible).

For $t > 0$, we have a non-generate trade-off, since τ_{Bt} enters positively into generation- t social welfare \tilde{V}_t and negatively into generation- $t - 1$ social welfare \tilde{V}_{t-1} . That is, generation- t zero-bequest receivers benefit from a higher τ_{Bt} (since this leads to a lower tax rate τ_{Lt} on their labor income), while generation- $t - 1$ zero-bequest receivers lose from a higher τ_{Bt} (since this reduces the utility from leaving a bequest to their children).

The marginal changes $d\tilde{V}_t$ and $d\tilde{V}_{t-1}$ corresponding to a small change $d\tau_{Bt}$ are given by:

$$d\tilde{V}_t = d\tilde{v}_t \cdot \tilde{v}_t^{-\Gamma} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'}$$

$$d\tilde{V}_{t-1} = d\tilde{v}_{t-1} \cdot \tilde{v}_{t-1}^{-\Gamma} \cdot y_{t-1}^{1-\Gamma} \cdot N_{t-1}^{1-\Gamma'}$$

$$\text{With: } d\tilde{v}_t = \frac{1 - \tau_{Bt}}{1 - \alpha - \tau + \tau_{Bt}b_{yt}} \cdot b_{yt} \cdot \tilde{v}_t \cdot \frac{d\tau_{Bt}}{1 - \tau_{Bt}}$$

$$d\tilde{v}_{t-1} = -s_{b0t-1} \cdot \tilde{v}_{t-1} \cdot \frac{d\tau_{Bt}}{1 - \tau_{Bt}}$$

$$\text{And: } s_{b0t} = \frac{E(v_{ti}^{1-\Gamma} \cdot s_{bi})}{E(v_{ti}^{1-\Gamma})}$$

Setting $dSWF = e^{\delta H(t-1)}d\tilde{V}_{t-1} + e^{\delta Ht}d\tilde{V}_t = 0$, we obtain:

$$\tau_{Bt} = \frac{1 - (1 - \alpha - \tau)s_{b0t-1}e^{\delta' H}(\tilde{v}_{t-1}/\tilde{v}_t)^{1-\Gamma}/b_{yt}}{1 + s_{b0t-1}e^{\delta' H}(\tilde{v}_{t-1}/\tilde{v}_t)^{1-\Gamma}}$$

With: $\delta' = \delta - (1 - \Gamma)g - (1 - \Gamma')n$

Note that with Cobb-Douglas utility functions and period-by-period budget constraint the aggregate b_{yt} path is unaffected by tax changes (so we do not need to take into account a db_{yt} term). I.e. starting with any initial b_{y0} , we have: $b_{yt+1} = s(1 - \tau_{Lt})(1 - \alpha)e^{(r-g-n)H} + s(1 - \tau_{Bt})e^{(r-g-n)H}b_{yt} = s(1 - \alpha - \tau)e^{(r-g-n)H} + s \cdot e^{(r-g-n)H}b_{yt}$. As $t \rightarrow +\infty$, $b_{yt} \rightarrow b_y = \frac{s(1 - \alpha - \tau)e^{(r-g-n)H}}{1 - s \cdot e^{(r-g-n)H}}$. That is, with Cobb-Douglas utility and period-by-period budget constraint, the elasticity e_B of the aggregate bequest flow with respect to tax changes is equal to zero, both in the short run and in the long run.

As $t \rightarrow +\infty$, we also have $\tilde{v}_t \rightarrow \tilde{v} = v \cdot (1 - \alpha - \tau + \tau_B b_y) \cdot (1 - \alpha) \cdot \tilde{\theta}$, with $v = [E(v_i^{1-\Gamma})]^{1/(1-\Gamma)}$ and $v_i = (1 - s_i)^{1-s_i} s_i^{s_i} [(1 - \tau_B)e^{rH}]^{s_{bi}}$, and $s_{b0t} \rightarrow s_{b0} = \frac{E(v_i^{1-\Gamma} \cdot s_{bi})}{E(v_i^{1-\Gamma})}$.

We therefore have the following formula for the asymptotic tax rate:

$$\tau_{Bt} \rightarrow \tau_B(\delta') = \frac{1 - (1 - \alpha - \tau)s_{b0}e^{\delta' H}/b_y}{1 + s_{b0}e^{\delta' H}}$$

Q.E.D.

C.8.2 Alternative proof.

The following alternative proof focuses on small long run tax changes and further clarifies the role played by social discount rates. In addition this alternative proof also applies to the general case with any long run elasticity e_B .

Consider a tax policy sequence $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$, and assume that this is intertemporal welfare optimum. Assume that as $t \rightarrow +\infty$, $\tau_{Bt} \rightarrow \tau_B$, $\tau_{Lt} \rightarrow \tau_L$, and $b_{yt} \rightarrow b_y$.

Consider a small, permanent change in the bequest tax rate occurring after some time $t_0 > 0$. I.e. for $t \geq t_0$, τ_{Bt} becomes $\tau_{Bt} + d\tau_B$, with $d\tau_B > 0$ or $d\tau_B < 0$. The labor tax rate needs to adjust to $\tau_{Lt} + d\tau_{Lt}$, (so as to maintain period-by-period budget balance), and the aggregate inheritance-output ratio adjusts to $b_{yt} + db_{yt}$. Period-by-period budget balance implies: $d\tau_{Lt} = -\frac{b_{yt}d\tau_B + \tau_{Bt}db_{yt}}{1 - \alpha}$.

By definition of the elasticity e_B , as $t \rightarrow +\infty$, $db_{yt} \rightarrow db_y = -e_B \cdot b_y \cdot \frac{d\tau_B}{1 - \tau_B}$ and $d\tau_{Lt} \rightarrow d\tau_L = -\frac{b_y d\tau_B}{1 - \alpha} \left(1 - \frac{e_B \tau_B}{1 - \tau_B}\right)$.¹¹²

The implied change on the welfare of generation t - zero bequest receivers is given by: $d\tilde{V}_t = d\tilde{v}_t \cdot \tilde{v}_t^{-\Gamma} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'}$

For $t \geq t_0$, we have: $d\tilde{v}_t = -s_{b0t} \cdot \tilde{v}_t \cdot \frac{d\tau_B}{1 - \tau_{Bt+1}} - \tilde{v}_t \cdot \frac{d\tau_{Lt}}{1 - \tau_{Lt}}$

With: $s_{b0t} = \frac{E(v_{ti}^{1-\Gamma} \cdot s_{bi})}{E(v_{ti}^{1-\Gamma})}$

Using period-by-period budget balance equations, and letting $t \rightarrow +\infty$, we have:

$d\tilde{v}_t \rightarrow d\tilde{v} = \tilde{v} \cdot \frac{d\tau_B}{1 - \tau_B} \cdot \left[\frac{1 - (1 + e_B)\tau_B}{1 - \alpha - \tau + \tau_B b_y} b_y - s_{b0} \right]$

With: $s_{b0} = \frac{E(v_i^{1-\Gamma} \cdot s_{bi})}{E(v_i^{1-\Gamma})}$

Define $\tau_B^* = \frac{1 - (1 - \alpha - \tau)s_{b0}/b_y}{1 + e_B + s_{b0}}$ the steady-state welfare optimum.

We have: $\frac{d\tilde{v}}{d\tau_B} > 0$ iff $\tau_B < \tau_B^*$.

That is, by taking t_0 large enough, one can increase the welfare of all generations $t \geq t_0$ by raising the long run bequest tax rate τ_B if it is smaller than τ_B^* , and by reducing the long run bequest tax rate τ_B if it is larger than τ_B^* .

This does not imply, however, that the asymptotic optimum tax rate τ_B has to be equal to

¹¹²Strictly speaking, with Cobb-Douglas utility, i.i.d. shocks, and period-by-period budget constraint, $e_B = 0$ (and the terms db_{yt} are equal to zero all along the adjustment path). Here we write the proof with any positive (or negative) e_B in order to show how it works in the general case. Note that we ignore the fact that e_B is not strictly constant overtime as the elasticity following a permanent reform at time t_0 builds up over time. It is possible to write the proof with a time varying e_B^t in which case the optimal formula depends on the average elasticity \bar{e}_B across time periods discounted by time factor $e^{-\delta' H t}$. This average elasticity is not strictly equal to the long-run steady state elasticity e_B used in the main text but should be quantitatively very close for small δ' . When $\delta' \rightarrow 0$ then naturally $\bar{e}_B \rightarrow e_B$ and we recover exactly the same formula as in the main text.

τ_B^* , because we also need to take into account the impact of $d\tau_B$ on the welfare of generation $t_0 - 1$.

$$\text{I.e.: } dSWF = e^{-\delta H(t_0-1)} d\tilde{V}_{t_0-1} + \sum_{t \geq t_0} d\tilde{V}_t e^{-\delta Ht}$$

$$\text{With: } d\tilde{v}_{t_0-1} = -s_{b0t_0-1} \cdot \tilde{v}_{t_0-1} \cdot \frac{d\tau_B}{1 - \tau_{Bt_0}}$$

E.g. if $\tau_B < \tau_B^*$, then raising τ_{Bt} to $\tau_{Bt} + d\tau_B$ for all $t \geq t_0$ will increase welfare \tilde{V}_t of all generations $t \geq t_0$ (for t_0 large enough), but will reduce the welfare \tilde{V}_{t_0-1} of generation $t_0 - 1$ zero bequest receivers (since the latter do not benefit from a reduction in their labor tax rate, but derive less utility from the bequest left to their children). With a corrected social discount rate that is arbitrarily close to zero, this negative effect on generation $t_0 - 1$ is negligible, and the asymptotic optimum tax rate is arbitrarily close to the steady-state optimum τ_B^* . But as long as the corrected social discount rate is strictly positive, then this negative effect cannot be neglected, implying that the asymptotic optimum is strictly larger than τ_B^* . To see this, one can re-arrange $dSWF$ in the following way:

$$dSWF = \sum_{t \geq t_0} \xi_t \cdot d\tau_B \cdot e^{-\delta' Ht} \cdot y_{t_0}^{1-\Gamma} \cdot N_{t_0}^{1-\Gamma'}$$

$$\text{With: } \xi_t = \frac{\tilde{v}_t}{1 - \tau_{Bt}} \cdot \left[\frac{1 - (1 + e_B)\tau_{Bt}}{1 - \alpha - \tau + \tau_{Bt}b_{yt}} b_{yt} - s_{b0t-1} \cdot (\tilde{v}_{t-1}/\tilde{v}_t) \cdot e^{\delta' H} \right]$$

$$\text{And: } \delta' = \delta - (1 - \Gamma)g - (1 - \Gamma')n$$

$$\text{As } t \rightarrow +\infty, \xi_t \rightarrow \xi = \frac{\tilde{v}}{1 - \tau_B} \cdot \left[\frac{1 - (1 + e_B)\tau_B}{1 - \alpha - \tau + \tau_B b_y} b_y - s_{b0} \cdot e^{\delta' H} \right]$$

$$\text{Define: } \tau_B(\delta') = \frac{1 - (1 - \alpha - \tau)s_{b0}e^{\delta' H}/b_y}{1 + s_{b0}e^{\delta' H} + e_B}$$

$$\text{We have: } \xi > 0 \text{ iff } \tau_B < \tau_B(\delta').$$

Now assume $\tau_B < \tau_B(\delta')$, so that $\xi > 0$.

Pick any $\varepsilon > 0, \varkappa > 0$ s.t. $\varepsilon < \tau_B(\delta') - \tau_B$ and $\varkappa < \xi$. Then $\exists t_0 \geq 0$ s.t. $\forall t \geq t_0$, $\tau_{Bt} < \tau_B(\delta') - \varepsilon$ and $\xi_t > \xi - \varkappa > 0$.

One can see that if one picks $d\tau_B = \varepsilon$, then moving from the tax policy sequence τ_{Bt}, τ_{Lt} to the sequence $\tau_{Bt} + d\tau_B, \tau_{Lt} - d\tau_{Lt}$ for $t \geq t_0$ does raise intertemporal social welfare:

$$dSWF \geq \sum_{t \geq t_0} (\xi - \varkappa) \cdot \varepsilon \cdot e^{-\delta' Ht} \cdot y_{t_0}^{1-\Gamma} \cdot N_{t_0}^{1-\Gamma'} = (\xi - \varkappa) \cdot \varepsilon \cdot e^{-\delta' Ht} \cdot y_{t_0}^{1-\Gamma} \cdot N_{t_0}^{1-\Gamma'} \cdot \frac{e^{-\delta' Ht_0}}{1 - e^{-\delta' H}} > 0.$$

This contradicts the fact that the tax policy sequence τ_{Bt}, τ_{Lt} maximizes intertemporal social welfare.

Conversely, if one assume $\tau_B > \tau_B(\delta')$, one can increase SWF by cutting the long run bequest tax rate, i.e. one can find $t_0 \geq 0$ and $d\tau_B < 0$ s.t. moving from the tax policy sequence τ_{Bt}, τ_{Lt} to the sequence $\tau_{Bt} + d\tau_B, \tau_{Lt} - d\tau_{Lt}$ for $t \geq t_0$ raises intertemporal social welfare.

Therefore we have shown that the long run bequest tax rate must be equal to $\tau_B = \tau_B(\delta')$.

Note that both proofs work for any $\delta, \Gamma, g, \Gamma', n$, as long as assumption 4 is satisfied, i.e. as long as $\delta' > 0$. **Q.E.D.**

C.9 Proof of Proposition C2

C.9.1 Part 1: $r < r^*$

Consider first the case $r < r^* = \delta + \Gamma g + \Gamma' n$. Take any tax policy sequence $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$ satisfying the intertemporal budget constraint and converging towards some asymptotic tax policy (τ_B, τ_L) as $t \rightarrow +\infty$. Under assumptions 1-6, $b_{yt} \rightarrow b_y = \frac{s(1-\tau_L)(1-\alpha)e^{(r-g)H}}{1-s(1-\tau_B)e^{(r-g)H}}$, and $\bar{\tau}_t \rightarrow \bar{\tau} = \tau_L(1-\alpha) + \tau_B b_y$. Assume that $\tau_L < 1$.

Consider a small, budget balanced tax change whereby the planner reduces the labor tax rate from τ_{Lt_0} to $\tau_{Lt_0} - d\tau$ at time $t_0 \geq 0$ (with $d\tau > 0$) and raises the labor tax rate from τ_{Lt_1} to $\tau_{Lt_1} + d\tau'$ at some future dates $t_1 > t_0$ (with $d\tau' > 0$). The bequest tax rate sequence τ_{Bt} is unchanged.

Part 1.1 Neglecting for the time being the impact of this labor tax change on the b_{yt} path (and therefore on the stream of bequest tax revenues), intertemporal budget balance requires:

$$\begin{aligned} e^{rH(t_1-t_0)} Y_{Lt_0} d\tau &= Y_{Lt_1} d\tau' = e^{(g+n)H(t_1-t_0)} Y_{Lt_0} d\tau' \\ \text{I.e. } d\tau' &= e^{(r-g-n)H(t_1-t_0)} d\tau \end{aligned}$$

The social welfare \tilde{V}_t of generation t -zero bequest receivers is given by:

$$\tilde{V}_t = \frac{\tilde{v}_t^{1-\Gamma}}{1-\Gamma} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'} = \frac{\tilde{v}_t^{1-\Gamma}}{1-\Gamma} \cdot y_0^{1-\Gamma} \cdot N_0^{1-\Gamma'} \cdot e^{(1-\Gamma)gHt + (1-\Gamma')nHt}$$

With: $\tilde{v}_t = v_t \cdot (1 - \tau_{Lt}) \cdot (1 - \alpha) \cdot \tilde{\theta}$

$$v_t = [E(v_{ti}^{1-\Gamma})]^{1/(1-\Gamma)},$$

$$v_{ti} = (1 - s_i)^{1-s_i} s_i^{s_i} [(1 - \tau_{Bt+1})e^{rH}]^{s_{bi}}$$

$$\tilde{\theta} = [E(\theta_i^{1-\Gamma})]^{1/(1-\Gamma)}$$

Following small tax changes we have: $d\tilde{V}_t = d\tilde{v}_t \cdot \tilde{v}_t^{-\Gamma} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'} = -\frac{d\tau_{Lt}}{1-\tau_{Lt}} \cdot \tilde{v}_t^{1-\Gamma} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'}$

The total change in intertemporal social welfare induced by the labor tax change can therefore be written:

$$\begin{aligned} dSWF &= e^{-\delta H t_0} d\tilde{V}_{t_0} + e^{-\delta H t_1} d\tilde{V}_{t_1} \\ \text{i.e. } dSWF &= e^{-\delta H t_0} \cdot \frac{\tilde{v}_{t_0}^{1-\Gamma}}{1-\tau_{Lt_0}} \cdot y_{t_0}^{1-\Gamma} \cdot N_{t_0}^{1-\Gamma'} \cdot [d\tau - d\tau' \cdot \xi_{t_0, t_1} \cdot e^{[(1-\Gamma)g + (1-\Gamma')n - \delta]H(t_1-t_0)}] \end{aligned}$$

With: $\xi_{t_0, t_1} = \frac{1 - \tau_{Lt_0}}{1 - \tau_{Lt_1}} \cdot \frac{\tilde{v}_{t_1}^{1-\Gamma}}{\tilde{v}_{t_0}^{1-\Gamma}} \rightarrow 1$ as $t_0, t_1 \rightarrow +\infty$

Since $d\tau' = e^{(r-g-n)H(t_1-t_0)} d\tau$, this can be rewritten:

$$dSWF = e^{-\delta H t} \cdot \frac{\tilde{v}_t^{1-\Gamma}}{1-\tau_{Lt}} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'} \cdot d\tau \cdot [1 - \xi_{t_0, t_1} \cdot e^{[r-r^*]H(t_1-t_0)}]$$

With : $r^* = \delta + \Gamma g + \Gamma' n$.

If $r < r^*$ then $\exists t_0^*$ s.t. $\forall t_1 > t_0 \geq t_0^*$, $\xi_{t_0, t_1} \cdot e^{[r-r^*]H(t_1-t_0)} < 1$, i.e. $dSWF > 0$

Therefore one can raise intertemporal social welfare by reducing τ_{Lt_0} and increasing τ_{Lt_1} , which contradicts the fact that the sequence τ_{Bt}, τ_{Lt} maximizes intertemporal welfare. It follows that the asymptotic labor tax rate must be equal to 1: as $t \rightarrow +\infty$, $\tau_{Lt} \rightarrow 1$.

Part 1.2 Taking into account the impact of the labor tax change on the b_{yt} path complicates the notations but does not alter the conclusion. The key reason is that both $d\tau$ and $d\tau'$ induce behavioral changes db_{yt} that are proportional to the initial mechanical changes $d\tau$ and $d\tau'$ and the proportion is the same for both $d\tau$ and $d\tau'$. Hence, the welfare consequence does not change.

To see this going from τ_{Lt_0} to $\tau_{Lt_0} - d\tau$ induces changes in b_{yt} for all $t > t_0$. Using the transition equation $b_{yt+1} = s(1 - \tau_{Lt})(1 - \alpha)e^{(r-g-n)H} + s(1 - \tau_{Bt})e^{(r-g-n)H}b_{yt}$, we have:

$$db_{yt} = s(1 - \alpha)e^{(r-g-n)H} \cdot d\tau \text{ if } t = t_0 + 1 \text{ and}$$

$$db_{yt} = s(1 - \alpha)e^{(r-g-n)H} \cdot d\tau \cdot \left(\prod_{t_0+1 \leq t' < t} s(1 - \tau_{Bt'})e^{(r-g-n)H} \right) \text{ if } t > t_0 + 1$$

The net present value at time t_0 of the total changes in tax revenues induced by going from τ_{Lt_0} to $\tau_{Lt_0} - d\tau$. can be written:

$$dT = -Y_{Lt_0} \cdot d\tau + \sum_{t > t_0} e^{-rH(t-t_0)} \cdot \tau_{Bt} \cdot Y_t \cdot db_{yt}$$

$$\text{I.e. } dT = -Y_{Lt_0} \cdot d\tau \cdot [1 - \varkappa_{t_0}].$$

$$\text{With: } \varkappa_{t_0} = \sum_{t > t_0} e^{-(r-g-n)H(t-t_0)} \cdot \tau_{Bt} \cdot s \cdot e^{(r-g-n)H} \cdot \left(\prod_{t_0+1 \leq t' < t} s(1 - \tau_{Bt'})e^{(r-g-n)H} \right)$$

$$\text{That is: } \varkappa_{t_0} = \sum_{t > t_0} \tau_{Bt} \cdot s \cdot \left(\prod_{t_0+1 \leq t' < t} s(1 - \tau_{Bt'}) \right)$$

$$\text{As } t \rightarrow +\infty, \tau_{Bt} \rightarrow \tau_B \text{ Therefore as } t_0 \rightarrow +\infty, \varkappa_{t_0} \rightarrow \frac{\tau_B \cdot s}{1 - s \cdot (1 - \tau_B)} < 1.$$

Similarly, going from τ_{Lt_1} to $\tau_{Lt_1} + d\tau'$. induces changes in b_{yt} for all $t > t_1$, namely:

$$db'_{yt} = -s(1 - \alpha)e^{(r-g-n)H} \cdot d\tau' \text{ if } t = t_1 + 1 \text{ and}$$

$$db'_{yt} = -s(1 - \alpha)e^{(r-g-n)H} \cdot d\tau' \cdot \left(\prod_{t_1+1 \leq t' < t} s(1 - \tau_{Bt'})e^{(r-g-n)H} \right) \text{ if } t > t_1 + 1$$

The net present value at time t_1 of the total changes in tax revenues induced by going from τ_{Lt_1} to $\tau_{Lt_1} + d\tau'$. can be written:

$$dT' = Y_{Lt_1} \cdot d\tau' + \sum_{t > t_1} e^{-rH(t-t_1)} \cdot \tau_{Bt} \cdot Y_t \cdot db'_{yt}$$

$$\text{I.e. } dT' = Y_{Lt_1} \cdot d\tau' \cdot [1 - \varkappa_{t_1}].$$

$$\text{With: } \varkappa_{t_1} = \sum_{t > t_1} \tau_{Bt} \cdot s \cdot \left(\prod_{t_1+1 \leq t' < t} s(1 - \tau_{Bt'}) \right) \rightarrow \frac{\tau_B \cdot s}{1 - s \cdot (1 - \tau_B)} \text{ as } t_1 \rightarrow +\infty.$$

The tax change $d\tau, d\tau'$ is budget balance if and only if $dT + e^{-r(t_1-t_0)}dT' = 0$, i.e. iff:

$$d\tau' = \frac{1 - \varkappa_{t_0}}{1 - \varkappa_{t_1}} \cdot e^{(r-g-n)H(t_1-t_0)} \cdot d\tau$$

The induced change in intertemporal social welfare can again be written:

$$dSWF = e^{-\delta Ht} \cdot \frac{\tilde{v}_t^{1-\Gamma}}{1 - \tau_{Lt}} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'} \cdot d\tau \cdot [1 - \xi_{t_0, t_1} \cdot e^{[r-r^*]H(t_1-t_0)}]$$

$$\text{With: } \xi_{t_0, t_1} = \frac{1 - \tau_{Lt_0}}{1 - \tau_{Lt_1}} \cdot \frac{\tilde{v}_{t_1}^{1-\Gamma}}{\tilde{v}_{t_0}^{1-\Gamma}} \cdot \frac{1 - \varkappa_{t_0}}{1 - \varkappa_{t_1}} \rightarrow 1 \text{ as } t_0, t_1 \rightarrow +\infty$$

So if $r < r^*$ we again have: $\exists t_0^*$ s.t. $\forall t_1 > t_0 \geq t_0^*$, $\xi_{t_0, t_1} \cdot e^{[r-r^*]H(t_1-t_0)} < 1$, i.e. $dSWF > 0$.

Part 1.3 In the same way, one can show that if $r < r^*$ and $\tau_{Bt} \rightarrow \tau_B < 1$, then one can increase intertemporal social welfare by reducing the bequest tax rate from τ_{Bt_0} to $\tau_{Bt_0} - d\tau$ at some time $t_0 \geq 0$ and raising the bequest tax rate from τ_{Bt_1} to $\tau_{Bt_1} + d\tau'$ at some future date $t_1 > t_0$ (where the small tax changes $d\tau, d\tau'$ are positive and budget balanced). It follows that if $r < r^*$ then we have both $\tau_{Lt} \rightarrow \tau_L = 1$ and $\tau_{Bt} \rightarrow \tau_B = 1$. Note that since $\tau_{Lt} \rightarrow \tau_L = 1$, $b_{yt} \rightarrow b_y = 0$, i.e. in the long run the bequest tax rate does not matter since there is nothing to tax. Finally, intertemporal budget balance implies that $a_t \rightarrow a = \frac{\tau - \bar{\tau}}{R} = -\frac{1 - \alpha - \tau}{R} < 0$.

C.9.2 Part 2: $r > r^*$

Conversely, in the case $r > r^*$, one can show in a similar way that one can increase intertemporal social welfare by raising the labor tax rate from τ_{Lt_0} to $\tau_{Lt_0} + d\tau$ at some time $t_0 \geq 0$ and reducing the labor tax rate from τ_{Lt_1} to $\tau_{Lt_1} - d\tau'$ at some future date $t_1 > t_0$ (where the small tax changes $d\tau, d\tau'$ are positive and budget balanced), or by raising the bequest tax rate from τ_{Bt_0} to $\tau_{Bt_0} + d\tau$ at some time $t_0 \geq 0$ and reducing the bequest tax rate from τ_{Bt_1} to $\tau_{Bt_1} - d\tau'$ at some future date $t_1 > t_0$. It follows that if $r > r^*$, then τ_{Lt}, τ_{Bt} must converge towards their minimal values $\underline{\tau}_L, \underline{\tau}_B$.

C.9.3 Part 3: $r = r^*$

Finally consider the knife-edge case $r = r^*$. Depending on the initial conditions and the specific parameters, the optimal asset and tax policy sequence $(a_t, \tau_{Bt}, \tau_{Lt})_{t \geq 0}$ might involve positive or negative government asset position in the long run: $a_t \rightarrow a > 0$ or < 0 . Taking as given the optimal sequence $(a_t)_{t \geq 0}$, one can derive the same proof as in Proposition C1 in order to derive the asymptotic properties of $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$. That is, taking $(a_t)_{t \geq 0}$ as given, the intertemporal budget constraint can be rewritten as a period-by-period budget constraint:

$$\begin{aligned} \bar{\tau}_t &= \tau_{Lt}(1 - \alpha) + \tau_{Bt}b_{yt} = \tau + a_{t+1}e^{(g+n)H} - a_t e^{rH} \\ \text{I.e. } : \quad \tau_{Lt} &= \frac{\tau - \tau_{Bt}b_{yt} + a_{t+1}e^{(g+n)H} - a_t e^{rH}}{1 - \alpha} \end{aligned}$$

Note also that for a given $(a_t)_{t \geq 0}$, changes in $(\tau_{Bt}, \tau_{Lt})_{t \geq 0}$ do not affect the b_{yt} path, since we have:

$$b_{yt+1} = s(1 - \alpha - \bar{\tau}_t)e^{(r-g-n)H} + s \cdot e^{(r-g-n)H}b_{yt}$$

It follows that if we take $(a_t)_{t \geq 0}$ as given, then any small bequest tax change $d\tau_{Bt}$ must be compensated by a labor tax change $d\tau_{Lt} = -b_{yt}d\tau_{Bt}/(1 - \alpha)$.

Using the same formulas for the social welfare \tilde{V}_t of generation t -zero bequest receivers, total social welfare SWF , and marginal welfare changes $d\tilde{V}_t$ and $dSWF$ as those given the proof of Proposition C1, we obtain the following results.

First, at period $t = 0$, we have $\tau_{B0} = 1$ and $\tau_{L0} = \frac{\tau - b_{y0} + a_1e^{(g+n)} - a_0e^{rH}}{1 - \alpha}$. This follows from the fact that τ_{B0} only enters -positively- into generation-0 social welfare \tilde{V}_0 . I.e. at time $t = 0$, capital is on the table and should again be taxed as much as possible).

For $t > 0$, we have a non-generate trade-off, since τ_{Bt} enters positively into generation- t social welfare \tilde{V}_t and negatively into generation- $t - 1$ social welfare \tilde{V}_{t-1} .

Setting $dSWF = e^{\delta H(t-1)}d\tilde{V}_{t-1} + e^{\delta Ht}d\tilde{V}_t = 0$, we obtain:

$$\tau_{Bt} = \frac{1 - (1 - \alpha - \tau)s_{b0t-1}e^{\delta'H}(\tilde{v}_{t-1}/\tilde{v}_t)^{1-\Gamma}/b_{yt}}{1 + s_{b0t-1}e^{\delta'H}(\tilde{v}_{t-1}/\tilde{v}_t)^{1-\Gamma}}$$

As $t \rightarrow +\infty$, we again have:

$$\tau_{Bt} \rightarrow \tau_B(\delta') = \frac{1 - (1 - \alpha - \tau)s_{b0}e^{\delta'H}/b_y}{1 + s_{b0}e^{\delta'H}}$$

For the asymptotic labor tax rate, the only difference with the previous formula is that we now have a $-\bar{R} \cdot a$ term:

$$\tau_{Lt} = \frac{\tau - \tau_{Bt}b_{yt} + a_{t+1}e^{(g+n)H} - a_t e^{rH}}{1 - \alpha} \rightarrow \tau_L(\delta') = \frac{\tau - \tau_B(\delta')b_y - \bar{R} \cdot a}{1 - \alpha}$$

With: $\bar{R} = e^{rH} - e^{(g+n)H} = 1 + R - (1 + G)(1 + N) = R - G - N - GN$. **Q.E.D.**

Additional Appendix References

- Abel, A., G. Mankiw, L. Summers and R. Zeckhauser** 1989. "Assessing Dynamic Efficiency: Theory and Evidence", *Review of Economic Studies* 56, 1-20.
- Allais, Maurice** 1947. *Economie et intérêt*, Paris: Imprimerie nationale.
- Allais, Maurice** 1962. "The Influence of the Capital-output Ratio on Real National Income", *Econometrica*, 30, 700-728.
- Blanchard Olivier and Stanley Fisher** 1989. *Lectures on Macroeconomics*, Cambridge: MIT Press, 650p.
- Bourguignon, François,** "Pareto Superiority of Unequalitarian Equilibria in Stiglitz' Model of Wealth Distribution with Convex Saving Function", *Econometrica*, Vol. 49, No. 6, 1981, pp. 1469-1475
- Brunner, Johann, and Suzanne Pech** 2011a. "Optimal Taxation of Bequests in a Model with Initial Wealth", Working Paper, University of Linz, forthcoming, *Scandinavian Journal of Economics*.
- Brunner, Johann, and Suzanne Pech** 2011b. "Optimal Taxation of Wealth Transfers when Bequests Are Motivated by Joy of Giving", Working Paper, University of Linz.
- Carroll C., J. Overland, and D.N. Weil** 2000. "Saving and Growth with Habit Formation", *American Economic Review* 90(3), 341-355.
- Diamond, Peter** 1965. "National Debt in a Neoclassical Growth Model", *American Economic Review* 55(5), 1126-1150.
- Guesnerie, Roger** 2004. "Calcul économique et développement durable", *Revue économique*, 55(5), 363-382.
- Hopehnayn, Hugo and Edward Prescott** 1992. "Stochastic Monotonicity and Stationary Distributions for Dynamic Economies", *Econometrica*, 60(6), 1387-1406.
- Malinvaud, Edmond** 1953. "Capital Accumulation and Efficient Allocation of Resources", *Econometrica*, 21, 233-268.
- Meade, James** 1978. *The Structure and Reform of Direct Taxation*, Report of a Committee chaired by J.E. Meade, Institute for Fiscal Studies, 551p.
- Milesi-Ferretti, Gian-Maria, and Nouriel Roubini** 1998. "Growth Effects of Income and Consumption Taxes", *Journal of Money, Credit and Banking*, 30(4), 721-744.
- Mirrlees, James, et al.** 2011. *Tax by Design: the Mirrlees Review*, Oxford University Press, 548p.
- Nirei, Makoto** 2009. "Pareto Distributions in Economic Growth Models", Working Paper, Hitotsubashi University, 34p.
- Nobel Committee** 2006. "Edmund Phelps's Contributions to Macroeconomics", The Royal Swedish Academy of Sciences, 33p.
- Nordhaus, William** 2007a. "Critical Assumptions in the Stern Review on Climate Change", *Science*, 317, 201-202.
- Nordhaus, William** 2007b. "A Review of the Stern Review on the Economics of Climate Change", *Journal of Economic Literature*, 45, 686-702.

- Phelps, Edmund** 1961. “The Golden Rule of Accumulation: A Fable for Growthmen”, *American Economic Review*, 51(4), 638-643.
- Phelps, Edmund** 1965. “Second Essay on the Golden Rule of Accumulation”, *American Economic Review*, 55(4), 793-814.
- Philippon, Thomas** 2011. “Has the U.S. Finance Industry Become Less Efficient?”, Working Paper, NYU, 28p.
- Piketty, Thomas** 1997. “The Dynamics of the Wealth Distribution and the Interest Rate with Credit-Rationing”, *Review of Economic Studies*, 64(2), 173-189.
- Saez, Emmanuel, and Stefanie Stantcheva** 2012. “Endogenous Social Welfare Weights for Optimal Tax Theory”, UC Berkeley Working Paper.
- Stern, Nicholas** 2007. *The Economics of Climate Change: The Stern Review*, Cambridge University Press, 712p.
- Sterner T., and U.M. Persson** 2008. “An Even Sterner Review: Introducing Relative Prices into the Discounting Debate”, *Review of Environmental Economics & Policy*, 2, 2008, 61-76.
- Stiglitz, Joseph E.** 1969. “Distribution of Income and Wealth among Individuals”, *Econometrica*, 37, 382–97.
- Stiglitz, Joseph E.** 1978. “Notes on Estate Taxes, Redistribution and the Concept of Balanced Growth Path Incidence”, *Journal of Political Economy*, 86(2), S137-S150.
- Von Neumann, J.** 1945. “A Model of General Economic Equilibrium”, *Review of Economic Studies*, 13, 1-9.