

Chapter 6 Inheritance and Gift Taxation

6.1 Introduction¹

Death taxes may be imposed for a variety of reasons. One of them, on which this paper focuses, is redistribution. The institution of inheritance is a major factor responsible for concentration of wealth and, indirectly, for income inequality. According to most recent estimates, inherited wealth accounts for almost half of the net worth of households². Death taxes, especially in the form of inheritance taxation, can thus (at least potentially) be used to moderate economic inequalities.

There are, however, more subtle effects of death taxation that might be significant; they are stressed by those who are less enthusiastic about such taxation. In addition to the alleged reduction of savings, there is the concern that death taxation may adversely affect equality. According to Becker (1974, 1991) and Tomes (1981), transfers between generations follow a ‘regression towards the mean’ mechanism with bequests and gifts flowing from well-to-do donors to less well-to-do recipients. Within each family, transfers thus tend to offset inequalities. Where this is true, taxation will mitigate the redistributive effect of wealth transfers.

The question one might raise at this point is why the government couldn’t directly effect the appropriate redistribution across and within generations. In a perfect information setting, it is clear that taxes on wealth transfers could be tuned in such a way that redistribution within families is not discouraged while redistribution across families is fostered. In an asymmetric information setting, however, this is less clear. Well to do families could be induced to leave lower bequests to avoid a too heavy tax burden; at the same time, the government could not be able to implement the ‘right’ redistribution because of imperfect information as to wealth holding.

It should be pointed out that the government can affect bequest behavior not only through the tax schedule, but also through the choice of the tax base or even through restrictions on estate sharing. Polar cases include estate taxes (based on the total amount which is bequeathed) and inheritance taxes (based on individual shares) or even accession taxes (based on individual shares plus other resources). Because tax schedules are typically non-linear (and progressive) the definition of the tax base is of crucial importance. In addition, bequests are subject to a variety of legal rules including more or less stringent equal sharing rules (amongst children).

¹ Section 1-3 draw heavily from Cremer and Pestieau (2001)

² For a survey of empirical studies on bequests, see Arrondel et al. (1997).

6.2 Optimal Inheritance Tax with Bequests in the Utility

6.2.1 Model³

We consider a dynamic economy with a discrete set of generations $0, 1, \dots, t, \dots$ and no growth. Each generation has measure 1, lives one period, and is replaced by the next generation. Individual ti (from dynasty i living in generation t) receives pre-tax inheritance $b_{ti} \geq 0$ from generation $t-1$ at the beginning of period t . The initial distribution of bequests b_{0i} is exogenously given. Inheritances earn an exogenous gross rate of return R per generation. We relax the no-growth and small open economy fixed factor price assumptions at the end of Section 6.2.3.

Individual Maximization

Individual ti has exogenous pre-tax wage rate w_{ti} , drawn from an arbitrary but stationary ergodic distribution (with potential correlation of individual draws across generations). Individual ti works l_{ti} , and earns $y_{Lti} = w_{ti}l_{ti}$ at the end of period and then splits lifetime resources (the sum of net-of-tax labor income and capitalized bequests received) into consumption c_{ti} and bequests left $b_{t+1i} \geq 0$. We assume that there is a linear labor tax at rate τ_{Lt} , a linear tax on capitalized bequests at rate τ_{Bt} , and a lump-sum grant E_t ⁴. Individual ti has utility function

$V^{ti}(c, \underline{b}, l)$ increasing in consumption $c = c_{ti}$ and net-of-tax capitalized bequests left

$\underline{b} = Rb_{t+1i}(1 - \tau_{Bt+1})$, and decreasing in labor supply $l = l_{ti}$. Like w_{ti} , preferences V^{ti} are also drawn from an arbitrary ergodic distribution. Hence, individual ti solves

$$\max_{l_{ti}, c_{ti}, b_{t+1i} \geq 0} V^{ti}(c_{ti}, Rb_{t+1i}(1 - \tau_{Bt+1}), l_{ti}) \quad \text{s.t.} \quad (1)$$

$$c_{ti} + b_{t+1i} = Rb_{ti}(1 - \tau_{Bt}) + w_{ti}l_{ti}(1 - \tau_{Lt}) + E_t.$$

The individual first order condition for bequests left b_{t+1i} is $V_c^{ti} = R(1 - \tau_{Bt+1})V_{\underline{b}}^{ti}$ if $b_{t+1i} > 0$.

Equilibrium Definition

³ This part is drawn heavily from Piketty and Saez (2013, pp.1853-73).

⁴ Note that τ_{Bt} taxes both the raw bequest received b_{ti} and the lifetime return to bequest $(R-1) \cdot b_{ti}$, so it should really be interpreted as a broad-based capital tax rather than as a narrow inheritance tax.

We denote by b_t, c_t, y_{Lt} aggregate bequests received, consumption, and labor income in generation t . We assume that the stochastic processes for utility functions V^i and for wage rates w_{it} are such that, with constant tax rates and lump-sum grant, the economy converges to a unique ergodic steady-state equilibrium independent of the initial distribution of bequests $(b_{0i})_i$. All we need to assume is an ergodicity condition for the stochastic process for V^i and w_{it} . Whatever parental taste and ability, one can always draw any other taste or productivity.⁵ In equilibrium, all individuals maximize utility as in (1) and there is a resulting steady-state ergodic equilibrium distribution of bequests and earnings (b_{it}, y_{L0i}) . In the long run, the position of each dynasty i is independent of the initial position (b_{0i}, y_{L0i}) .

6.2.2 Steady-State Welfare Maximization

For pedagogical reasons, we start with the case where the government considers the long-run steady-state equilibrium of the economy and chooses steady-state long-run policy E, τ_L, τ_B to maximize *steady-state* social welfare, defined as a weighted sum of individual utilities with Pareto weights $\omega_i \geq 0$, subject to a period-by-period budget balance

$$E = \tau_B Rb_t + \tau_L y_{Lt} :$$

$$SWF = \max_{\tau_L, \tau_B} \int \omega_i V^i (Rb_{it}(1 - \tau_B) + \omega_i l_{it}(1 - \tau_L) + E - b_{it+1}, Rb_{it+1}(1 - \tau_B), l_{it}). \quad (2)$$

In the ergodic equilibrium, social welfare is constant over time. Taking the lump-sum grant E as fixed, τ_L and τ_B are linked to meet the budget constraint, $E = \tau_B Rb_t + \tau_L y_{Lt}$. As we shall see, the optimal τ_B depends on the size of behavioral responses to taxation captured by elasticities, and the combination of social preferences and the distribution of bequests and earnings captured by distributional parameters, which we introduce in turn.

Elasticity Parameters

The aggregate variable b_t is a function of $1 - \tau_B$ (assuming that τ_L adjusts), and y_{Lt} is a function of $1 - \tau_L$ (assuming that τ_B adjusts). Formally, we can define the corresponding long-run elasticities as

⁵ See Piketty and Saez (2012) for a precise mathematical statement and concrete examples. Random taste shocks can generate Pareto distributions with realistic levels of wealth concentration – which are difficult to generate with labor productivity shocks alone. Random shocks to rates of return would work as well.

$$\begin{aligned} \text{Long-run Elasticities: } e_B &= \frac{1-\tau_B}{b_t} + \frac{db_t}{d(1-\tau_B)} \Big|_E \quad \text{and} \\ e_L &= \frac{1-\tau_L}{y_{Lt}} + \frac{dy_{Lt}}{d(1-\tau_L)} \Big|_E . \end{aligned} \tag{3}$$

That is, e_B is the long-run elasticity of aggregate bequest flow (i.e., aggregate capital accumulation) with respect to the net-of-bequest-tax rate $1-\tau_B$, while e_L is the long-run elasticity of aggregate labor supply with respect to the net-of-labor-tax rate $1-\tau_L$. Importantly, those elasticities are *policy elasticities* (Hendren (2013)) that capture responses to a joint and budget neutral change (τ_B, τ_L) . Hence, they incorporate both own- and cross-price effects. Empirically, e_L and e_B can be estimated directly using budget neutral joint changes in (τ_L, τ_B) or indirectly by decomposing e_L and e_B into own- and cross-price elasticities, and estimating these separately.

Distributional Parameters

We denote by $g_{ti} = \omega_i V_c^{ii} / \int_j \omega_j V_c^{jj}$ the *social marginal welfare weight* on individual ti .

The weights g_{ti} are normalized to sum to 1. g_{ti} measures the social value of increasing consumption of individual ti by \$1 (relative to distributing the \$1 equally across all individuals). Under standard redistributive preferences, g_{ti} is low for the well-off (those with high bequests received or high earnings) and high for the worse-off. To capture *distributional parameters* of earnings, bequests received, bequests left, we use the ratios – denoted with an upper bar – of the population average weighted by social marginal welfare weights g_{ti} to the unweighted population average (recall that the g_{ti} weights sum to 1). Formally, we have

$$\begin{aligned} \text{Distributional Parameters: } \bar{b}^{\text{received}} &= \frac{\int_i g_{ti} b_{ti}}{b_t}, \quad \bar{b}^{\text{left}} = \frac{\int_i g_{ti} b_{t+1i}}{b_{t+1}}, \quad \text{and} \\ \bar{y}_L &= \frac{\int_i g_{ti} y_{Lti}}{y_{Lt}} . \end{aligned} \tag{4}$$

Each of those ratios is below 1 if the variable is lower for those with high social marginal welfare weights. With standard redistributive preferences, the more concentrated the variable is among the well-off, the lower the distributional parameter.

Optimal τ_B Derivation

To obtain a formula for the optimal τ_B (taking τ_L as given), we consider a small reform $d\tau_B > 0$. Budget balance with $dE = 0$ requires $d\tau_L < 0$ such that $Rb_t d\tau_B + \tau_B Rdb_t + y_{Lt} d\tau_L + \tau_{Lt} dy_{Lt} = 0$. Using the elasticity definitions (3), this implies

$$Rb_t d\tau_B \left(1 - e_B \frac{\tau_B}{1 - \tau_B} \right) = -d\tau_L y_{Lt} \left(1 - e_L \frac{\tau_L}{1 - \tau_L} \right). \quad (5)$$

Using the fact that b_{t+1i} and l_{ti} are chosen to maximize individual utility, and applying the envelope theorem, the effect of the reform $d\tau_B, d\tau_L$ on steady-state social welfare (2) is

$$dSWF = \int_i \omega_{ii} V_c^{ii} \cdot (Rdb_{ii}(1 - \tau_B) + Rb_{ii} d\tau_B - d\tau_L y_{Lti}) + \omega_{ii} V_b^{ii} \cdot (-d\tau_B Rb_{t+1i}).$$

At the optimum τ_B , $dSWF = 0$. Using the individual first order condition $V_c^{ii} = R(1 - \tau_B)V_b^{ii}$

when $b_{t+1i} > 0$, expression (5) for $d\tau_L$, and the definition of $g_{ii} = \omega_{ii} V_c^{ii} / \int_j \omega_{ij} V_c^{ij}$, we have

$$0 = \int_j g_{ij} \cdot \left(-d\tau_B Rb_{ii}(1 + e_{Bii}) + \frac{1 - e_B \tau_B / (1 - \tau_B)}{1 - e_L \tau_L / (1 - \tau_L)} \frac{y_{Lti}}{y_{Lt}} Rb_t d\tau_B - d\tau_B \frac{b_{t+1i}}{1 - \tau_B} \right), \quad (6)$$

Where we have expressed db_{ii} using $e_{Bii} = \frac{1 - \tau_B}{b_{ii}} \frac{db_{ii}}{d(1 - \tau_B)} \Big|_E$ the individual elasticity of

bequest received (e_B is the bequest-weighted population average of e_{Bii}).

The first term in (6) captures the negative effect of $d\tau_B$ on bequest received (the direct effect and the dynamic effect via reduced pre-tax bequests), the second term captures the positive effect of reduced labor income tax, and the third term captures the negative effect on bequest leavers.

Finally, let \hat{e}_B be the average of e_{Bii} weighted by $g_{ii} b_{ii}$ ⁶. Dividing (6) by $Rb_t d\tau_B$, and

⁶ \hat{e}_B is equal to e_B (b_{ii} -weighted average of e_{Bii}) if individual bequest elasticities are uncorrelated with g_{ii} .

using the distributional parameters from (4), the first order condition (6) can be rewritten as

$$0 = -\bar{b}^{\text{received}}(1 + \hat{e}_B) + \frac{1 - e_B \tau_B / (1 - \tau_B)}{1 - e_L \tau_L / (1 - \tau_L)} \bar{y}_L - \frac{\bar{b}^{\text{left}}}{R(1 - \tau_B)},$$

hence, re-arranging, we obtain.

Steady-state Optimum: For a given τ_L , the optimal tax rate τ_B that maximizes long-run steady-state social welfare with period-by-period budget balance is given by

$$\tau_B = \frac{1 - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \cdot \left[\frac{\bar{b}^{\text{received}}}{\bar{y}_L} (1 + \hat{e}_B) + \frac{1}{R} \frac{\bar{b}^{\text{left}}}{\bar{y}_L} \right]}{1 + e_B - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \frac{\bar{b}^{\text{received}}}{\bar{y}_L} (1 + \hat{e}_B)}, \quad (7)$$

with e_B and e_L the aggregate elasticities of bequests and earnings with respect to $1 - \tau_B$ and $1 - \tau_L$ defined in (3), and with $\bar{b}^{\text{received}}$, \bar{b}^{left} , and \bar{y}_L the distributional parameters defined in (4).

Five important points are worth noting about the economics behind formula (7):

1. *Role of R .* The presence of R in formula (7) is a consequence of steady-state maximization, that is, no social discounting. As shown in Section 6.2.3, with social discounting at rate $\Delta < 1$, R should be replaced by $R\Delta$. Furthermore, in a closed economy with government debt, dynamic efficiency implies that the Modified Golden Rule, $R\Delta = 1$, holds. Hence, formula (7) continues to apply in the canonical case with discounting and dynamic efficiency by replacing R by 1 in equation (7). This also remains true with exogenous economic growth. Therefore, if one believes that the natural benchmark is dynamic efficiency and no social discounting ($\Delta = 1$), then formula (7) can be used with $R = 1$. As we shall discuss, it is unclear, however, whether this is the most relevant case for numerical calibrations.
2. *Endogeneity of right-hand-side parameters.* As with virtually all optimal tax formulas e_B , e_L , \bar{b}^{left} , $\bar{b}^{\text{received}}$, and \bar{y}_L depend on tax rates τ_B , τ_L and hence are endogenous⁷. For calibration, assumptions need to be made on how those parameters vary with tax rates. Formula (7) can also be used to evaluate bequest tax reform around current tax rates. If current τ_B is lower than (7), then it is desirable to increase τ_B (and decrease τ_L) and vice versa. Formula (7) is valid for any τ_L meeting the government budget (and does not require τ_L to be optimal).

⁷ Multiple tax equilibria might also satisfy formula (7), with only one characterizing the global optimum.

3. *Comparative statics.* τ_B decreases with the elasticity e_B for standard efficiency reasons and increases with e_L as a higher earnings elasticity makes it more desirable to increase τ_B to reduce τ_L . τ_B naturally decreases with the distributional parameters $\bar{b}^{\text{received}}$ and \bar{b}^{left} , that is, the social weight put on bequests receivers and leavers. Under a standard utilitarian criterion with decreasing marginal utility of disposable income, welfare weights g_{ii} are low when bequests and/or earnings are high. As bequests are more concentrated than earnings (Piketty (2011)), we expect $\bar{b}^{\text{received}} < \bar{y}_L$ and $\bar{b}^{\text{left}} < \bar{y}_L$. When bequests are infinitely concentrated, $\bar{b}^{\text{received}}, \bar{b}^{\text{left}} \ll \bar{y}_L$ and (7) boils down to $\tau_B = 1/(1 + e_B)$, the revenue maximizing rate. Conversely, when the g_{ii} 's put weight on large inheritors, then $\bar{b}^{\text{received}} > 1$ and τ_B can be negative.
4. *Pros and cons of taxing bequests.* Bequest taxation differs from capital taxation in a standard OLG model with no bequests in two ways. First, τ_B hurts both donors (\bar{b}^{left} effect) and donees ($\bar{b}^{\text{received}}$ effect), making bequests taxation relatively less desirable. Second, bequests introduce a new dimension of lifetime resources inequality, lowering $\bar{b}^{\text{received}} / \bar{y}_L$, $\bar{b}^{\text{left}} / \bar{y}_L$ and making bequests taxation more desirable. This intuition is made precise in Section 6.2.4 where we specialize our model to the Farhi-Werning two-period case with uni-dimensional inequality.
5. *General social marginal welfare weights.* General social marginal welfare weights allow great flexibility in the social welfare criterion choice (Saez and Stantcheva (2013)). One normatively appealing concept is that individuals should be compensated for inequality they are not responsible for – such as bequests received – but not for inequality they are responsible for – such as labor income (Fleurbaey (2008)). This amounts to setting social welfare weights g_{ii} to zero for all bequest receivers and setting them positive and uniform on zero-bequests receivers. About half the population in France or the United States receives negligible bequests. Hence, this “Meritocratic Rawlsian” optimum has broader appeal than the standard Rawlsian case.

Meritocratic Rawlsian Steady-State Optimum: *The optimal tax rate τ_B that maximizes long-run welfare of zero-bequests receivers with period-by-period budget balance is given by*

$$\tau_B = \frac{1 - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \frac{1}{R} \frac{\bar{b}^{\text{left}}}{\bar{y}_L}}{1 + e_B}, \quad (8)$$

with \bar{b}^{left} , \bar{y}_L the ratios of average bequests left and earnings of zero-receivers to

population averages.

In that case, even when zero-receivers have average labor earnings (i.e., $\bar{y}_L = 1$), if bequests are quantitatively important in lifetime resources, zero-receivers will leave smaller bequests than average, so that $\bar{b}^{\text{left}} < 1$. Formula (8) then implies $\tau_B > 0$ even with $R = 1$ and $e_L = 0$.

In the inelastic labor case, formula (8) further simplifies to $\tau_B = \frac{1 - \bar{b}^{\text{left}} / (R\bar{y}_L)}{1 + e_B}$. If we further assume $e_B = 0$ and $R = 1$ (benchmark case with dynamic efficiency and $\Delta = 1$), the optimal tax rate $\tau_B = 1 - \frac{\bar{b}^{\text{left}}}{\bar{y}_L}$ depends only on distributional parameters, namely the relative position of zero-bequest receivers in the distributions of bequests left and labor income. For instance, if $\bar{b}^{\text{left}} / \bar{y}_L = 50\%$, for example, zero-bequest receivers expect to leave bequests that are only half of average bequest and to receive average labor income, then it is in their interest to tax bequests at rate $\tau_B = 50\%$. Intuitively, with a 50% bequest tax rate, the distortion on the “bequest left” margin is so large that the utility value of one additional dollar devoted to bequests is twice larger than one additional dollar devoted to consumption. For the same reasons, if $\bar{b}^{\text{left}} / \bar{y}_L = 100\%$, but $R = 2$, then $\tau_B = 50\%$. If the return to capital doubles the value of bequests left at each generation, then it is in the interest of zero-receivers to tax capitalized bequest at a 50% rate, even if they plan to leave as many bequests as the average. These intuitions illustrate the critical importance of distributional parameters – and also of perceptions. If everybody expects to leave large bequests, then subjectively optimal τ_B will be fairly small – or even negative.

6.2.3 Social Discounting, Government Debt, and Dynamic Efficiency

In this section, the government chooses policy $(\tau_{Bt}, \tau_{Lt})_t$ to maximize a discounted stream of social welfare across periods with generational discount rate $\Delta \leq 1$ (Section 6.2.2 was the special case $\Delta = 1$). We derive the long-run optimum τ_B , that is, when all variables have converged:

$$SWF = \sum_{t \geq 0} \Delta^t \int_i \omega_{it} V^i (Rb_{it}(1 - \tau_{Bt}) + \omega_{it} l_{it}(1 - \tau_{Lt}) + E_t - b_{t+1i}, Rb_{t+1i}(1 - \tau_{Bt+1}), l_{it}).$$

Budget Balance and Open Economy

Let us first keep period-by-period budget balance, so that $E_t = \tau_{Bt} Rb_t + \tau_{Lt} y_{Lt}$, along with the open economy R exogenous assumption. Consider again a reform $d\tau_B$ so that $d\tau_{Bt} = d\tau_B$ for all $t \geq T$ (and correspondingly $d\tau_{Lt}$ to maintain budget balance and keeping E_t constant) with T large (so that all variables have converged),

$$dSWF = \sum_{t \geq T} \Delta^t \int_i \omega_{ii} V_c^{ii} \cdot (Rdb_{ii}(1 - \tau_B) - Rb_{ii}d\tau_B - d\tau_{Lt} y_{Lti}) \\ + \sum_{t \geq T-1} \Delta^t \int_i \omega_{ii} V_b^{ii} \cdot (-d\tau_B Rb_{t+1i}).$$

In contrast to steady-state maximization, we have to sum effects for $t \geq T$. Those terms are not identical, as the response to the permanent small tax change might build across generations $t \geq T$. However, we can define average discounted elasticities e_B, \hat{e}_B, e_L to parallel our earlier analysis. The necessity of defining such discounted elasticities complicates the complete presentation of the discounted welfare case relative to steady-state welfare maximization. The key additional difference with steady-state maximization is that the reform starting at T also hurts generation $T - 1$ bequest leavers. We formally derive the following formula:

Long-Run Optimum with Social Discounting: *The optimal long-run tax rate τ_B that maximizes discounted social welfare with period-by-period budget balance is given by*

$$\tau_B = \frac{1 - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \cdot \left[\frac{\bar{b}^{\text{received}}}{\bar{y}_L} (1 + \hat{e}_B) + \frac{1}{R\Delta} \frac{\bar{b}^{\text{left}}}{\bar{y}_L} \right]}{1 + e_B - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \frac{\bar{b}^{\text{received}}}{\bar{y}_L} + (1 + \hat{e}_B)}, \quad (9)$$

with e_B, \hat{e}_B , and e_L the discounted aggregate bequest and earnings elasticities, and with $\bar{b}^{\text{received}}, \bar{b}^{\text{left}}$, and \bar{y}_L defined in (4).

The only difference with (7) is that R is replaced by $R\Delta$ in the denominator of the term, reflecting the utility loss of bequest leavers. The intuition is transparent: the utility loss of bequest leavers has a multiplicative factor $1/\Delta$ because bequest leavers are hurt one generation in advance of the tax reform. Concretely, a future inheritance tax increase 30 years away does not generate any revenue for 30 years and yet already hurts the current adult population who will leave bequests in 30 or more years. Naturally, with $\Delta = 1$, formulas (7) and (9) coincide.

Government Debt in the Closed Economy

Suppose now that the government can use debt (paying the same rate of return R) and hence can transfer resources across generations. Let a_t be the net asset position of the government. If $R\Delta > 1$, reducing consumption of generation t to increase consumption of generation $t + 1$ is

desirable (and vice versa). Hence, if $R\Delta > 1$, the government wants to accumulate infinite assets. If $R\Delta < 1$, the government wants to accumulate infinite debts. In both cases, the small open economy assumption would cease to hold. Hence, a steady-state equilibrium only exists if the *Modified Golden Rule* $R\Delta = 1$ holds.

Therefore, it is natural to consider the closed-economy case with endogenous capital stock $K_t = b_t + a_t$, CRS production function $F(K_t, L_t)$, where L_t is the total labor supply, and where rates of returns on capital and labor are given by $R_t = 1 + F_K$ and $w_t = F_L$. Denoting by $\underline{R}_t = R_t(1 - \tau_{Bt})$ and $\underline{w}_t = w_t(1 - \tau_{Lt})$ the after-tax factor prices, the government budget dynamics is given by $a_{t+1} = R_t a_t + (R_t - \underline{R}_t)b_t + (w_t - \underline{w}_t)L_t - E_t$. Two results can be obtained in that context.

First, going back or an instant to the budget balance case, it is straight-forward to show that formula (9) carries over unchanged in this case. This is a consequence of the standard optimal tax result of Diamond and Mirrlees (1971) that optimal tax formulas are the same with fixed prices and endogenous prices. The important point is that the elasticities e_B and e_L are pure supply elasticities (i.e., keeping factor prices constant). Intuitively, the government chooses the net-of-tax prices \underline{R}_t and \underline{w}_t and the resource constraint is $0 = b_t + F(b_t, L_t) - \underline{R}_t b_t - \underline{w}_t L_t - E_t$, so that the pre-tax factors effectively drop out of the maximization problem and the same proof goes through (see the Supplemental Material (Piketty and Saez (2013b)) for complete details). Second, and most important, moving to the case with debt, we can show that the long-run optimum takes the following form.

Long-Run Optimum with Social Discounting, Closed Economy, and Government Debt: *In the long-run optimum, the Modified Golden Rule holds, so that $R\Delta = 1$. The optimal long-run tax rate τ_B continues to be given by formula (9) with $R\Delta = 1$,*

$$\tau_B = \frac{1 - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \cdot \left[\frac{\bar{b}^{\text{received}}}{\bar{y}_L} (1 + \hat{e}_B) + \frac{\bar{b}^{\text{left}}}{\bar{y}_L} \right]}{1 + e_B - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \frac{\bar{b}^{\text{received}}}{\bar{y}_L} + (1 + \hat{e}_B)}, \quad (10)$$

Proof: We first establish that the Modified Golden Rule holds in the long run. Consider a small reform $d\underline{w}_T = d\underline{w} > 0$ for a single T large (so that all variables have converged). Such a reform has an effect $dSWF$ on discounted social welfare (measured as a period T) and da on long-term government debt (measured as of period T). Both $dSWF$ and da are proportional to $d\underline{w}$.

Now consider a second reform $d\underline{w}_{T+1} = R d\underline{w} < 0$ at $T+1$ only. By linearity of small changes, this reform has welfare effect $dSWF' = -R\Delta dSWF$, as it is $-R$ times larger and happens one period after the first reform. The effect on government debt is $da' = -Rda$ measured as of period $T+1$, and hence $-da$ measured as of period T (i.e., the same absolute effect as the initial reform). Hence, the sum of the two reforms would be neutral for government debt. Therefore, if social welfare is maximized, the sum has to be neutral from a social welfare perspective as well, implying that $dSWF + dSWF' = 0$ so that $R\Delta = 1$.

Next, we can easily extend the result above that the optimal tax formula takes the same form with endogenous factor prices. Hence, (9) applies with $R\Delta = 1$. *Q.E.D.*

This result shows that dynamic efficiency considerations (i.e., optimal capital accumulation) are conceptually orthogonal to cross-sectional redistribution considerations. That is, whether or not dynamic efficiency prevails, there are distributional reasons pushing for inheritance taxation, as well as distortionary effects pushing in the other direction, resulting in an equity-efficiency trade-off that is largely independent from aggregate capital accumulation issues⁸.

One natural benchmark would be to assume that we are at the Modified Golden Rule (though this is not necessarily realistic). In that case, the optimal tax formula (10) is independent of R and Δ and depends solely on elasticities e_B, e_L and the distributional factors $\bar{b}^{\text{received}}, \bar{b}^{\text{left}}, \bar{y}_L$.

If the Modified Golden Rule does not hold (which is probably more plausible) and there is too little capital, so that $R\Delta > 1$, then the welfare cost of taxing bequests left is smaller and the optimal tax rate on bequests should be higher (everything else being equal). The intuition for this result is simple: if $R\Delta > 1$, pushing resources toward the future is desirable. Taxing bequests more in period T hurts period $T-1$ bequest leavers and benefits period T labor earners, effectively creating a transfer from period $T-1$ toward period T . This result and intuition depend on our assumption that bequests left by generation $t-1$ are taxed in period t as part of generation t lifetime resources. This fits with actual practice, as bequest taxes are paid by definition at the end of the lives of bequest leavers and paid roughly in the middle of the adult life of bequest receivers⁹. If we assume instead that period t taxes are $\tau_{Bt} b_{t+1} + \tau_{Lt} y_{Lt}$, then formula (9) would have no $R\Delta$ term dividing \bar{b}^{left} , but all the terms in $\bar{b}^{\text{received}}$ would be multiplied by $R\Delta$. Hence, in the Meritocratic Rawlsian optimum where $\bar{b}^{\text{received}} = 0$, we can obtain (10) by considering steady-state maximization subject to $\tau_{Bt} b_{t+1} + \tau_{Lt} y_{Lt} = E_t$ and without the need to consider dynamic efficiency issues.

The key point of this discussion is that, with government debt and dynamic efficiency ($R\Delta = 1$), formula (10) no longer depends on the timing of tax payments.

⁸ The same decoupling results have been proved in the OLG model with only life-cycle savings with linear Ramsey taxation and a representative agent per generation (King (1980), Atkinson and Sandmo (1980)).

⁹ Piketty and Saez (2012) made this point formally with a continuum of overlapping cohorts. With accounting budget balance, increasing bequest taxes today allows to reduce labor taxes today, hurting the old who are leaving bequests and benefiting current younger labor earners (it is too late to reduce the labor taxes of the old).

Economic Growth

Normatively, there is no good justification for discounting the welfare of future generations, that is, for assuming $\Delta < 1$. However, with $\Delta = 1$, the Modified Golden Rule implies that $R = 1$ so that the capital stock should be infinite. A standard way to eliminate this unappealing result as well as making the model more realistic is to consider standard labor augmenting economic growth at rate $G > 1$ per generation. Obtaining a steady state where all variables grow at rate G per generation requires imposing standard homogeneity assumptions on individual utilities, so that $V^i(c, \underline{b}, l) = \frac{(U^i(c, \underline{b})e^{-h_i(l)})^{1-\gamma}}{1-\gamma}$, with $U^i(c, \underline{b})$ homogeneous of degree 1. In that case, labor supply is unaffected by growth. The risk aversion parameter γ reflects social value for redistribution both within and across generations¹⁰. We show that the following hold:

First, the steady-state optimum formula (7) carries over in the case with growth by just replacing R by R/G . The intuition is simple. Leaving a relative bequest b_{t+1}/b_t requires making a bequest G times larger than leaving the same relative bequest b_{t+1}/b_t . Hence, the relative cost of taxation to bequest leavers is multiplied by a factor G .

Second, with social discounting at rate Δ , marginal utility of consumption grows at rate $G^{-\gamma} < 1$, as future generations are better off and all macroeconomic variables grow at rate G . This amounts to replacing Δ by $\Delta G^{1-\gamma}$ in the social welfare calculus $dSWF$. Hence, with those two new effects, formula (9) carries over simply replacing ΔR by $\Delta(R/G)G^{1-\gamma} = \Delta R G^{-\gamma}$.

Third, with government debt in a closed economy, the Modified Golden Rule becomes $\Delta R G^{-\gamma} = 1$ (equivalent to $r = \delta + \gamma g$ when expressed in conventional net instantaneous returns). The well-known intuition is the following. One dollar of consumption in generation $T+1$ is worth $\Delta G^{-\gamma}$ dollars of consumption in generation t because of social discounting Δ and because marginal utility in generation $t+1$ is only $G^{-\gamma}$ times the marginal utility of generation t . At the dynamic optimum, this must equal the rate of return R on government debt. Hence, with the Modified Golden Rule, formula (10) carries over unchanged with growth.

Role of R and G

Which formula should be used? From a purely theoretical viewpoint, it is more natural to replace R by $\Delta R G^{-\gamma} = 1$ in formula (7), so as to entirely separate the issue of optimal capital accumulation from that of optimal redistribution. In effect, optimal capital accumulation is equivalent to removing all returns to capital in the no-growth model ($R = 1$). However, from a

¹⁰ In general, the private risk aversion parameter might well vary across individuals, and differ from the social preferences for redistribution captured by γ . Here we ignore this possibility to simplify notations.

practical policy viewpoint, it is probably more justified to replace R by R/G in formula (7) and to use observed R and G to calibrate the formula. The issue of optimal capital accumulation is very complex, and there are many good reasons why the Modified Golden Rule $\Delta RG^{-\gamma} = 1$ does not seem to be followed in the real world. In practice, it is very difficult to know what the optimal level of capital accumulation really is. Maybe partly as a consequence, governments tend not to interfere too massively with the aggregate capital accumulation process and usually choose to let private forces deal with this complex issue (net government assets – positive or negative – are typically much smaller than net private assets). One pragmatic viewpoint is to take these reasons as given and impose period-by-period budget constraint (so that the government does not interfere at all with aggregate capital accumulation), and consider steady-state maximization, in which case we obtain formula (7) with R/G .

Importantly, the return rate R and the growth rate G matter for optimal inheritance rates even in the case with dynamic efficiency. A larger R/G implies a higher level of aggregate bequest flows (Piketty (2011)), and also a higher concentration of inherited wealth. Therefore, a larger R/G leads to smaller $\bar{b}^{\text{received}}$ and \bar{b}^{left} and hence a higher τ_B .

6.2.4 Role of Bi-Dimensional Inequality: Contrast With Farhi-Werning

Our results on positive inheritance taxation (under specific redistributive social criteria) hinge crucially on the fact that, with inheritances, labor income is no longer a complete measure of lifetime resources, that is, our model has bi-dimensional (labor income, inheritance) inequality.

To see this, consider the two-period mode of Farhi and Werning (2010), where each dynasty lasts for two generations with working parents starting with no bequests and children receiving bequests and never working. In this model, all parents have the same utility function, hence earnings and bequests are perfectly correlated so that inequality is uni-dimensional (and solely due to the earnings ability of the parent). This model can be nested within the class of economies we have considered by simply assuming that each dynasty is a succession of (non-overlapping) two-period-long parent-child pairs, where children have zero wage rates and zero taste for bequests. Formally, preferences of parents have the form $V^P(c, \underline{b}, l)$, while preferences of children have the simpler form $V^C(c)$. Because children are totally passive and just consume the net-of-tax bequests they receive, parents' utility functions are de facto altruistic (i.e., depend on the utility of the child) in this model¹¹. In general equilibrium, the parents and children are in equal proportion in any cross-section. Assuming dynamic efficiency $R\Delta = 1$, our previous formula (10) naturally applies

¹¹ This assumes that children do not receive the lump-sum grant E_t (that accrues only to parents). Lump-sum grants to children can be considered as well and eliminated without loss of generality if parents' preferences are altruistic and hence take into account the lump-sum grant their children get, that is, the parents' utility is $V^{ii}(c_{ii}, Rb_{t+1i}(1 - \tau_{Bt+1}) + E_{t+1}^{child}, J_{ii})$. Farhi and Werning (2010) considered this altruistic case.

to this specific model.

Farhi and Werning (2010) analyzed the general case with nonlinear taxation with weakly separable parents' utilities of the form $U^i(u(c, \underline{b}), l)$. If social welfare puts weight only on parents (the utility of children is taken into account only through the utility of their altruistic parents), the Atkinson-Stiglitz theorem applies and the optimal inheritance tax rate is zero. If social welfare puts additional direct weight on children, then the inheritance tax is less desirable and the optimal tax rate becomes naturally negative¹². We can obtain the linear tax counterpart of these results if we further assume that the sub-utility $u(c, \underline{b})$ is homogeneous of degree 1. This assumption is needed to obtain the linear tax version of Atkinson-Stiglitz (Deaton (1979)).

Optimal Bequest Tax in the Farhi-Werning Version of Our Model: *In the parent-child model with utilities of parents such that $V^i(c, \underline{b}, l) = U^i(u(c, \underline{b}), l)$ with $u(c, \underline{b})$ homogeneous of degree 1 and homogeneous in the population and with dynamic efficiency ($R\Delta = 1$):*

- *If the social welfare function puts zero direct weight on children, then $\tau_B = 0$ is optimal.*
- *If the social welfare function puts positive direct weight on children, then $\tau_B < 0$ is optimal.*

The proof is in Piketty and Saez (2013a), where we show that any tax system (τ_B, τ_L, E) can be replaced by a tax system $(\tau'_B = 0, \tau'_L, E')$ that leaves all parents as well off and raises more revenue. The intuition can be understood using our optimal formula (10). Suppose for simplicity here that there is no lump-sum grant. With $u(c, \underline{b})$ homogeneous, bequest decisions are linear in lifetime resources so that $b_{t+1i} = s \cdot y_{Lti}(1 - \tau_L)$, where s is homogeneous in the population.

This immediately implies that $E[\omega_i V_c^i b_{t+1i}] / b_{t+1} = E[\omega_i V_c^i y_{Lti}] / y_L$ so that $\bar{b}^{\text{left}} = \bar{y}_L$.

Absent any behavioral response, bequest taxes are equivalent to labor taxes on distributional grounds because there is only one dimension of inequality left. Next, the bequest tax τ_B also reduces labor supply (as it reduces the use of income) exactly in the same proportion as the labor tax. Hence, shifting from the labor tax to the bequest tax has zero net effect on labor supply and $e_L = 0$. As parents are the zero-receivers in this model, we have $\bar{b}^{\text{received}} = 0$ when social welfare counts only parents' welfare. Therefore, optimal tax formula (10) with $\bar{b}^{\text{left}} = \bar{y}_L$ and $e_L = 0$ implies that $\tau_B < 0$. If children (i.e., bequest receivers) also enter social welfare, then $\bar{b}^{\text{received}} > 0$. In that case, formula (10) with $\bar{b}^{\text{left}} = \bar{y}_L$ and $e_L = 0$ implies that $\tau_B < 0$.

As our analysis makes clear, however, the Farhi-Werning (2010) two-period model only provides an incomplete characterization of the bequest tax problem because it fails to capture the fact that lifetime resources inequality is bi-dimensional, that is, individuals both earn and receive bequests.

¹² Farhi and Werning (2010) also obtained valuable results on the progressivity of the optimal bequest tax subsidy that cannot be captured in our linear framework.

This key bi-dimensional feature makes positive bequest taxes desirable under some redistributive social welfare criteria. An extension to our general model would be to consider nonlinear (but static) earnings taxation. The Atkinson-Stiglitz zero tax result would no longer apply as, conditional on labor earnings, bequests left are a signal for bequests received, and hence correlated with social marginal welfare weights, violating Assumption 1 of Saez's (2002) extension of Atkinson-Stiglitz to heterogeneous populations. The simplest way to see this is to consider the case with uniform labor earnings: Inequality arises solely from bequests, labor taxation is useless for redistribution, and bequest taxation is the only redistributive tool.

6.2.5 Accidental Bequests or Wealth Lovers

Individuals also leave bequests for non-altruistic reasons. For example, some individuals may value wealth per se (e.g., it brings social prestige and power), or for precautionary motives, and leave accidental bequests due to imperfect annuitization. Such non-altruistic reasons are quantitatively important (Kopczuk and Lupton (2007)). If individuals do not care about the after-tax bequests they leave, they are not hurt by bequest taxes on bequests they leave. Bequest receivers continue to be hurt by bequest taxes. This implies that the last term \bar{b}^{left} in the numerator of our formulas, capturing the negative effect of τ_B on bequest leavers, ought to be discounted. Formally, it is straightforward to generalize the model to utility functions $V^i(c, b, \underline{b}, l)$, where b is pre-tax bequest left, which captures wealth loving motives. The individual first order condition becomes $V_c^{ti} = R(1 - \tau_{Bt+1})V_{\underline{b}}^{ti} + V_b^{ti}$ and

$v_{ii} = R(1 - \tau_{Bt+1})V_{\underline{b}}^{ti} / V_c^{ti}$ naturally captures the relative importance of altruism in bequests motives. All our formulas carry over by simply replacing \bar{b}^{left} by $v \cdot \bar{b}^{\text{left}}$, with v the population average of v_{ii} (weighted by $g_{ii} b_{t+1i}$). Existing surveys can be used to measure the relative importance of altruistic motives versus other motives to calibrate the optimal τ_B . Hence, our approach is robust and flexible to accommodate such wealth loving effects that are empirically first order.

6.3 Optimal Inheritance Tax in the Dynastic Model

6.3.1 The Dynastic Model

The Barro-Becker dynastic model has been widely used in the analysis of optimal capital/inheritance taxation. Our sufficient statistics formula approach can also fruitfully be used in that case, with minor modifications. In the dynastic model, individuals care about the utility of

their heirs V^{t+li} instead of the after-tax capitalized bequests $R(1 - \tau_{Bt+1})b_{t+li}$ they leave. The standard assumption is the recursive additive form $V^{ti} = u^{ti}(c, l) + \delta V^{t+li}$, where $\delta < 1$ is a uniform discount factor. We assume again a linear and deterministic tax policy $(\tau_{Bt}, \tau_{Lt}, E_t)_{t \geq 0}$.

Individual ti chooses b_{t+li} and l_{ii} to maximize $u^{ti}(c_{ii}, l_{ii}) + \delta E_t V^{t+li}$ subject to the individual budget $c_{ii} + b_{t+li} = Rb_{ii}(1 - \tau_{Bt}) + w_{ii}l_{ii} + E_t$ with $b_{t+li} \geq 0$, where $E_t V^{t+li}$ denotes expected utility of individual $t + li$ (based on information known in period t). The first order condition for b_{t+li} implies the Euler equation $u_c^{ti} = \delta R(1 - \tau_{Bt+1})E_t u_c^{t+li}$ (whenever $b_{t+li} > 0$).

With stochastic ergodic processes for wages w_{ii} and preferences u^{ti} , standard regularity assumptions, this model also generates an ergodic equilibrium where long-run individual outcomes are independent of initial position. Assuming again that the tax policy converges to (τ_L, τ_B, E) , the long-run aggregate bequests and earnings b_t, y_{Lt} also converge and depend on asymptotic tax rates τ_L, τ_B . We show in Piketty and Saez (2013a) that this model generates finite long-run elasticities e_B, e_L defined as in (3) that satisfy (5) as in Section 6.2. The long-run elasticity e_B becomes infinite when stochastic shocks vanish. Importantly, as b_{t+li} is known at the end of period t , the individual first order condition in b_{t+li} implies that (regardless of whether $b_{t+li} = 0$):

$$u_c^{ti} \cdot b_{t+li} = \delta R(1 - \tau_{Bt+1})E_t [b_{t+li} u_c^{t+li}] \text{ and hence } \bar{b}_{t+1}^{\text{left}} = \delta R(1 - \tau_{Bt+1})\bar{b}_{t+1}^{\text{received}}, \quad (11)$$

with $\bar{b}_t^{\text{received}} = \frac{\int_i \omega_{0i} u_c^{ti} b_{ii}}{b_t \int_i \omega_{0i} u_c^{ti}}$ and $\bar{b}_{t+1}^{\text{left}} = \frac{\int_i \omega_{0i} u_c^{ti} b_{t+li}}{b_{t+1} \int_i \omega_{0i} u_c^{ti}}$ as in (4) for any dynastic Pareto weights $(\omega_{0i})_i$.

Paralleling the analysis of Section 6.2, we start with steady-state welfare maximization in Section 6.3.2 and then consider discounted utility maximization in Section 6.3.3.

6.3.2 Optimum Long-Run τ_B in Steady-State Welfare Maximization

We start with the utilitarian case (uniform Pareto weights $\omega_{0i} \equiv 1$). We assume that the economy is in steady-state ergodic equilibrium with constant tax policy τ_B, τ_L, E set such that the government budget constraint $\tau_B Rb_t + \tau_L y_{Lt} = E$ holds each period. As in Section 6.2.2, the

government chooses τ_B (with τ_L adjusting to meet the budget constraint and with E exogenously given) to maximize discounted steady-state utility:

$$\max_{\tau_B} EV_{\infty} = \sum_{t \geq 0} \delta^t E[u^t (Rb_{t+1}(1 - \tau_B) + w_{t+1}l_{t+1}(1 - \tau_L) + E - b_{t+1}l_{t+1})],$$

where we assume (w.l.o.g.) that the steady state has been reached in period 0. b_{0i} is given to the individual (but depends on τ_B), while b_{ti} for $t \geq 1$ and l_{ti} for $t \geq 0$ are chosen optimally so that the envelope theorem applies. Therefore, first order condition with respect to τ_B is

$$0 = E[u_c^{0i} \cdot R(1 - \tau_B)db_{0i}] - E[u_c^{0i} \cdot Rb_{0i}d\tau_B] \\ - \sum_{t \geq 0} \delta^{t+1} E[u_c^{t+1i} \cdot Rb_{t+1i}d\tau_B] - \sum_{t \geq 0} \delta^t E[u_c^t \cdot y_{Lti}d\tau_L],$$

where we have broken out into two terms the effect of $d\tau_B$. Using (5) linking $d\tau_L$ to, $d\tau_B$,

$$e_{Bi} = \frac{1 - \tau_B}{b_{0i}} \frac{db_{0i}}{d(1 - \tau_B)}, \text{ and the individual first order condition}$$

$$u_c^t b_{t+1i} = \delta R(1 - \tau_B) E_t u_c^{t+1i} b_{t+1i},$$

$$0 = -E[u_c^{0i} Rb_{0i} (e_{Bi} + 1)] + \sum_{t \geq 0} \delta^t \left(-\frac{E[u_c^t b_{t+1i}]}{1 - \tau_B} + E \left[u_c^t Rb_t \frac{1 - e_B \tau_B / (1 - \tau_B)}{1 - e_L \tau_L / (1 - \tau_L)} \frac{y_{Lti}}{y_{Lt}} \right] \right). \quad (12)$$

The sum in (12) is a repeat of identical terms because the economy is in ergodic steady state. Hence, the only difference with (6) in Section 6.2 is that the second and third terms are repeated (with discount factor δ), hence multiplied by $1 + \delta + \delta^2 + \dots = 1/(1 - \delta)$. Hence, this is equivalent to discounting the first term (bequest received effect) by a factor $1 - \delta$, so that we only need to replace $\bar{b}^{\text{received}}$ by $(1 - \delta)\bar{b}^{\text{received}}$ in formula (7). Hence, conditional on elasticities and distributional parameters, the dynastic case makes the optimal τ_B larger because double counting costs of taxation are reduced relative to the bequests in the utility model of Section 6.2.

Dynastic Model Long-Run Optimum, Steady-state Utilitarian Perspective:

$$\tau_B = \frac{1 - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \cdot \left[\frac{(1 - \delta) \bar{b}^{\text{received}}}{\bar{y}_L} (1 + \hat{e}_B) + \frac{1}{R} \frac{\bar{b}^{\text{left}}}{\bar{y}_L} \right]}{1 + e_B - \left[1 - \frac{e_L \tau_L}{1 - \tau_L} \right] \frac{(1 - \delta) \bar{b}^{\text{received}}}{\bar{y}_L} (1 + \hat{e}_B)}. \quad (13)$$

Hence, conditional on the sufficient statistics elasticities and distributional parameters, the dynastic model hardly changes the form of the optimal steady-state welfare maximizing τ_B relative to the bequests in the utility model of Section 6.2. Under the standard utilitarian social objective we have used, with enough curvature of utility functions, the distributional parameters $\frac{\bar{b}^{\text{received}}}{\bar{y}_L}$ and $\frac{\bar{b}^{\text{left}}}{\bar{y}_L}$ will be low if bequests are more concentrated than earnings. This realistic feature is difficult to obtain with only shocks to productivity (the standard model), but can be obtained with taste shocks. The dynastic utility model also generates large elasticities e_B when stochastic shocks are small. Indeed, the elasticity is infinite in the limit case with no stochastic shocks as in the Chamley-Judd model (see our discussion below). Therefore, the dynastic model leads to small optimal steady state τ_B only when it is (unrealistically) calibrated to generate either modest concentration of bequests (relative to earnings) or large elasticities of bequests with respect to $1 - \tau_B$. Our approach shows that, once these key sufficient statistics are known, the primitives of the model (dynastic vs. bequest loving) are largely irrelevant.

We can also consider general Pareto weights ω_{0t} . In (12), the sums over t are no longer identical terms, as the correlation of social marginal welfare weights $w_{0t} u_c^{it}$ with b_{t+1i} and y_{Lti}

changes with t . Hence, in that case, $\frac{1}{1 - \delta}$, \bar{b}^{left} , and \bar{y}_L have to be replaced by

$$\frac{1}{1 - \delta} = \sum_{t \geq 0} \delta^t \frac{E[\omega_{0t} u_c^{it}]}{E[\omega_{0t} u_c^{0i}]}, \quad \bar{b}^{\text{left}} = \frac{\sum_{t \geq 0} \delta^t E[\omega_{0t} u_c^{it} b_{t+1i}]}{\sum_{t \geq 0} \delta^t E[\omega_{0t} u_c^{it}]}, \quad \bar{y}_L = \frac{\sum_{t \geq 0} \delta^t E[\omega_{0t} u_c^{it} y_{Lti}]}{\sum_{t \geq 0} \delta^t E[\omega_{0t} u_c^{it}]}.$$

In the zero-receiver Meritocratic Rawlsian optimum, $\bar{b}^{\text{received}}$ vanishes, so that the simpler formula (8) applies in that case.

If stochastic shocks vanish, then $e_B = \infty$ (see Piketty and Saez (2013a) for a proof) and hence $\tau_B = 0$ even in the Meritocratic Rawlsian case with $\bar{b}^{\text{received}} = 0$ discussed above. This nests

the steady-state maximization version of Chamley and Judd (presented in Piketty (2000, p.444)) that delivers a zero τ_B optimum when the supply elasticity of capital is infinite even when the government cares only about workers with zero wealth.

Finally, it is possible to write a fully general model $V^i = u^i(c, \underline{b}, b, l) + \delta V^{i+1}$ that encompasses many possible bequest motivations. The optimal formula in the steady state continues to take the same general shape we have presented, although notations are more cumbersome.

6.3.3 Optimum Long-Run τ_B From Period Zero Perspective

Next, we consider maximization of period 0 dynastic utility, which has been the standard in the literature, and we solve for the long-run optimal τ_B . The key difference with Section 6.2.3 is that bequest behavior can change generations in advance of an anticipated tax change¹³.

To understand the key intuitions in the most pedagogical way, let us first assume inelastic earnings y_{Lti} . Because labor supply is inelastic, we assume without loss of generality that $\tau_L = 0$ and that bequest taxes fund the lump-sum grant so that $E_t = \tau_{Bt} R b_t$. Initial bequests $(b_{0i})_i$ are given. Let $(\tau_{Bt})_{t \geq 0}$ be the tax policy maximizing EV_0 , that is, expected utility of generation 0:

$$EV_0 = \sum_{t \geq 0} \delta^t E u^i (R b_{it} (1 - \tau_{Bt}) + \tau_{Bt} R b_t + y_{Lti} - b_{t+1i}).$$

Assume that τ_{Bt} converges to τ_B . Consider a small reform $d\tau_B$ for all $t \geq T$ where T is large so that all variables have converged to their limit. Using the envelope theorem for b_{it} , we have

$$dEV_0 = R d\tau_B \sum_{t \geq T} \delta^t E [u_c^i \cdot (b_t - b_{it})] + R \sum_{t \geq T} \delta^t E [u_c^i] \tau_{Bt} db_t.$$

The first term is the mechanical welfare effect (absent any behavioral response), while the second term reflects the welfare effect due to behavioral responses in bequest behavior affecting tax revenue (and hence the lump-sum grant). Importantly, note that the second sum starts at $t \geq 1$, as bequests may be affected before the reform takes place in anticipation. At the optimum,

$$0 = \frac{1}{R} \frac{dEV_0}{d\tau_B} = \sum_{t \geq T} \delta^t E [u_c^i \cdot (b_t - b_{it})] - \sum_{t \geq 1} \delta^t E [u_c^i] b_t \frac{\tau_{Bt}}{1 - \tau_{Bt}} e_{Bt}, \quad (14)$$

¹³ Recall that, in the bequest in the utility model of Section 6.2.3, a future bequest tax change at date T has no impact on behavior until the first generation of donors (i.e., generation $T-1$) is hit.

with $e_{b_t} = \frac{1-\tau_{B_t}}{b_t} \frac{db_t}{d(1-\tau_B)}$ the elasticity of b_t with respect to the small reform $d\tau_B$ (for all $t \geq T$).

For $t \geq T$, τ_{B_t} changes by $d\tau_B$ and the bequest decision is directly affected. When $t \rightarrow \infty$, e_{B_t} converges to the long-run elasticity e_B of b_t with respect to $1-\tau_B$ as in Section 6.3.1¹⁴. For $t < T$, τ_{B_t} does not change, hence bequest decisions are only affected in anticipation of the future tax increase. In a model with no stochastic shocks (as in Chamley-Judd), the full path of consumption is shifted up for $t < T$ and then decreases faster for $t \geq T$. This implies that bequests start responding from period 1 even for a very distant tax reform. In the stochastic model, however, the anticipation response is attenuated as individuals hit the zero wealth constraint almost certainly as the horizon grows (see Piketty and Saez (2013a)). Therefore, we can assume that e_{B_t} is nonzero only for t large at a point where τ_{B_t} , b_t , and c_{it} have converged to their long-run distribution. Hence, we can define the total elasticity e_B^{pdv} as the sum of the post-reform response elasticity e_B^{post} and the pre-reform anticipatory elasticity $e_B^{\text{anticip.}}$ as follows:

$$e_B^{\text{pdv}} = e_B^{\text{post}} + e_B^{\text{anticip.}} \quad \text{with}$$

$$e_B^{\text{post}} = (1-\delta) \sum_{t \geq T} \delta^{t-T} e_{B_t} \quad \text{and} \quad e_B^{\text{anticip.}} = (1-\delta) \sum_{t < T} \delta^{t-T} e_{B_t}. \quad (15)$$

e_B^{pdv} is the elasticity of the present discounted value of the tax base with respect to a distant tax rate increase. e_B^{post} is the standard (discounted) average of the post-reform elasticities e_{B_t} , while

$e_B^{\text{anticip.}}$ is the sum of all the pre-reform behavioral elasticities e_{B_t} . We show in Piketty and Saez (2013a) that $e_B^{\text{anticip.}}$ becomes infinite when stochastic shocks disappear as in Chamley-Judd. Importantly, in that case, $e_B^{\text{anticip.}}$ is infinite even in situations where the long-run elasticity e_B and hence e_B^{post} is finite, as in the endogenous discount factor case of Judd (1985, Theorem 5, p.79) (see Piketty and Saez (2013a)). However, this elasticity is finite in the Aiyagari (1995) model with stochastic shocks. Naturally, $e_B^{\text{pdv}} \rightarrow e_B$ when $\delta \rightarrow 1$. Numerical simulations could shed light on how $e_B^{\text{anticip.}}$, e_B^{post} , e_B change with the model specification and the structure of stochastic shocks.

As all terms in (14) have converged, dividing by $b_t E u_c^{it}$, and using (15), we rewrite (14) as

¹⁴ This long-run elasticity e_B is calculated assuming that tax revenue is rebated lump-sum period by period.

$$0 = \sum_{t \geq T} \delta^t \left[1 - \frac{E[u_c^i b_{it}]}{E[u_c^i] b_t} \right] - \frac{\tau_B}{1 - \tau_B} \sum_{t \geq 1} \delta^t e_{Bt}, \quad \text{hence}$$

$$0 = 1 - \frac{E[u_c^i b_{it}]}{b_t E[u_c^i]} - \frac{\tau_B}{1 - \tau_B} e_B^{\text{pdv}}.$$

Using the definition $\bar{b}^{\text{received}} = \frac{E[u_c^i b_{it}]}{b_t E[u_c^i]}$ and $\bar{b}^{\text{left}} = \delta R (1 - \tau_B) \bar{b}^{\text{received}}$ from (11), we therefore

obtain the following:

Dynastic Model Long-Run Optimum, Period 0 Perspective, Inelastic Labor Supply:

$$\tau_B = \frac{1 - \bar{b}^{\text{received}}}{1 - \bar{b}^{\text{received}} e_B^{\text{pdv}}} \quad \text{or equivalently} \quad \tau_B = \frac{1 - \frac{1}{\delta R} \bar{b}^{\text{left}}}{1 + e_B^{\text{pdv}}}, \quad (16)$$

where e_B^{pdv} , defined in (15), is the total (post-reform and anticipatory) elasticity of the present discounted value of aggregate bequests to a long-term distant pre-announced bequest tax increase.

Six points are worth noting about formula (16). First, it shows that the standard equity-efficiency approach also applies to the standard dynastic model. The first expression in (16) takes the standard optimal linear tax rate form, decreasing in the elasticity e_B^{pdv} and decreasing with the distributional parameter $\bar{b}^{\text{received}}$. The key is to suitably define the elasticity e_B^{pdv} . As argued above, this elasticity is infinite in the Chamley-Judd model with uncertainty, so that our analysis nests the Chamley-Judd zero tax result. However, whenever the elasticity e_B^{pdv} is finite, the optimal tax rate is positive as long as $\bar{b}^{\text{received}} < 1$, that is bequests received are negatively correlated with marginal utility u_c^i , which is the expected case. This point on the sign of optimal long-run bequest taxation was made by Chamley (2001), although he did not derive an optimal tax formula. He also crafted an example showing that $\bar{b}^{\text{received}} > 1$ is theoretically possible.

Second, there is no double counting in the dynastic model from period 0 perspective. Hence, the cost of bequest taxation can be measured either on bequest receivers (first formula in (16)) or, equivalently, on bequest leavers (second formula in (16)). This shows that the optimal τ_B in the dynastic model takes the same form as (9), the long-run optimum with social discounting from

Section 6.2, ignoring the welfare effect on bequest receivers, that is, setting $\bar{b}^{\text{received}} = 0$.¹⁵

Third, we can add labor supply decisions. Considering a $d\tau_B$, $d\tau_L$ trade-off modifies the optimal tax rate as expected. $\bar{b}^{\text{received}}$ and \bar{b}^{left} in (16) need to be replaced by $\frac{\bar{b}^{\text{received}}}{\bar{y}_L} \left[1 - \frac{e_L^{\text{pvd}} \tau_L}{1 - \tau_L} \right]$ and $\frac{\bar{b}^{\text{left}}}{\bar{y}_L} \left[1 - \frac{e_L^{\text{pvd}} \tau_L}{1 - \tau_L} \right]$, with e_L^{pvd} the elasticity of aggregate PDV earnings (see Section S.1.3).

Fourth, optimal government debt management in the closed economy would deliver the Modified Golden Rule $\delta R = 1$ and the same formulas continue to hold (see Section S.1.4).

Fifth, we can consider heterogeneous discount rate δ_{ii} . Formula (16) still applies with $\bar{b}^{\text{received}} = \lim_T \frac{\sum_{t \geq T} E[\delta_{i1} \cdots \delta_{it} u_c^t b_{it}]}{\sum_{t \geq T} E[\delta_{i1} \cdots \delta_{it} u_c^t] b_t}$. Hence, $\bar{b}^{\text{received}}$ puts weight on consistently altruistic dynasties, precisely those that accumulate wealth so that $\bar{b}^{\text{received}} > 1$ and $\tau_B < 0$ is likely. In that case, the period 0 criterion puts no weight on individuals who had non-altruistic ancestors. This fits with aristocratic values, but is the polar opposite of realistic modern meritocratic values. Hence, the dynastic model with the period zero objective generates unappealing normative recommendations when there is heterogeneity in tastes for bequests.

Sixth, adding Pareto weights ω_{0i} that depend on initial position delivers exactly the same formula, as the long-run position of each individual is independent of the initial situation. This severely limits the scope of social welfare criteria in the period 0 perspective model relative to the steady-state welfare maximization model analyzed in Section 6.3.2.

¹⁵ Naturally, $\tau_L = e_L = 0$ here. Note also that \bar{y}_L is replaced by 1 because the trade-off here is between the bequest tax and the lump-sum grant (instead of the labor tax as in Section 6.2).

References

- Abel, A., G. Mankiw, L. Summers, and R. Zeckhauser (1989) "Assessing dynamic efficiency: Theory and evidence." *Review of Economic Studies*, 56: 1-20.
- Atkinson, A., and A. Sandmo (1980) "Welfare implications of the taxation of savings." *Economic Journal*, 90: 529-49.
- Auerbach, A., and L. Kotlikoff (1987) *Dynamic Fiscal Policy*, Cambridge University Press.
- Blumkin T. and E. Sadka (2003) "Estate taxation with intended and accidental bequests." *Journal of Public Economics*, 88: 1-21.
- Chamley, C. (1986) "Optimal taxation of capital income in general equilibrium with infinite lives." *Econometrica*, 54: 607-22.
- Chamley, C. (2001) "Capital income taxation, wealth distribution and borrowing constraints." *Journal of Public Economics*, 79: 55-69.
- Cremer H. and P. Pestieau (2001) "Non-linear taxation of bequests, equal sharing rules and the tradeoff between intra- and inter-family inequalities." *Journal of Public Economics*, 79: 35-53.
- Diamond, P. (1965) "National debt in a neoclassical growth model." *American Economic Review*, 55:1126-50.
- Gordon, R. (1986) "Taxation of investment and savings in a world economy." *American Economic Review*, 76: 1086-1102.
- Hilman, A.L. (2003) *Public Finance and Public Policy*, Cambridge; Cambridge University Press.
- Judd, K. (1985) "Redistributive taxation in a simple perfect foresight model." *Journal of Public Economics*, 28: 59-83.
- Krusell, P., L. Ohanian, J.-V. Rios-Rull, and G. Violante (2000) "Capital-skill complementarity and inequality: A macroeconomic analysis." *Econometrica*, 68: 1029-53.
- Kopczuk, W. (2013) "Taxation of Transfers and Wealth" in *Handbook of Public Economics*, vol.5, Amsterdam: North-Holland. [1852]
- Ordoover, J., and E. Phelps (1979) "The concept of optimal taxation in the overlapping generations model of capital and wealth." *Journal of Public Economics*, 12: 1-26.
- Piketty, T., and E. Saez (2012) "A Theory of Optimal Capital Taxation" Working Paper 17989, NBER. [1854, 1862, 1879 1880]
- Piketty, T., and E. Saez (2013a) "A Theory of Optimal Inheritance Taxation" *Econometrica*, 81(5): 1851-86.
- Piketty, T., and E. Saez (2013b) "Supplement to 'A Theory of Optimal Capital Taxation'," *Econometrica* Supplemental Material, 81, http://www.econometricsociety.org/ecta/supmat/10712_proofs.pdf; http://www.econometricsociety.org/ecta/supmat/10712_programs_and_data.zip. [1861]
- Salanié, B. (2003) *The Economics of Taxation*, Cambridge, MA; The MIT Press
- Samuelson, P. (1958) "An exact consumption-loan model of interest with or without the social contrivance of money." *Journal of Political Economy*, 66: 467-82.
- Stiglitz, J. (1985) "Inequality and capital taxation." IMSSS Technical Report 457, Stanford University.
- Summers, L. (1981) "Taxation and capital accumulation in a life cycle growth model." *American Economic Review*, 71: 533-54.
- Tobin, J. (1958) "Liquidity preference as behavior towards risk." *Review of Economic Studies*, 25: 65-8.
- Tuomala M. (1985) "Uncertain lifetime, life insurance, and the theory of consumer." *Review of Economic Studies*, 32: 137-50.