NBER WORKING PAPER SERIES

# A THEORY OF OPTIMAL CAPITAL TAXATION

Thomas Piketty
Emmanuel Saez

Working Paper 17989
http://www.nber.org/papers/w17989

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2012

We are grateful to seminar participants at the Paris School of Economics, the London School of Economics, the National Bureau of Economic Research (Public Economics Program), the University of California at Berkeley, the Massachussetts 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2012 by Thomas Piketty and Emmanuel Saez. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

# Appendix (not for publication)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 $b_{yt} = \dfrac{e^{rH} b_t}{y_t}$. In the special case with i.i.d. taste and productivity shocks, $(s_{wti}, s_{bti})$ and $\theta_{ti}$ 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 $b_{ti}$:

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

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

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

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 $t \to +\infty$, 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 $t \to +\infty$, 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 $\mu = s(1-\tau_B)e^{(r-g)H} < 1$):

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

In the long-run, the minimal normalized inheritance level $z_0$ is given by: $z_0 = \dfrac{s_0(1-\mu)\theta_0}{s - s_0\mu} < 1$. 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 $proba(z_{it+T} > z' \mid z_{it} = z) > \varepsilon$ and $proba(z_{it+T} < z' \mid z_{it} = z'') > \varepsilon$. (consider a sufficiently long sequence of positive 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:

$$\begin{aligned} \forall(s_{wti}, s_{bti}) &\in S, (s_{wt+1i}, s_{bt+1i}) \in S, g(s_{wt+1i}, s_{bt+1i} \mid s_{wti}, s_{bti}) > 0 \\ (\text{with} &: s_{wti}, s_{bti} = \text{parental tastes, } s_{wt+1i}, s_{bt+1i} = \text{children tastes}) \end{aligned}$$

And:

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

[63] One 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.

$$\begin{aligned}\forall \theta_{ti} &\in \Theta, \theta_{t+1i} \in \Theta, h(\theta_{t+1i} \mid \theta_{ti}) > 0 \\ \text{(with} &: \theta_{ti} = \text{parental productivity}, \theta_{t+1i} = \text{children productivity)}\end{aligned}$$

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 $t \to +\infty$. 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 $t \to +\infty$, 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: $z_0 = \dfrac{s_0(1-\mu_z)\theta_0}{s - s_0\mu}$ and $z_1 = \dfrac{s_1(1-\mu_z)\theta_1}{s - s_1\mu}$ (assuming $\dfrac{s_1}{s}\mu < 1$; 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

[64]Note 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:

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

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 $\mu_z$) can be self-fulfilling because it implies higher steady-state inequality of inheritance, via the transition equation for $z_{ti}$: higher $\mu_z$ 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 $s \to \overline{s}(s) = E(s_{t+1i} \mid s_{ti} = s)$ 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]

[65]In 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

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

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

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

It follows that as $z \to +\infty$, $\log[1 - \Phi(z)] \approx a \cdot [\log(\omega) - \log(z)]$, i.e. $1 - \Phi(z) \approx \left(\dfrac{\omega}{z}\right)^a$, with a constant term $\omega = \dfrac{1-\mu}{\mu - p}$, a Pareto coefficient $a = \dfrac{\log(1/p)}{\log(\mu/p)} > 1$ and an inverted Pareto coefficient $b = \dfrac{a}{a-1} = \dfrac{\log(1/p)}{\log(1/\mu)} > 1$.

As $\mu \to 1^-$ (for given $p < \mu$), $a \to 1^+$ and $b \to +\infty$ (infinite inequality). Intuitively, the multiplicative inheritance effect becomes infinite as compared to the equalizing labor income effect. The same occurs as $p \to 0^+$ (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. $p = \dfrac{p_0}{1 + p_0 - p_1} \in [p_0, p_1]$. 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)

$z = z_1 = \dfrac{1-\mu_z}{p} > 0$ 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.

[66] In the binomial model, one can directly compute the "empirical" inverted Pareto coefficient $b' = \dfrac{E(z \mid z \geq z_k)}{z_k} \to \dfrac{1-p}{1-\mu}$ as $k \to +\infty$. 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: $z_k = \frac{1-\mu_z}{\mu - p} \cdot [(\frac{\mu}{p})^k - 1]$ 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, $z_0 \to 0$ and $a \to 1$ (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:

$$\begin{aligned} V_{ti} &= \max V_i(c_{ti}, w_{ti}, \bar{b}_{ti}) = c_{ti}^{1-s_i} w_{ti}^{s_{wi}} \bar{b}_{ti}^{s_{bi}} \quad \text{s.t.} \quad c_{ti} + w_{ti} \leq \tilde{y}_{ti} = (1-\tau_B) z_{ti} b_t e^{rH} + (1-\tau_L)\theta_{ti} y_{Lt} \\ &\to \; c_{ti} = (1-s_i) \cdot \tilde{y}_{ti},\; w_{ti} = s_i \cdot \tilde{y}_{ti},\; \bar{b}_{ti} = (1-\tau_B)e^{rH} \cdot s_i \cdot \tilde{y}_{ti} \\ \text{I.e.} \quad V_{ti} &= v_i \cdot \tilde{y}_{ti}, \text{ with } \; v_i = (1-s_i)^{1-s_i} s_i^{s_i} [(1-\tau_B)e^{rH}]^{s_{bi}}, \text{ and } V_{ci} = v_i \end{aligned}$$

[67] The 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 $\widetilde{y}_{ti} = (1-\tau_L)\theta_i y_{Lt}$. 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: $\max_{\tau_B} \tau_B b_y \iff \tau_B = \dfrac{1}{1+e_B}$.

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

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

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 $s_{b0} = \dfrac{E(v_i \cdot s_{bi})}{E(v_i)}$. 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: $SWF = \iint_{z\geq 0, \theta_0 \leq \theta \leq \theta_1} \omega_{p_z p_\theta} \dfrac{V_{z\theta}^{1-\Gamma}}{1-\Gamma} d\Psi(z,\theta)$, 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 $\widetilde{y}_{ti} = \widetilde{y}_{tz\theta} = (1-\tau_B)zb_t e^{rH} + (1-\tau_L)\theta y_{Lt}$) but with different taste parameters $s_{wi}$, $s_{bi}$. In effect, we are implicitly assuming that the welfare weights $\omega_i$ are the same for all individuals $i$ with the same ranks $p_z, p_\theta$ 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: $V_{z\theta} = v \cdot \widetilde{y}_{tz\theta}$, with $v = E(v_i)$; and $\tau_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}}$, with $s_{b0} = \dfrac{E(v_i \cdot s_{bi})}{E(v_i)}$. 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: $s_{b0} = \dfrac{p \cdot v_1 \cdot s_{b1}}{1-p+p \cdot v_1}$(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 $\omega_i$ 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: $\tau_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}}$ with $s_{b0} = \dfrac{E(\omega_i \cdot v_i \cdot s_{bi})}{E(\omega_i \cdot v_i)} = E(s_{bi}) = s_b$. 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 $\tau_B$ (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: $V_{z\theta} = v' \cdot [(1-\tau_B)e^{rH}]^{s_b} \cdot \widetilde{y}_{tz\theta}$, 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 $\tau_B$ leads directly to: $\tau_B = \dfrac{1-(1-\alpha-\tau)s_{b0}/b_y}{1+e_B+s_{b0}}$, with $s_{b0} = s_b$.

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

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

We have: $\tau_B > 0$ iff $b_y > s_{b0}(1-\alpha-\tau)$. Intuitively, if we start from $\tau_B = 0$ and $\tau_L = \tau/(1-\alpha)$, then $s_{b0}(1-\alpha-\tau) = s_{b0}(1-\alpha)(1-\tau_L)$ is the bequest-motive-driven fraction of income that zero-receivers are going to leave to their children; this measures how much $\tau_B$ 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 $b_y = \dfrac{s(1-\tau-\alpha)e^{(r-g)H}}{1-se^{(r-g)H}}$ into the $\tau_B$ formula, we obtain: $\tau_B = \dfrac{1+s_b-(s_b/s)e^{-(r-g)H}}{1+e_B+s_b}$. We get the following condition

[68] For 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 $\tau_B$ small).

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

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

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

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

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

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

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

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

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

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 $\theta_i$, and $\theta_z$ is the average of $\theta_i$ 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 $\theta_z$ is the simple average of $\theta_i$ 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 $b_y = \dfrac{s(1-\tau-\alpha)e^{(r-g)H}}{1-se^{(r-g)H}}$ into the $\tau_B$ formula, we obtain: $\tau_B = \dfrac{1-e^{-(r-g)H}(1-se^{(r-g)H})s_b/s - (1+e_B+s_b)z}{(1+e_B+s_b)(1-z)}$. 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 $\tau_B$ 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 $10^{th}$ 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 $10^{th}$ percentile rises by $dz > 0$. Define $e_{p_z}$ the elasticity of the $p_z$ percentile normalized inheritance $z$ with respect to $\tau_B$ (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 $\tau_B$ (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:

$z = z_k = \dfrac{1-\mu}{\mu-p} \cdot \left[\left(\dfrac{\mu}{p}\right)^k - 1\right]$ with probability $(1-p)\cdot p^k$ (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 $e_{p_z} = \dfrac{dz_1/z_1}{d\tau_B/(1-\tau_B)} = \dfrac{\mu}{1-\mu} > 0$ . I.e. the $dz$ effect raises the optimal tax rate.

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

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:

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

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

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

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

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 $\Gamma \to +\infty$, we have: $\bar{s}_b \to s_{b0} = E(s_{bi}|z_i = 0, \theta_i = \theta_0)$ and $\bar{z}/\bar{\theta} \to 0$, 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 re-arranged 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:

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

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

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

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

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

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

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

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

**(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 ($\nu_z$), and divided by average total resources of $z$-receivers/average total resources ($\pi_z$).[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 $\tau_B < 1/(1+e_B^*)$).

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

[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 $\pi_i = \tilde{y}_{ti}/\tilde{y}_t$) = average total resources of $z$-receivers/average total resources. In the general case, $\pi_z = \frac{\int_{z_i=z} V_{ci}\theta_i \pi_i d\Psi}{\int_{z_i=z} V_{ci}\theta_i d\Psi}$= average of $\pi_i$ 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 $-(1+R)V_{\bar{b}i}(b_{t+1i} - b_{t+1}^*)^+ d\tau_B$. 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 $V_{ci} = V_{wi} + (1-\tau_B)(1+R)V_{\bar{b}i}$, 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 $s_{bi} = s_i(1-\tau_B)(1+R)V_{\bar{b}i}/V_{ci}$. Hence, we have:

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

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

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 $\varepsilon_{ti}$ 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 $\xi$ 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 $\xi$ measures the extent to which rates of return are responsive to such efforts. When $\xi$ is close to 0, $R_{ti}$ is almost a pure noise: returns are determined by luck. Conversely when $\xi$ is large (as compared to the mean and variance of $\varepsilon_{ti}$), $R_{ti}$ 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 $C(e_{ti}) = (1-\widetilde{\tau}_B)b_{ti}c(e_{ti})$, where $(1-\widetilde{\tau}_B)b_{ti}$ is portfolio size (net-of-tax bequest) and $c(e_{ti})$ is a convex, increasing function of effort.[73]

[72] In 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 conventionaly 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.

[73] It would be interesting to introduce scale economies in portfolio management (i.e. by assuming that cost rises less than proportionally with portfolio size), so as to generate the realistic prediction that higher portfolios tend to get higher returns (at least over some range; 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 $\xi$ 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 $\xi$ 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 $\tau_K$ depends negatively upon the elasticity $e_R$. If $e_R$ is close to zero, then providing full insurance by taxing capital income at rate $\tau_K = 100\%$ is optimal in our model. Conversely, if $e_R$ is sufficiently large, then the disincentive effects of taxing capital income are so large that one zero capital income tax ($\tau_K = 0\%$) becomes optimal. 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 $\tau_K$ and a labor income tax $\tau_L$ such that:*

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

*(b) If $e_R$ is sufficiently small, then $\tau_K > \tau_L$; if $e_R$ is sufficiently large, then $\tau_K < \tau_L$*

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

**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 $\tau_K$ is arbitrarily close to 100%. The same continuity reasoning

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

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

applies to $e_R = \overline{e}_R$ and $\tau_K = 0\%$. Note that $\overline{e}_R$ is finite because 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 $b_y$.

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 $\varepsilon_{it}$ is equal to $-\varepsilon_0 < 0$. We assume $\varepsilon_0$ 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 $\varepsilon_{ti}$ 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, $\varepsilon_0$ 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

$$V^i_c = V^i_w + V^i_{\tilde{b}}(1-\tau_B)(1+(R-\varepsilon_0)(1-\tau_K)-c(R)) \quad \text{hence} \quad \nu_i = V^i_{\tilde{b}}(1-\tau_B)(1+(R-\varepsilon_0)(1-\tau_K)-c(R))/V^i_c$$

We also make the Cobb-Douglas utility assumption and assume that $s_i$ is orthogonal to $\theta_i$ 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 $\tau_B$ (for fixed $\tau$) so that $e_B = 0$ and depends upon $\tau_K$ 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 $\tau_K$ only but $b_t$ now depends in a complex way on both $\tau_K$ and $\tau_B$ (for a given $\tau$), 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 $\tau_B$.** 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 $\tau_K$, 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

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

Using the definition of $\nu_i = V_{\bar{b}}^i(1-\tau_B)(1+(R-\varepsilon_0)(1-\tau_K)-c(R))/V_c^i$, 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 $\tau_B$ 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 $\tau_B$ is smaller relative to proposition 2.

**Optimum $\tau_K$.** 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 $\tau_K$ 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

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

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

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

which leads to the fairly complex optimal tax formula for $\tau_K$:

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

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

If $e_R > 0$, then $\tau_K < 100\%$ and $\tau_B$ 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 would 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 $h_i(.)$ is increasing and convex (and could differ across individuals).

Individual $i$ labor income is $y_{Lti} = v_t \theta_i l_i$ where $\theta_i$ is individual productivity (with mean one across the population) and $v_t = v_o e^{gHt}$ is the average wage rate of generation $t$.[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 $\tau_L$ as a function of $\tau_B$ 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 $\tau_L$ and $\tau_B$ 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 $p_z > 0$:*

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

[76] As 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×labor income).*

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

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

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

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

**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 $\tau_B, \tau_L$). 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:

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

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

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

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

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:

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

[77] With 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 $\tau_B$.

Intuitively, a higher labor supply elasticity makes high labor taxation less efficient, 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* $\tau_L = 33\%$.

[78] This 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 $(\tau_B, \tau_L)$, we again have a unique long run steady-state: $b_{yt} \to b_y$, $R_t \to R$, $\Psi_t \to \Psi$ (Proposition 1). This follows from the fact in the open economy case the long run $b_y$ is an increasing function of the exogenous rate of return $R$ (i.e. long run capital supply is upward sloping). Since the demand for capital is downward sloping, there exists a unique long run rate of return $R$ clearing the capital market: $\frac{R}{1+R} = \frac{\alpha}{b_y}$.

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 $\tau_B, \tau_L$ 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 $b_{yt} = \dfrac{e^{rH} B_t}{Y_t}$ (where $B_t = N_t \cdot b_t$ is the aggregate bequest flow received by generation $t$) becomes:

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

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

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, optimal capital taxes are lower in high population growth countries, 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 $\tau_C$ in addition to the labor income tax $\tau_L$ and the capitalized bequest tax $\tau_B$. Whether $\tau_C$ 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 $\tau_C$ 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 $\tau_B$, so that we are constrained to have $\tau_B = 0$. Then it is clearly optimal to have some positive level of consumption tax $\tau_C$ in addition to the labor income tax $\tau_L$, 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 $\tau_C$ 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 program in presence of a consumption tax $\tau_C$ can then be written as follows:

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

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

$c_{ti}$ = consumption

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

$\overline{w}_{ti} = (1-\tau_C)w_{ti}$ = 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

[79] Kaldor 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 cannot be measured well) can also be found in Meade (1978) and King (1980).

[80] Unfortunately 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.

[81] In a model where 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 valuable 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 $(\tau_C, \tau_B, \tau_L)$ is equivalent to a tax mix with zero consumption tax $(\overline{\tau}_C = 0, \overline{\tau}_B, \overline{\tau}_L)$, where the corrected tax rates $\overline{\tau}_B, \overline{\tau}_L$ are given by: $1 - \overline{\tau}_B = (1 - \tau_C)(1 - \tau_B)$ and $1 - \overline{\tau}_L = (1 - \tau_C)(1 - \tau_L)$. It is also equivalent to a tax mix with zero bequest taxes $(\overline{\overline{\tau}}_C, \overline{\overline{\tau}}_B = 0, \overline{\overline{\tau}}_L)$, where the corrected tax rates $\overline{\overline{\tau}}_C, \overline{\overline{\tau}}_L$ are given by: $1 - \overline{\overline{\tau}}_C = (1 - \tau_C)(1 - \tau_B)$ and $1 - \overline{\overline{\tau}}_L = (1 - \tau_L)/(1 - \tau_B)$.

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 $\tau_C$, 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 $\widehat{\tau}_C$, then the consequences for individual welfare are the same as with the broad tax, but the government budget constraint now looks as follows:

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

In the broad consumption tax formulation $\tau_C$, 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 $\overline{\overline{\tau}}_B = 0$, $\overline{\overline{\tau}}_C = 67\%$ and $\overline{\overline{\tau}}_L = -115\%$. 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 valuable.[82]

[82] Capital 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 valuable, 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 $\tau_B$ 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 $\tau_B$ and a labor income tax $\tau_L$, not a consumption tax $\tau_C$.

## 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 $\theta_{ti}$ matters for $\theta_{t+1i}$). 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 $\theta_{ti}$. 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

[83]Assuming $r > g$, then in steady-state we have $b_t e^{rH} > b_{t+1} = b_t e^{gH}$, i.e. $\tau_B$ 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 would be lower, i.e. if $r > g$ then $\widehat{\tau}_B < \tau_B$.

[84]This 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 $\tau_L$ and the bequest tax $\tau_B$ 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 $\tau_B$.[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

[85] As 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).

[86] To obtain meaningful theoretical formulas for inheritance flows (i.e. formulas that can be used with real numbers), we need a dynamic model with a realistic age structure. Models with infinitely lived agents or perpetual youth models will not do, and standard two or three-period OLG models will not do either. Here we follow the continuous-time OLG model introduced by Piketty (2010, sections 5-7; 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 $y_{Ltia}$
$\tau_B$ = bequest tax rate, $\tau_L$ = labor income tax rate

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

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

[87] In effect we assume a flat, cross-sectional age-productivity profile at the aggregate level. The $\lambda$ 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, $\lambda$ corrects for differences between the lifetime profiles of labor income flows vs. inheritance flows (dollars received earlier in life are worth more). When labor income flows acrue earlier in life than inheritance flows then $\lambda > 1$ (and $\lambda < 1$ conversely with early inheritance). In practice, inheritance tends to happen around mid-life, and $\lambda$ 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:

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

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:

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

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 $b_{yt}$:

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

In case assumption 3 is satisfied, then $b_{yt} \to b_y = \frac{s(1-\tau_L)\lambda(1-\alpha)e^{(r-g)H}}{1 - s(1-\tau_B)e^{(r-g)H}}$ as $t \to +\infty$. Hence, we obtain exactly the same steady-state formula as in the discrete-time, one-period model, except for the correcting factor $\lambda$ (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:

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

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

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

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

[88] For detailed empirical calibrations and theoretical extensions of the $\lambda$ 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 $\widetilde{N}_t$ 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 $\sigma$ is relatively small, the impact on our optimal capital tax results is 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 $\sigma$ 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

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

with $q_{tia}$ = individual consumption habit stock

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

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

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

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

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

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

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

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

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

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

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

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

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

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

In the Cobb-Douglas case ($V(U,w,\bar{b}) = U^{1-s_{bi}-s_{wi}} w^{s_{wi}} \bar{b}^{s_{bi}}$), the $\mu$ 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 $\mu$ term does not disappear, but the point is that it does not depend on tax rates $\tau_B$ and $\tau_L$, so in effect it cancels out from the first-order condition for optimal tax rates.[89]

[89] $\mu$ depends on $r$ and hence would depend on the annual capital income tax rate $\tau_K$ when $\tau_K > 0$ making the analysis of $\tau_K$ 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 have lower optimal savings. 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 $\mu = \dfrac{\mu_c}{\tilde{\mu}}$ given by the same formulas as before, except that one needs to replace $g_c = \sigma(r - \delta)$ by $g_c = g$ (case 2a) or $g_c = g + \sigma(r - \delta)$(case 2b).

## B.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 $\psi_t$. 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 $\tau_{Lt}, \tau_{Bt}, \tau_{Kt}$ would now vary over time, and in particular would 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 would 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 $\tau_B$. 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

[90]See Piketty (1997) for a similar steady-state multiplicity.
[91]See 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 would 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 valuable, 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 $\tau_L = \dfrac{\tau}{1-\alpha} = 43\%$. 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 $e_L$

[92]See example 4 with $x_z = 50\%$, $e_B = 0.2$; note that with $e_L > 0$, $\tau_L$ would be even smaller and $\tau_B$ 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 it is socially optimal to accumulate infinite assets in order to have zero taxes or maximal subsidies in the long run. Conversely, in case $r < r^*$, it is optimal to 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 $\tau_B$ formula with $\delta' = \delta + (\Gamma - 1)g$. This correction appears because $\tau_{Bt}$ 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 correction. Second, the formula for

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

$\tau_L$ 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 would 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):

$$SWF = \iint_{z\geq 0,\theta\geq 0} \omega_{p_z p_\theta} \frac{V_{z\theta}^{1-\Gamma}}{1-\Gamma} dz d\theta \tag{21}$$

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 $p_z, p_\theta$ in the steady-state distribution of normalized inheritance $z$ and productivity $\theta$

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

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

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

[94] See the discussion agove about Kaldor (1955).

[95] Those 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.

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

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:

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

With: $V_{tz\theta} = E(V_{ti} \mid z_{ti} = z, \theta_{ti} = \theta)$ = 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. $c_{ti} + w_{ti} \leq \widetilde{y}_{ti} = (1-\tau_{Bt}) z_{ti} b_t e^{rH} + (1-\tau_{Lt})\theta_{ti} y_{Lt}$

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 $SWF = \sum_{t\geq 0} \widetilde{V}_t e^{-\delta H t}$ might be infinite. In order to ensure that the sum converges, we need to put constraints on parameters.

First, note that for any $z, \theta$, the average utility level $V_{tz\theta}$ grows at the same rate as per capita output $y_t$ as $t \to +\infty$, 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:

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

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 $\widetilde{y}_{tz\theta} = (1-\tau_B) z b_t e^{rH} + (1-\tau_L)\theta y_{Lt}$.[99]

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

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

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

[97] In 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 $\delta$ 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.

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

[99] See 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 $\Gamma \to +\infty$, 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 $\omega_{tp_zp_\theta}$ 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 $\omega_{tp_zp_\theta} = \omega_t \cdot \omega_{p_zp_\theta}$, with:

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

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

[100] One 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.

[101] The 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 $t \to +\infty$, the social welfare of generation $t$ grows at instantaneous rate $(1-\Gamma)g + (1-\Gamma')n$:

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

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

It follows that the intertemporal sum $SWF = \sum_{t \geq 0} \widetilde{V}_t e^{-\delta H t}$ 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** $\quad \delta' = \delta - (1-\Gamma)g - (1-\Gamma')n > 0$

In what follows, we constantly assume that assumption 5 is satisfied (otherwise the intertemporal social welfare function would not be well defined). Intuitively, $\delta'$ can be viewed as the "modified" social discounted rate, i.e. the difference between the "raw" social discount rate $\delta$ 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.

[102] On 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 $\Gamma'$ makes little difference.

We now assume that the social planner seeks to maximize the intertemporal sum $SWF = \sum_{t\geq 0} \widetilde{V}_t e^{-\delta H t}$. 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 $t \to +\infty$.

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 $\tau_{Bt}$ early on and then to let $\tau_{Bt}$ decline over time. This simply comes from the fact that the short run elasticity of the bequest flow is smaller than the long run elasticity. Indeed the elasticity of the initial bequest flow $B_0$ is equal to zero: capital in on the table and can be taxed at no efficiency cost, 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 $\tau_{Bt}, \tau_{Lt}$ always involves confiscatory (or quasi-confiscatory) bequest tax rates during the first time periods, and converges towards some long run, stationary tax policy $\tau_B, \tau_L$ as $t \to +\infty$.

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 $t \to +\infty$, $\tau_{Bt} \to \tau_B(\delta') = \frac{1-(1-\alpha-\tau)s_{b0}e^{\delta' H}/b_y}{1+s_{b0}e^{\delta' H}}$ and $\tau_{Lt} \to \tau_L(\delta') = \frac{\tau - \tau_B(\delta')b_y}{1-\alpha}$*

*with: $\delta' = \delta - (1-\Gamma)g - (1-\Gamma')n =$ modified social discount factor*

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

**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 $\delta' \to 0$, 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 optimum 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 $\delta \to 0$.

Conversely, if $\delta' \to +\infty$, then $\tau_B(\delta') \to -\dfrac{1-\alpha-\tau}{b_y}$ and $\tau_L(\delta') \to 1$. 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 $\tau_{B0} = 1, \tau_{L0} = \dfrac{\tau - b_{y0}}{1-\alpha}$ to $\tau_{B1} = -\dfrac{1-\alpha-\tau}{b_y}, \tau_{L1} = 1$. 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 $\delta' \to +\infty$ can arise either because $\delta \to +\infty$ (the planner is infinitely impatient) or because $\Gamma \to +\infty$ (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 $\overline{\tau}_t = \tau_{Lt}(1-\alpha) + \tau_{Bt}b_{yt}$ the aggregate tax rate imposed on generation $t$, this can be rewritten as follows:

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

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 $r$ has to 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 would 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) would rise so as to restore $r > g + n$.

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

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

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

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

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

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

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 would 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:

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

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

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

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

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 $t \to +\infty$, $\tau_{Lt} \to 1$, $\tau_{Bt} \to 1$, $b_{yt} \to 0$, $\overline{\tau}_t \to 1-\alpha$, and $a_t \to \underline{a} = -\dfrac{1-\alpha-\tau}{\overline{R}} < 0$.*

(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 $t \to +\infty$, $\tau_{Lt} \to \underline{\tau}_L$, $\tau_{Bt} \to \underline{\tau}_B$, $b_{yt} \to \overline{b}_y$, $\overline{\tau}_t \to \underline{\tau} \leq 0$, and $a_t \to \overline{a} = \dfrac{\tau - \underline{\tau}}{\overline{R}} > 0$.*

(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* $t \to +\infty$, $\tau_{Bt} \to \tau_B(\delta') = \dfrac{1-(1-\alpha-\tau)s_{b0}e^{\delta' H}/b_y}{1+s_{b0}e^{\delta' H}}$ *and* $\tau_{Lt} \to \tau_L(\delta', a) = \dfrac{\tau - \tau_B(\delta')b_y - \overline{R}\cdot a}{1-\alpha}$

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

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

**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 $\tau_{Lt}$, $\tau_{Bt}$ 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 ($b_{yt} \to 0$, i.e. the domestic capital stock is entirely owned by foreigners, just like the public debt), and the aggregate tax rate converges towards $\overline{\tau} = 1 - \alpha$ (i.e. the labor share is taxed at 100%). By assumption 4, this is sufficient to cover public spending (i.e. $\overline{\tau} = 1 - \alpha > \tau$), and extra tax revenue $\overline{\tau} - \tau$ allows the government to finance its debt service and stabilize its (negative) asset position at $\underline{a} = -\frac{1-\alpha-\tau}{\overline{R}} < 0$.

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 $\tau_{Lt}$, $\tau_{Bt}$ 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 $\overline{b}_y$ and aggregate tax rate $\underline{\tau}$. This in turn determines the long run positive asset position $\overline{a} = \frac{\tau - \underline{\tau}}{\overline{R}}$. So for instance if $\underline{\tau}_L = \underline{\tau}_B = \underline{\tau} = 0$, then $a = \frac{\tau}{\overline{R}}$.

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 ($a_t \to +\infty$) so as to be able to distribute infinite subsidies ($\tau_{Lt} \to -\infty$, $\tau_{Bt} \to -\infty$). Private wealth accumulation would also follow an exploding path. That is, as $t \to +\infty$, $s(1 - \tau_{Bt})e^{(r-g-n)H} > 1$, and therefore $b_{yt} \to +\infty$. 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 $(\tau_B, \tau_L)$. 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 $-\overline{R} \cdot a$ term now entering into the $\tau_L$ formula.[105]

[105]I.e. if $a > 0$ then $\tau_L$ 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: $\beta_t = \dfrac{K_t}{Y_t} = b_{yt} e^{-r_t H} + a_t =$ 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 $t \to +\infty$, $r_t \to r^* = \delta + \Gamma' n + \Gamma g$. (or, equivalently, $1 + R_t = e^{r_t H} \to 1 + R^* = e^{r^* H} = (1+\Delta)(1+N)^{\Gamma'}(1+G)^{\Gamma}$). That is, the government will accumulate assets until the point where $\beta_t \to \beta^* = \dfrac{\alpha}{R^*}$.

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

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

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

Under this assumption, the intertemporal social optimum necessarily involves $r_t \to r^*$ and $\beta_t \to \beta^*$. 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 $\tau_L$ 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:* $r_t \to r^* = \delta + \Gamma' n + \Gamma g$. *The capital-output ratio converges towards the corresponding level:* $\beta_t = b_{yt} e^{-r_t H} + a_t \to \beta^* = \dfrac{\alpha}{R^*} = \dfrac{\alpha}{e^{r^* H} - 1}$.

*(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:*

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

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

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

**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 ($a_t \to a > 0$) or the accumulation of public debt ($a_t \to a < 0$) ? 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 $\beta^* = \dfrac{\alpha}{R^*} = \dfrac{\alpha}{e^{r^* H} - 1}$ (with $r^* = \delta + \Gamma' n + \Gamma g$) depends on the capital share $\alpha$ and on the parameters of the social welfare function $\delta, \Gamma', \Gamma$ . Typically, a more patient planner ($\delta \to 0$) will accumulate more capital, while a more concave planner ($\Gamma \to +\infty$) 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 ($\beta^* \to 0$ as $\Gamma \to +\infty$ , as long as $g > 0$), since they will be richer than us anyway.[106] Conversely, in case

[106] Of 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 $\Gamma$ 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 $\delta$) 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 $\Gamma$: 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: $\beta^* \to +\infty$ as $\delta \to 0$. Note that the socially optimal capital-output ratio $\beta^*$ 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 $\tau_L, \tau_B$, we have: $b_{yt} \to b_y = \dfrac{s(1-\tau_L)(1-\alpha)e^{(r-g-n)H}}{1-s(1-\tau_B)e^{(r-g-n)H}}$.

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

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

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 $\widehat{\beta}$ depends positively on the saving taste $s$ and negatively on the growth rate $g+n$. In case $s \to 1$ and $g+n \to 0$, then $\widehat{\beta} \to +\infty$. Conversely, in case $s \to 0$ then unsurprisingly $\widehat{\beta} \to 0$.

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

In case $\widehat{\beta} > \beta^*$, 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: $a_t \to a < 0$.

Conversely, in case $\widehat{\beta} < \beta^*$, 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: $a_t \to a > 0$.

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 $\widehat{\beta}$ are so large than $\widehat{r} < g+n$. 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 $\alpha < \widetilde{s}$ (one simply needs to multiply both sides by $\beta$). 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 would optimally 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

[107]The 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.

[108]In 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).

[109]The 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 optimal return to capital has to 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

to make 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 $\delta, \Gamma', \Gamma$, 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 $\tau_B(\delta') = \dfrac{1-(1-\alpha-\tau)s_{b0}e^{\delta' H}/b_y}{1+s_{b0}e^{\delta' H}}$ 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: $\forall t \geq 0$, $\tau_{Lt} = \dfrac{\tau - \tau_{Bt} b_{yt}}{1-\alpha}$.

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

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

With: $\widetilde{v}_t = v_t \cdot (1-\tau_{Lt}) \cdot (1-\alpha) \cdot \widetilde{\theta} = v_t \cdot (1-\alpha-\tau+\tau_{Bt}b_{yt}) \cdot \widetilde{\theta}$
$v_t = [E(v_{ti}^{1-\Gamma})]^{1/(1-\Gamma)}$,
$v_{ti} = (1-s_i)^{1-s_i} s_i^{s_i} [(1-\tau_{Bt+1})e^{rH}]^{s_{bi}}$
$\widetilde{\theta} = [E(\theta_i^{1-\Gamma})]^{1/(1-\Gamma)}$

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

$\Gamma' = 0$, 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 $c_t = (1-s) \cdot f(\widetilde{k}) \cdot e^{gHt} = [f(\widetilde{k}) - (n+g) \cdot \widetilde{k}] \cdot e^{gHt}$, with $\widetilde{k}_t = K_t/L_t$ = capital per efficiency labor unit, which yields $r^* = f'(\widetilde{k}^*) = n + g$. In effect this is assuming a social welfare function with a logarithmic form ($\Gamma = 1$).

[110] This 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 than 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).

[111] Note 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 $\delta' = \delta - (1-\Gamma)g - (1-\Gamma')n \to 0$, then $r^* = \delta + \Gamma g + \Gamma' n \to g + n$. 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 $\tau_{L0} = \dfrac{\tau - b_{y0}}{1-\alpha}$. This follows from the fact that $\tau_{B0}$ only enters -positively- into generation-0 social welfare $\widetilde{V}_0$. I.e. at time $t = 0$, capital is on the table and can be taxed as much as possible).

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

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

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

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

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

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

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

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

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

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 $t \to +\infty$, $b_{yt} \to b_y = \dfrac{s(1-\alpha-\tau)e^{(r-g-n)H}}{1-s\cdot e^{(r-g-n)H}}$. 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 $t \to +\infty$, we also have $\widetilde{v}_t \to \widetilde{v} = v \cdot (1-\alpha-\tau+\tau_B b_y) \cdot (1-\alpha) \cdot \widetilde{\theta}$, with $v = \left[E(v_i^{1-\Gamma})\right]^{1/(1-\Gamma)}$ and $v_i = (1-s_i)^{1-s_i} s_i^{s_i} [(1-\tau_B)e^{rH}]^{s_{bi}}$, and $s_{b0t} \to s_{b0} = \dfrac{E(v_i^{1-\Gamma} \cdot s_{bi})}{E(v_i^{1-\Gamma})}$.

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

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

**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 $t \to +\infty$, $\tau_{Bt} \to \tau_B$, $\tau_{Lt} \to \tau_L$, and $b_{yt} \to b_y$.

Consider a small, permanent change in the bequest tax rate occurring after some time $t_0 > 0$. I.e. for $t \geq t_0$, $\tau_{Bt}$ 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: $d\tau_{Lt} = -\dfrac{b_{yt} d\tau_B + \tau_{Bt} db_{yt}}{1-\alpha}$.

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

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

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

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

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

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

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

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

We have:. $\dfrac{d\widetilde{v}}{d\tau_B} > 0$ 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 $\tau_B$ if it is smaller than $\tau_B^*$, and by reducing the long run bequest tax rate $\tau_B$ if it is larger than $\tau_B^*$.

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

[112] Strictly 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 would be quantitatively very close for small $\delta'$. When $\delta' \to 0$ then naturally $\bar{e}_B \to e_B$ and we recover exactly the same formula as in the main text.

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

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

With: $d\widetilde{v}_{t_0-1} = -s_{b0t_0-1} \cdot \widetilde{v}_{t_0-1} \cdot \dfrac{d\tau_B}{1-\tau_{Bt_0}}$

E.g. if $\tau_B < \tau_B^*$, then raising $\tau_{Bt}$ to $\tau_{Bt} + d\tau_B$ for all $t \geq t_0$ will increase welfare $\widetilde{V}_t$ of all generations $t \geq t_0$ (for $t_0$ large enough), but will reduce the welfare $\widetilde{V}_{t_0-1}$ 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 $\tau_B^*$. 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 $\tau_B^*$. To see this, one can re-arrange $dSWF$ in the following way:

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

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

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

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

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

We have:$\xi > 0$ 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 $\tau_{Bt}, \tau_{Lt}$ to the sequence $\tau_{Bt} + d\tau_B, \tau_{Lt} - d\tau_{Lt}$ for $t \geq t_0$ does raise intertemporal social welfare:

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

This contradicts the fact that the tax policy sequence $\tau_{Bt}, \tau_{Lt}$ 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 $\tau_{Bt}, \tau_{Lt}$ 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 $(\tau_B, \tau_L)$ as $t \to +\infty$. Under assumptions 1-6, $b_{yt} \to b_y = \dfrac{s(1-\tau_L)(1-\alpha)e^{(r-g)H}}{1-s(1-\tau_B)e^{(r-g)H}}$, and $\overline{\tau}_t \to \overline{\tau} = \tau_L(1-\alpha) + \tau_B b_y$ Assume that $\tau_L < 1$.

Consider a small, budget balanced tax change whereby the planner reduces the labor tax rate from $\tau_{Lt_0}$ to $\tau_{Lt_0} - d\tau$ at time $t_0 \geq 0$ (with $d\tau > 0$) and raises the labor tax rate from $\tau_{Lt_1}$ to $\tau_{Lt_1} + d\tau'$.at some future dates $t_1 > t_0$ (with $d\tau' > 0$). The bequest tax rate sequence $\tau_{Bt}$ 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 $\widetilde{V}_t$ of generation $t$-zero bequest receivers is given by:

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

With: $\widetilde{v}_t = v_t \cdot (1-\tau_{Lt}) \cdot (1-\alpha) \cdot \widetilde{\theta}$
$v_t = [E(v_{ti}^{1-\Gamma})]^{1/(1-\Gamma)}$,
$v_{ti} = (1-s_i)^{1-s_i} s_i^{s_i} [(1-\tau_{Bt+1})e^{rH}]^{s_{bi}}$
$\widetilde{\theta} = [E(\theta_i^{1-\Gamma})]^{1/(1-\Gamma)}$

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

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

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

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

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

$$\begin{aligned} dSWF &= e^{-\delta Ht} \cdot \frac{\widetilde{v}_t^{1-\Gamma}}{1-\tau_{Lt}} \cdot y_t^{1-\Gamma} \cdot N_t^{1-\Gamma'} \cdot d\tau \cdot [1 - \xi_{t_0,t_1} \cdot e^{[r-r^*]H(t_1-t_0)}] \\ \text{With } &: \; r^* = \delta + \Gamma g + \Gamma' n. \end{aligned}$$

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

**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 $\tau_{Lt_0}$ 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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The tax change $d\tau, d\tau'$ is budget balance if and 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:

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

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

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 $\tau_{Bt} \to \tau_B < 1$, then one can increase intertemporal social welfare by reducing the bequest tax rate from $\tau_{Bt_0}$ to $\tau_{Bt_0} - d\tau$ at some time $t_0 \geq 0$ and raising the bequest tax rate from $\tau_{Bt_1}$ 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 $\tau_{Lt} \to \tau_L = 1$ and $\tau_{Bt} \to \tau_B = 1$. Note that since $\tau_{Lt} \to \tau_L = 1$, $b_{yt} \to b_y = 0$, 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 $a_t \to a = \frac{\tau - \overline{\tau}}{\overline{R}} = -\frac{1-\alpha-\tau}{\overline{R}} < 0$.

### 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 $\tau_{Lt_0}$ to $\tau_{Lt_0} + d\tau$ at some time $t_0 \geq 0$ and reducing the labor tax rate from $\tau_{Lt_1}$ 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 $\tau_{Bt_0}$ to $\tau_{Bt_0} + d\tau$ at some time $t_0 \geq 0$ and reducing the bequest tax rate from $\tau_{Bt_1}$ to $\tau_{Bt_1} - d\tau'$ at some future date $t_1 > t_0$. It follows that if $r > r^*$, then $\tau_{Lt}, \tau_{Bt}$ 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: $a_t \to a > 0$ 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:

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

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:

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

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 $\widetilde{V}_t$ of generation $t$-zero bequest receivers, total social welfare $SWF$, and marginal welfare changes $d\widetilde{V}_t$ 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 $\tau_{L0} = \dfrac{\tau - b_{y0} + a_1 e^{(g+n)} - a_0 e^{rH}}{1-\alpha}$. This follows from the fact that $\tau_{B0}$ only enters -positively- into generation-0 social welfare $\widetilde{V}_0$. I.e. at time $t = 0$, capital is on the table and can again be taxed as much as possible).

For $t > 0$, we have a non-generate trade-off, since $\tau_{Bt}$ enters positively into generation-$t$ social welfare $\widetilde{V}_t$ and negatively into generation-$t-1$ social welfare $\widetilde{V}_{t-1}$.

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

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

As $t \to +\infty$, we again have:

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

For the asymptotic labor tax rate, the only difference with the previous formula is that we now have a $-\overline{R} \cdot a$ term:

$$\tau_{Lt} = \frac{\tau - \tau_{Bt} b_{yt} + a_{t+1} e^{(g+n)H} - a_t e^{rH}}{1-\alpha} \to \tau_L(\delta') = \frac{\tau - \tau_B(\delta') b_y - \overline{R} \cdot a}{1-\alpha}$$

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 Unegalitarian 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.