WILEY

A THEORY OF OPTIMAL INHERITANCE TAXATION
Author(s): Thomas Piketty and Emmanuel Saez

A THEORY OF OPTIMAL INHERITANCE TAXATION

By Thomas Piketty and Emmanuel Saez[1]

This paper derives optimal inheritance tax formulas that capture the key equity-efficiency trade-off, are expressed in terms of estimable sufficient statistics, and are robust to the underlying structure of preferences. We consider dynamic stochastic models with general and heterogeneous bequest tastes and labor productivities. We limit ourselves to simple but realistic linear or two-bracket tax structures to obtain tractable formulas. We show that long-run optimal inheritance tax rates can always be expressed in terms of aggregate earnings and bequest elasticities with respect to tax rates, distributional parameters, and social preferences for redistribution. Those results carry over with tractable modifications to (a) the case with social discounting (instead of steady-state welfare maximization), (b) the case with partly accidental bequests, (c) the standard Barro–Becker dynastic model. The optimal tax rate is positive and quantitatively large if the elasticity of bequests to the tax rate is low, bequest concentration is high, and society cares mostly about those receiving little inheritance. We propose a calibration using micro-data for France and the United States. We find that, for realistic parameters, the optimal inheritance tax rate might be as large as 50%–60%—or even higher for top bequests, in line with historical experience.

Keywords: Optimal taxation, inheritance, wealth mobility.

1. INTRODUCTION

THERE IS SUBSTANTIAL CONTROVERSY both in the public policy debate and among economists about the proper level of taxation of inherited wealth. The public debate centers around the equity versus efficiency trade-off. In the economic debate, there is a disparate set of models and results on optimal inheritance taxation. Those models differ primarily in terms of preferences for savings/bequests and the structure of economic shocks. In the dynastic interpretation of the infinite horizon model of Chamley (1986) and Judd (1985) with no stochastic shocks, the optimal inheritance tax is zero in the long run, because a constant inheritance tax rate creates a growing distortion on intertemporal choices. However, many subsequent studies have shown that this famous zero tax result can be overturned by relaxing each of the key hypotheses.[2] In a two-generation model with parents starting with no wealth but having heteroge-

---

[2]The most studied extensions leading to nonzero inheritance taxes are: (a) presence of idiosyncratic labor income shocks, (b) accidental bequests, (c) bequests givers caring about pre-tax or post-tax bequests rather than the utility of heirs, (d) long-run steady-state welfare maximization,

neous working abilities and leaving bequests to children (with no earnings), bequest taxes are useless with an optimal earnings tax on parents if social welfare is measured solely from the parents' perspective (Atkinson and Stiglitz (1976)). If children's utilities also enter directly social welfare, then a negative bequest tax is desirable (Kaplow (2001), Farhi and Werning (2010)). Hence, the theory of optimal inheritance taxation is scattered with no clear policy implications, as different—yet difficult to test—assumptions for bequest behavior lead to different formulas and magnitudes.

In this paper, we make progress on this issue by showing that optimal inheritance tax formulas can be expressed in terms of estimable "sufficient statistics" including behavioral elasticities, distributional parameters, and social preferences for redistribution. Those formulas are robust to the underlying primitives of the model and capture the key equity-efficiency trade-off in a transparent way. This approach has been fruitfully used in the analysis of optimal labor income taxation. (Piketty and Saez (2013a) provided a recent survey.) We follow a similar route and show that the equity-efficiency trade-off logic also applies to inheritance taxation. This approach successfully brings together many of the existing scattered results from the literature.

We first consider dynamic stochastic models with general and heterogeneous preferences for bequests and ability for work, where donors care solely about the net-of-tax bequest they leave to their heirs, and where the planner maximizes long-run steady-state welfare (Section 2.2). This is the simplest case to illustrate the key equity-efficiency trade-off transparently. Importantly, our results carry over with tractable modifications to (a) the case with social discounting instead of steady-state welfare maximization (Section 2.3), (b) the case with partly accidental bequests (Section 2.5), (c) the standard Barro–Becker dynastic model with altruism (Section 3).

In all cases, the problem can be seen as an equity-efficiency trade-off, where the optimal inheritance tax rate decreases with the elasticity of aggregate bequests to the net-of-tax bequest tax rate (defined as 1 minus the tax rate), and decreases with the value that society puts on the marginal consumption of bequest receivers and bequest leavers. The optimal tax rate is positive and quantitatively large if the elasticity is low, bequests are quantitatively large and highly concentrated, and society cares mostly about those receiving little inheritance. In contrast, the optimal tax rate can be negative when society cares mostly about inheritors.

As in the public debate, the desirability of taxing bequests hinges primarily on wealth inequality and mobility and how social marginal welfare weights are distributed across groups. The optimal tax rate is zero when the elasticity of bequests is infinite nesting the zero tax Chamley–Judd result. In contrast to

(e) time-invariant taxes, (f) lack of government commitment. (Cremer and Pestiau (2004) and Kopczuk (2013) provided recent surveys.)

Farhi and Werning (2010), inheritance taxation is positive even with optimal labor taxation because, in our model with bequests, inequality is bi-dimensional and earnings are no longer the unique determinant of lifetime resources. As a result, the famous Atkinson and Stiglitz (1976) zero tax result breaks down.[3]

Importantly, we limit ourselves to extremely simple linear (or two-bracket) tax structures on inheritances and labor income to be able to obtain tractable formulas in models with very heterogeneous preferences. The advantages are that, by necessity, our tax system is well within the realm of current practice and the economic trade-offs appear transparently. This "simple tax structure" approach is in contrast to the recent new dynamic public finance (NDPF) literature (Kocherlakota (2010) provided a recent survey) which considers the fully optimal mechanism given the informational structure. The resulting tax systems are complex—even with strong homogeneity assumptions for individual preferences—but potentially more powerful to increase welfare. Therefore, we view our approach as complementary to the NDPF approach.

As an illustration of the use of our formulas in sufficient statistics for policy recommendations, we propose a numerical simulation calibrated using microdata for the case of France and the United States (Section 4). For realistic parameters, the optimal inheritance tax rate might be as large as 50%–60%— or even higher for top bequests, in line with historical experience.

2. OPTIMAL INHERITANCE TAX WITH BEQUESTS IN THE UTILITY

2.1. Model

We consider a dynamic economy with a discrete set of generations $0, 1, \ldots,$ $t, \ldots$ 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 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

[3]Formally, our model can nest the Farhi–Werning two-period model (Section 2.5). In that case, inequality is uni-dimensional and we obtain the (linear tax version of) Farhi–Werning's results. The optimal inheritance tax rate is zero when maximizing parents' welfare and negative if the social planner also puts weight on children.

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

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

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

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

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^{ti} and for wage rates w_{ti} 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^{ti} and w_{ti}. 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 $(b_{ti}, y_{Lti})_i$. In the long run, the position of each dynasty i is independent of the initial position (b_{0i}, y_{L0i}).

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 $\omega_{ti} \geq 0$,

---

[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.

[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.

subject to a period-by-period budget balance $E = \tau_B Rb_t + \tau_L y_{Lt}$:

$$(2) \qquad SWF = \max_{\tau_L, \tau_B} \int_i \omega_{ti} V^{ti}\big(Rb_{ti}(1 - \tau_B) + w_{ti}l_{ti}(1 - \tau_L) + E - b_{t+1i},$$

$$Rb_{t+1i}(1 - \tau_B), l_{ti}\big).$$

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

$$(3) \qquad \textit{Long-run Elasticities}: \quad e_B = \frac{1 - \tau_B}{b_t} \frac{db_t}{d(1 - \tau_B)}\bigg|_E \quad \text{and}$$

$$e_L = \frac{1 - \tau_L}{y_{Lt}} \frac{dy_{Lt}}{d(1 - \tau_L)}\bigg|_E.$$

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 $g_{ti} = \omega_{ti} V_c^{ti} / \int_j \omega_{tj} V_c^{tj}$ 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

(4)   *Distributional Parameters*: $\bar{b}^{received} = \dfrac{\displaystyle\int_i g_{ti} b_{ti}}{b_t}$,

$$\bar{b}^{left} = \frac{\displaystyle\int_i g_{ti} b_{t+1i}}{b_{t+1}}, \quad \text{and} \quad \bar{y}_L = \frac{\displaystyle\int_i g_{ti} y_{Lti}}{y_{Lt}}.$$

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 $Rb_t\, d\tau_B + \tau_B R\, db_t + y_{Lt}\, d\tau_{Lt} + \tau_{Lt}\, dy_{Lt} = 0$. Using the elasticity definitions (3), this implies

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

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

$$dSWF = \int_i \omega_{ti} V_c^{ti} \cdot \left(R\, db_{ti}(1 - \tau_B) - Rb_{ti}\, d\tau_B - d\tau_L y_{Lti}\right)$$
$$+ \omega_{ti} V_{\underline{b}}^{ti} \cdot (-d\tau_B R b_{t+1i}).$$

At the optimum τ_B, $dSWF = 0$. Using the individual first order condition $V_c^{ti} = R(1 - \tau_B)V_{\underline{b}}^{ti}$ when $b_{t+1i} > 0$, expression (5) for $d\tau_L$, and the definition of $g_{ti} = \omega_{ti} V_c^{ti} / \int_j \omega_{tj} V_c^{tj}$, we have

(6) $0 = \displaystyle\int_i g_{ti} \cdot \Bigg(-d\tau_B R b_{ti}(1 + e_{Bti})$

$$+ \frac{1 - e_B \tau_B/(1 - \tau_B)}{1 - e_L \tau_L/(1 - \tau_L)} \frac{y_{Lti}}{y_{Lt}} Rb_t\, d\tau_B - d\tau_B \frac{b_{t+1i}}{1 - \tau_B}\Bigg),$$

where we have expressed db_{ti} using $e_{Bti} = \frac{1-\tau_B}{b_{ti}} \frac{db_{ti}}{d(1-\tau_B)}|_E$ the individual elasticity of bequest received (e_B is the bequest-weighted population average of e_{Bti}).

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_{Bti} weighted by $g_{ti}b_{ti}$.[6] Dividing (6) by $Rb_t \, d\tau_B$, and 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*

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

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 $\bar{b}^{\text{received}}$, $\bar{b}^{\text{left}}$, 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 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.

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

2. *Endogeneity of right-hand-side parameters.* As with virtually all optimal tax formulas, e_B, e_L, $\bar{b}^{\text{left}}$, $\bar{b}^{\text{received}}$, and $\bar{y}_L$ depend on tax rates τ_B, τ_L and hence are endogenous.[7] For calibration, assumptions need to be made on how those parameters vary with tax rates. To show the usefulness of those sufficient statistics formulas, we propose such an exercise in Section 4 using the actual joint micro-distributions of ($\bar{b}^{\text{received}}$, $\bar{b}^{\text{left}}$, y_L) for the United States and France. 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).

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 $\bar{b}^{\text{left}}$, 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_{ti} 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_{ti}'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 ($\bar{b}^{\text{left}}$ 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 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_{ti} 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 (Section 4). Hence, this "Meritocratic Rawlsian" optimum has broader appeal than the standard Rawlsian case.

---

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

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*

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

with $\bar{b}^{\text{left}}$, $\bar{y}_L$ 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 bequests 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.

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 $\Delta \le 1$ (Section 2.2 was the special case $\Delta = 1$). We derive the long-run optimum τ_B, that is, when all variables have converged:

$$SWF = \sum_{t \ge 0} \Delta^t \int_i \omega_{ti} V^{ti} \big(R b_{ti}(1 - \tau_{Bt}) + w_{ti}l_{ti}(1 - \tau_{Lt}) + E_t - b_{t+1i},$$

$$R b_{t+1i}(1 - \tau_{Bt+1}), l_{ti}\big).$$

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),

$$dSWF = \sum_{t \geq T} \Delta^t \int_i \omega_{ti} V_c^{ti} \cdot \left(R \, db_{ti}(1 - \tau_B) - Rb_{ti} \, d\tau_B - d\tau_{Lt} y_{Lti} \right)$$

$$+ \sum_{t \geq T-1} \Delta^t \int_i \omega_{ti} V_{\underline{b}}^{ti} \cdot (-d\tau_B Rb_{t+1i}).$$

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 (see Appendix A.1, equations (A.2) and (A.3), for exact and complete definitions). 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. In Appendix A.1, 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*

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

with e_B, $\hat{e}_B$, and e_L the discounted aggregate bequest and earnings elasticities defined in Appendix A.1, equations (A.2) and (A.3), 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 for an instant to the budget balance case, it is straightforward 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 Section S.1.1 of 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 $R\Delta = 1$,*

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

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 of period T) and da on long-term government debt (measured as of period T). Both $dSWF$ and da are proportional to dw.

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 $da' = -R\,da$ 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 (Section S.1.1 of the Supplemental Material). 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 issues.[8]

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 receivers.[9] If we assume instead that period t taxes are $\tau_{Bt}b_{t+1} + \tau_{Lt}y_{Lt}$, then

---

[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 to-

formula (9) would have no $R\Delta$ term dividing $\bar{b}^{\text{left}}$, 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 (see Section S.2 of the Supplemental Material).

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.

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 $V^{ti}(c, \underline{b}, l) = \frac{(U^{ti}(c,\underline{b})e^{-h_{ti}(l)})^{1-\gamma}}{1-\gamma}$, with $U^{ti}(c, \underline{b})$ 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 generations.[10] We show in Section S.1.2 of the Supplemental Material 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+1i}/b_{t+1} requires making a bequest G times larger than leaving the same relative bequest b_{t+1i}/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

---

day, 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).

[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.

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 $\Delta RG^{-\gamma} = 1$ 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 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 $\bar{b}^{\text{left}}$ and hence a higher τ_B.

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 model 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 model.[11] 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 to this specific model (Appendix A.2).

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 negative.[12] 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 $V^{ti}(c, \underline{b}, l) = U^{ti}(u(c, \underline{b}), l)$ 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 Appendix A.2, 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 $b_{t+1i} = s \cdot y_{Lti}(1 - \tau_{Lt})$, where s is homogeneous in the population. This immediately implies that $E[\omega_{ti} V_c^{ti} b_{t+1i}]/b_{t+1} = E[\omega_{ti} V_c^{ti} y_{Lti}]/y_{Lt}$ so that $\bar{b}^{\text{left}} = \bar{y}_L$.

---

[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 $V^{ti}(c_{ti}, Rb_{t+1i}(1 - \tau_{Bt+1}) + E_{t+1}^{\text{child}}, l_{ti})$. Farhi and Werning (2010) considered this altruistic case.

[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.

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. 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.

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 $\bar{b}^{\text{left}}$ 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 $V^{ti}(c, b, \underline{b}, l)$, 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 $\nu_{ti} = R(1 - \tau_{Bt+1})V_{\underline{b}}^{ti}/V_c^{ti}$ naturally captures the relative importance of altruism in bequests motives. All our formulas carry over by simply replacing $\bar{b}^{\text{left}}$ by $\nu \cdot \bar{b}^{\text{left}}$, with ν the population average of ν_{ti} (weighted by $g_{ti}b_{t+1i}$). As we shall see in Section 4, 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.

3. OPTIMAL INHERITANCE TAX IN THE DYNASTIC MODEL

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+1i} instead of the after-tax capitalized bequests $R(1 - \tau_{Bt+1})b_{t+1i}$ they leave. The standard assumption is the recursive additive form $V^{ti} = u^{ti}(c, l) + \delta V^{t+1i}$, 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+1i} and l_{ti} to maximize $u^{ti}(c_{ti}, l_{ti}) + \delta E_t V^{t+1i}$ subject to the individual budget $c_{ti} + b_{t+1i} = Rb_{ti}(1 - \tau_{Bt}) + w_{ti}l_{ti} + E_t$ with $b_{t+1i} \geq 0$, where $E_t V^{t+1i}$ denotes expected utility of individual $t + 1i$ (based on information known in period t). The first order condition for b_{t+1i} implies the Euler equation $u_c^{ti} = \delta R(1 - \tau_{Bt+1})E_t u_c^{t+1i}$ (whenever $b_{t+1i} > 0$).

With stochastic ergodic processes for wages w_{ti} 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 Appendix A.3 that this model generates finite long-run elasticities e_B, e_L defined as in (3) that satisfy (5) as in Section 2. The long-run elasticity e_B becomes infinite when stochastic shocks vanish. Importantly, as b_{t+1i} is known at the end of period t, the individual first order condition in b_{t+1i} implies that (regardless of whether $b_{t+1i} = 0$):

$$(11) \qquad u_c^{ti} \cdot b_{t+1i} = \delta R(1 - \tau_{Bt+1})E_t\left[b_{t+1i}u_c^{t+1i}\right] \quad \text{and hence}$$

$$\bar{b}_{t+1}^{\text{left}} = \delta R(1 - \tau_{Bt+1})\bar{b}_{t+1}^{\text{received}},$$

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

Paralleling the analysis of Section 2, we start with steady-state welfare maximization in Section 3.2 and then consider discounted utility maximization in Section 3.3.

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 $\tau_B R b_t + \tau_L y_{Lt} = E$ holds each period. As in Section 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:

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

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

$$0 = E \big[u_c^{0i} \cdot R(1 - \tau_B) \, db_{0i} \big] - E \big[u_c^{0i} \cdot R b_{0i} \, d\tau_B \big]$$

$$- \sum_{t \geq 0} \delta^{t+1} E \big[u_c^{t+1i} \cdot R b_{t+1i} \, d\tau_B \big] - \sum_{t \geq 0} \delta^t E \big[u_c^{ti} \cdot y_{Lti} \, d\tau_L \big],$$

where we have broken out into two terms the effect of $d\tau_B$. Using (5) linking $d\tau_L$ to $d\tau_B$, $e_{Bi} = \frac{1 - \tau_B}{b_{0i}} \frac{db_{0i}}{d(1 - \tau_B)}$, and the individual first order condition $u_c^{ti} b_{t+1i} = \delta R(1 - \tau_B) E_t u_c^{t+1i} b_{t+1i}$,

(12) $0 = -E \big[u_c^{0i} R b_{0i}(e_{Bi} + 1) \big]$

$$+ \sum_{t \geq 0} \delta^t \left(-\frac{E[u_c^{ti} b_{t+1i}]}{1 - \tau_B} + E \left[u_c^{ti} R b_t \frac{1 - e_B \tau_B/(1 - \tau_B)}{1 - e_L \tau_L/(1 - \tau_L)} \frac{y_{Lti}}{y_{Lt}} \right] \right).$$

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 2 is that the second and third terms are repeated (with discount factor δ), hence multiplied by $1 + \delta + \delta^2 + \cdots = 1/(1 - \delta)$. 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 2.

DYNASTIC MODEL LONG-RUN OPTIMUM, STEADY-STATE UTILITARIAN PERSPECTIVE:

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

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 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_{0i}. In (12), the sums over t are no longer identical terms, as the correlation of social marginal welfare weights $\omega_{i0}u_c^{ti}$ with b_{t+1i} and y_{Lti} changes with t. Hence, in that case, $\frac{1}{1-\delta}$, $\bar{b}^{\text{left}}$, and $\bar{y}_L$ have to be replaced by

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

$$\bar{y}_L = \frac{\sum_{t \geq 0} \delta^t E[\omega_{0i}u_c^{ti}y_{Lti}]}{\sum_{t \geq 0} \delta^t E[\omega_{0i}u_c^{ti}]y_{Lt}}.$$

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 Appendix A.3 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 $V^{ti} = u^{ti}(c, \underline{b}, b, l) + \delta V^{t+1i}$ 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.

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 2.3 is that bequest behavior can change generations in advance of an anticipated tax change.[13]

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:

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

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_{ti}, we have

$$dEV_0 = R \, d\tau_B \sum_{t \geq T} \delta^t E\big[u_c^{ti} \cdot (b_t - b_{ti}) \big] + R \sum_{t \geq 1} \delta^t E\big[u_c^{ti} \big] \tau_{Bt} \, db_t.$$

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,

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

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

For $t \geq T$, τ_{Bt} changes by $d\tau_B$ and the bequest decision is directly affected. When $t \to \infty$, e_{Bt} converges to the long-run elasticity e_B of b_t with respect to $1 - \tau_B$ as in Section 3.1.[14] For $t < T$, τ_{Bt} 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

---

[13]Recall that, in the bequest in the utility model of Section 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.

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

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 Appendix A.3). Therefore, we can assume that e_{Bt} is nonzero only for t large at a point where τ_{Bt}, b_t, and c_{ti} 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 $e_B^{\mathrm{anticip.}}$ as follows:

$$(15) \qquad e_B^{\mathrm{pdv}} = e_B^{\mathrm{post}} + e_B^{\mathrm{anticip.}} \quad \text{with}$$

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

e_B^{pvd} 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_{Bt}, while $e_B^{\mathrm{anticip.}}$ is the sum of all the pre-reform behavioral elasticities e_{Bt}. We show in Appendix A.3 that $e_B^{\mathrm{anticip.}}$ becomes infinite when stochastic shocks disappear as in Chamley–Judd. Importantly, in that case, $e_B^{\mathrm{anticip.}}$ 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 Appendix A.3). However, this elasticity is finite in the Aiyagari (1995) model with stochastic shocks. Naturally, $e_B^{\mathrm{pvd}} \to e_B$ when $\delta \to 1$. Numerical simulations could shed light on how $e_B^{\mathrm{anticip.}}$, 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 $b_t E u_c^{ti}$, and using (15), we rewrite (14) as

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

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

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

DYNASTIC MODEL LONG-RUN OPTIMUM, PERIOD 0 PERSPECTIVE, INELASTIC LABOR SUPPLY:

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

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^{pvd} and decreasing with the distributional parameter $\bar{b}^{received}$. The key is to suitably define the elasticity e_B^{pvd}. As argued above, this elasticity is infinite in the Chamley–Judd model with no uncertainty, so that our analysis nests the Chamley–Judd zero tax result. However, whenever the elasticity e_B^{pvd} is finite, the optimal tax rate is positive as long as $\bar{b}^{received} < 1$, that is, bequests received are negatively correlated with marginal utility u_c^{ti}, 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 $\bar{b}^{received} > 1$ 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 2, ignoring the welfare effect on bequest receivers, that is, setting $\bar{b}^{received} = 0$.[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. $\bar{b}^{received}$ and $\bar{b}^{left}$ in (16) need to be replaced by $\frac{\bar{b}^{received}}{\bar{y}_L}[1 - \frac{e_L^{pvd}\tau_L}{1-\tau_L}]$ and $\frac{\bar{b}^{left}}{\bar{y}_L}[1 - \frac{e_L^{pvd}\tau_L}{1-\tau_L}]$, 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 rates δ_{ti}. Formula (16) still applies with $\bar{b}^{received} = \lim_T \frac{\sum_{t\geq T} E[\delta_{1i}\cdots\delta_{ti}u_c^{ti}b_{ti}]}{\sum_{t\geq T} E[\delta_{1i}\cdots\delta_{ti}u_c^{ti}]b_t}$. Hence, $\bar{b}^{received}$ puts weight on consistently altruistic dynasties, precisely those that accumulate wealth so that $\bar{b}^{received} > 1$ 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 unappeal-

---

[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 2).

ing 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 3.2.

4. NUMERICAL CALIBRATIONS

We use wealth surveys for France (Enquête Patrimoine 2010) and the United States (Survey of Consumer Finances 2010) to calibrate the general steady-state formula (see Section S.3 for details)

$$(17) \qquad \tau_B = \frac{1 - \left[1 - \dfrac{e_L \tau_L}{1 - \tau_L}\right] \cdot \left[\dfrac{\bar{b}^{\text{received}}}{\bar{y}_L}(1 + \hat{e}_B) + \dfrac{\nu}{R/G}\dfrac{\bar{b}^{\text{left}}}{\bar{y}_L}\right]}{1 + e_B - \left[1 - \dfrac{e_L \tau_L}{1 - \tau_L}\right]\dfrac{\bar{b}^{\text{received}}}{\bar{y}_L}(1 + \hat{e}_B)},$$

which incorporates growth G (Section 2.3) and wealth loving motivations ν (Section 2.5). We consider the following benchmark values for the parameters: $e_B = \hat{e}_B = 0.2$, $e_L = 0.2$, $\tau_L = 30\%$, $R/G = e^{(r-g)H} = 1.82$ with $r - g = 2\%$ and $H = 30$ years, and $\nu = 1$ (pure bequests motives).[16] We discuss these parameter choices and carry out sensitivity analysis in Table I.

We use the joint micro-level distribution of bequests received, bequests left, and lifetime labor earnings $(b_{ti}, b_{t+1i}, y_{Lti})$ from the survey data to compute the distributional parameters $\bar{b}^{\text{received}}$, $\bar{b}^{\text{left}}$, and $\bar{y}_L$ using definition (4). This requires specifying social welfare weights g_{ti}. To be agnostic and explore heterogeneity in optimal τ_B across the distribution, we consider percentile p-weights which concentrate uniformly the weights g_{ti} on percentile p of the distribution of bequests received. Hence, for p-weights, $\bar{b}^{\text{received}}$, $\bar{b}^{\text{left}}$, and $\bar{y}_L$ are the average of bequests received, bequests left, earnings (relative to population averages) among pth percentile bequest receivers. By definition, $\bar{b}^{\text{received}}$ increases with p. As we shall see, $\bar{b}^{\text{left}}$ also increases with p, as large receivers tend to leave higher than average bequests themselves, while $\bar{y}_L$ only mildly increases with p.

Those distributional parameters are computed within the population of individuals aged 70 and above.[17] We use retrospective questions about bequest

---

[16]We use $\nu = 1$ in Figures 1 and 2. A more realistic (and still conservative) value based on estimates from Kopczuk and Lupton (2007) is probably $\nu = 0.7$ (see our discussion in Table I).

[17]We focus on older cohorts because they have already received bequests from their parents, and will soon leave bequests to their children. Hence, we can estimate the distribution of bequests both received and left. Wealth at age 70 and above overestimates bequests left because of late life

TABLE I

OPTIMAL INHERITANCE TAX RATE τ_B CALIBRATIONS[a]

	Elasticity $e_B = 0$ (Low-End Estimate)		Elasticity $e_B = 0.2$ (Middle-End Estimate)		Elasticity $e_B = 0.5$ (High-End Estimate)		Elasticity $e_B = 1$ (Extreme Estimate)	
	France	U.S.	France	U.S.	France	U.S.	France	U.S.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0. Basic Specification: Optimal Tax for Zero Receivers (Bottom 50%), $r - g = 2\%$ ($R/G = 1.82$), $\nu = 70\%$, $e_L = 0.2$, No Exemption (Linear Tax τ_B)								
P0–50, $r - g = 2\%$, $\nu = 70\%$, $e_L = 0.2$	76%	70%	63%	59%	50%	47%	38%	35%
1. Optimal Linear Tax Rate for Other Groups by Percentile of Bequests Received								
P50–70	75%	70%	62%	59%	48%	47%	35%	35%
P70–90	45%	60%	31%	46%	16%	31%	2%	17%
P90–95	-283%	-43%	-330%	-84%	-376%	-126%	-423%	-167%
2. Sensitivity to Capitalization Factor $R/G = e^{(r-g)H}$								
$r - g = 0\%$ ($R/G = 1$) or dynamic efficiency	56%	46%	46%	38%	37%	31%	28%	23%
$r - g = 3\%$ ($R/G = 2.46$)	82%	78%	68%	65%	55%	52%	41%	39%
3. Sensitivity to Bequests Motives ν								
$\nu = 1$ (100% bequest motives)	65%	58%	54%	48%	43%	39%	33%	29%
$\nu = 0$ (no bequest motives)	100%	100%	83%	83%	67%	67%	50%	50%
4. Sensitivity to Labor Income Elasticity e_L								
$e_L = 0$	73%	68%	61%	56%	49%	45%	37%	34%
$e_L = 0.5$	79%	75%	66%	62%	53%	50%	40%	37%
5. Optimal Linear Tax Rate in Rentier Society (France 1872–1937) for Zero Receivers (Bottom 80%) With $b^{left} = 25\%$ and $\tau_L = 15\%$								
P0–80, $r - g = 2\%$, $\nu = 70\%$, $e_L = 0.2$	90%		75%		60%		45%	
6. Optimal Top Tax Rate Above Positive Exemption Amount for Zero Receivers (Bottom 50%)								
Exemption amount: 500,000	88%	73%	65%	58%	46%	44%	32%	31%
Exemption amount: 1,000,000	92%	73%	66%	57%	46%	43%	30%	31%

[a]This table presents simulations of the optimal inheritance tax rate τ_B using formula (17) from the main text for France and the United States and various parameter values. In formula (17), we use $\tau_L = 30\%$ (labor income tax rate), except in Panel 5. Parameters $b^{received}$, b^{left}, y_L are obtained from the survey data (SCF 2010 for the U.S., Enquête Patrimoine 2010 for France, and Piketty, Postel-Vinay, and Rosenthal (2011) for panel 5).
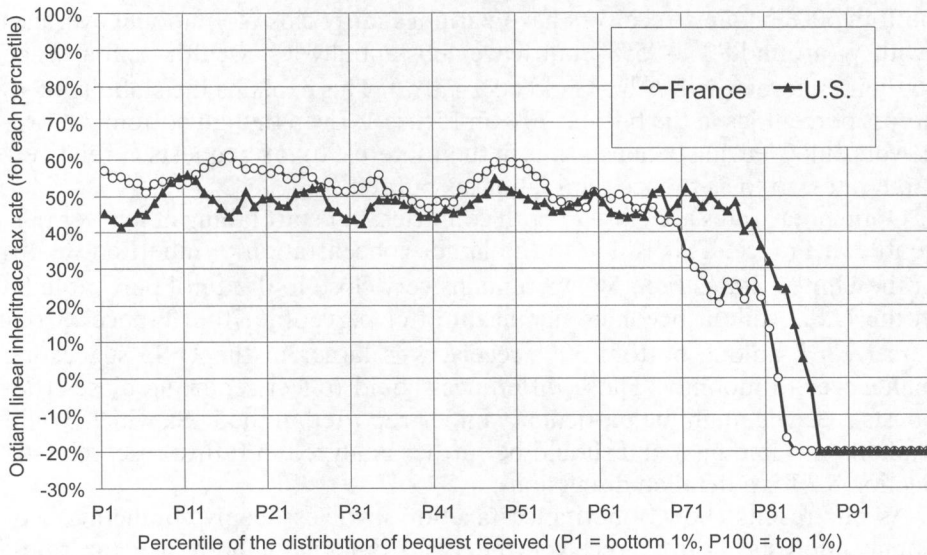
FIGURE 1.—Optimal linear inheritance tax rate (by percentile of bequest received). The figure reports the optimal linear tax rate τ_B from the point of view of each percentile of bequest receivers based on formula (17) in text using as parameters: $e_B = 0.2$, $e_L = 0.2$, $\tau_L = 30\%$, $\nu = 1$ (pure bequest motives), $R/G = 1.8$, y_L, b^{received} and b^{left} estimated from micro-data for each percentile (SCF 2010 for the U.S., Enquête Patrimoine 2010 for France).

and gift receipts available in both surveys to compute $\bar{b}^{\text{received}}$, questions about current net wealth to estimate $\bar{b}^{\text{left}}$, and the sum of wage, self-employment, and pension income (usually proportional to past earnings) to compute $\bar{y}_L$. Wealth of married individuals is defined as household wealth divided by two. Bequest received is defined as the sum of bequests and gifts received by both spouses divided by 2.[18]

Figure 1 depicts the optimal linear inheritance tax rate τ_B from the perspective of each percentile p of the distribution of *bequest received*. We find that, in both countries, the optimal tax rate is about 50% for the bottom 70% of the population, then falls abruptly and becomes negative within the top 20% of inheritors (particularly for the top 10%).[19] Because of the very large concentration of inherited wealth, the bottom 50% receive only about 5% of total bequests in both the United States and France. Hence, $\bar{b}^{\text{received}}$ is close to 0% for the bottom 50%, and barely higher for the next 20%. In both countries,

---

consumption and charitable giving. We repeated the computations separately for individuals aged 60–69, 70–79, 80–89, with almost identical results.

[18]Using transmissible net wealth (excluding pension funds) rather than net wealth or using information on past occupation to estimate $\bar{y}_L$ had very small effects on estimates.

[19]We put a lower bound $\tau_B = -20\%$ for readability, as the optimum is infinitely negative in upper percentiles.

bottom 50% bequest receivers have earnings fairly close to national averages (with $\bar{y}_L$ around 90%–95%), but leave substantially less wealth than average to their children (with $\bar{b}^{\text{left}}$ around 60%–70%). This explains the stability of τ_B across percentiles in the bottom 70% in Figure 1. Even though bottom 70% receivers enjoy leaving bequests, it is in their interest to tax bequests at relatively large rates, so to as reduce their labor tax burden.

Optimal tax rates are close in both countries, but start falling at a lower percentile in France. This is due to the larger concentration of inherited wealth in the United States (i.e., $\bar{b}^{\text{received}}$ remains very close to 0% until percentile 80 in the U.S., while it becomes significant after percentile 70 in France). Conversely, $\bar{b}^{\text{left}}$ among bottom 50% receivers is larger in the U.S., suggesting higher wealth mobility. Those differences could reflect reporting biases (bequests received might be particularly under-reported in the U.S., which would explain both findings) and should be further analyzed in future research (see Section S.3 for a detailed discussion).

As our results show, inheritance taxation involves deeply conflicting economic interests: bottom receivers benefit from high inheritance tax rates, but relatively large groups at the top would benefit from inheritance subsidies. Beliefs about wealth mobility are also key. Over-optimism about the prospect of leaving large bequests would lower perceived optimal tax rates.

Next, we explore, in Table I, sensitivity of optimal τ_B with the key parameters around the benchmark case, $e_L = 0.2$, $\tau_L = 30\%$, $R/G = e^{(r-g)H} = 1.82$, $\nu = 0.7$, and the Meritocratic Rawlsian optimum, that is, maximizing the welfare of bottom 50% bequest-receivers (who receive negligible bequests). In all panels, we display optimal tax rates for France and the United States for various values of the bequest elasticity $e_B = \hat{e}_B = 0, 0.2, 0.5, 1$, the key elasticity parameter.

First, regarding this key bequest elasticity e_B, we chose a benchmark value $e_B = 0.2$ in Figure 1. Using U.S. time and cross-section variations, Kopczuk and Slemrod (2001) found elasticities e_B around 0.1–0.2. There remains considerable uncertainty about e_B. More empirical work on this key parameter would be valuable. With $e_B = 0$, the optimal inheritance tax rate for bottom receivers would be about 70% (rather than 60%). With $e_B = 0.5$, it would be about 50%. Even with an elasticity $e_B = 1$, which seems implausibly high, the optimal inheritance tax rate would still be about 35% in both countries (Table I, panel 1, and Figures S.1–S.2).

Second, we chose a benchmark value $R/G = e^{(r-g)H} = 1.82$, which corresponds to $r - g = 2\%$ and $H = 30$ years. Historically, the difference between the average annual rate of return to wealth and the growth rate has been closer to 3%–4% or even higher (Piketty (2011, Table II, p. 1122)). With $r - g = 3\%$, optimal inheritance tax rates would be close to 70%, both in France and in the United States. Conversely, assuming $r - g = 0\%$, that is, $R/G = 1$, which

can be interpreted as the case with dynamic efficiency and optimal capital accumulation, optimal inheritance tax rates fall to about 40% in both countries (Table I, panel 2, and Figures S.3–S.4).

Third, Table I uses a benchmark value $\nu = 70\%$ for bequest motives strength. Kopczuk and Lupton (2007) showed that there is substantial heterogeneity in the distribution of motives for wealth accumulation. The average fraction of the population with a bequest motive is between one half and two thirds, hence $\nu = 70\%$ is on the high end. With $\nu = 0\%$, that is, in the complete absence of bequest motives, e_B is the sole limiting factor for optimal tax rates, which would then be over 80%. Conversely, with $\nu = 100\%$, that is, wealth accumulation is fully driven by bequest motives, the optimal tax rate would fall to about 50% (Table I, panel 3, and Figures S.5–S.6).

Fourth, Table I, panel 4 shows that optimal bequest tax rates increase with the labor elasticity e_L but moderately so: the optimal τ_B is only slightly higher for $e_L = 0.5$ (upper bound on the labor elasticity based on a large empirical literature) than for $e_L = 0$.

Fifth, to illuminate the crucial role played by wealth inequality and mobility, we also provide estimates using the micro files of estate tax returns collected by Piketty, Postel-Vinay, and Rosenthal (2011) in the Paris archives over the 1872–1937 period, a time characterized by large inheritance flows and extreme wealth concentration (with over 90% of aggregate inheritance received by top 10% successors). These highly reliable, exhaustive administrative data cover wealth over two generations. We find that $\bar{b}^{\text{left}}$ is as low as 20%–30% for the bottom 80% receivers (roughly the zero-receivers), implying very high optimal inheritance tax rates for zero-receivers—typically above 75% for moderate elasticities e_B (Table I, panel 5, and Section S.3).

Sixth, it is possible to extend the optimal linear tax formula to nonlinear bequest taxation that takes the form of a simple two-bracket tax with a flat tax rate above an exemption threshold, a reasonable approximation to actual schedules. Our formula carries over virtually unchanged by replacing bequests by taxable bequests above the exemption threshold in our formulas (see Section S.4).[20] Figure 2 shows that, in both countries, the optimal top tax rate above an exemption level (of 1m \$ or €) is roughly comparable to the optimal linear inheritance tax rate. It is slightly higher in France, where bottom 50% bequest receivers have a relatively small probability to leave bequests above such levels. This difference between France and the United States could again be partly due to reporting biases (see Section S.3). Table I, panel 6 shows that optimal nonlinear tax rates are higher than optimal linear tax rates for moderate e_B (but no longer for large e_B). It is worth noting that these high top inheritance

---

[20]It is computationally more difficult to solve for the optimal exemption threshold (and even more so for the optimal many-bracket nonlinear tax schedule). Hence, in Figure 2 and Table I, panel 6, we take the exemption threshold as given at 500,000 or 1 million \$ (for the U.S.) or € (for France).
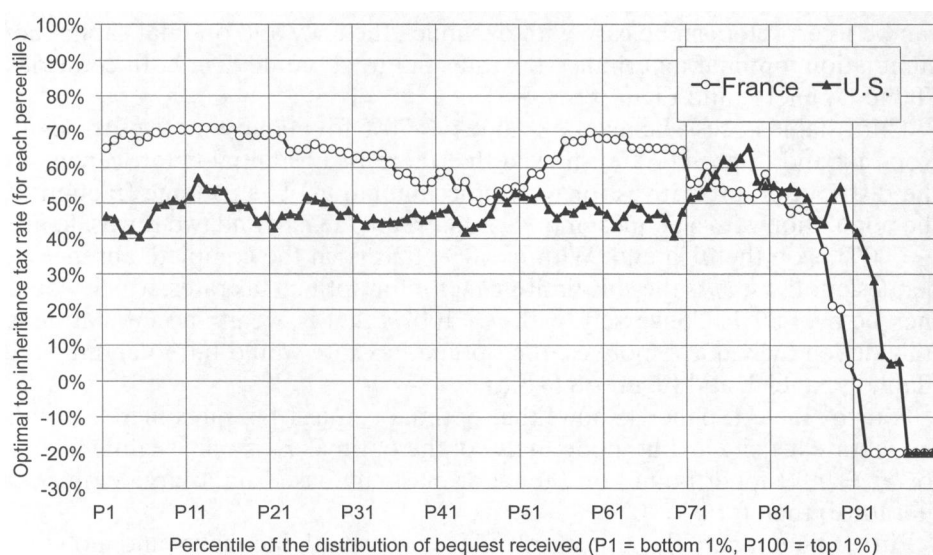
FIGURE 2.—Optimal top inheritance tax rate (by percentile of bequest received). Bequests are taxed at flat rate only above exemption threshold of 1m ($ or €). Optimal top tax rate τ_B (above exemption threshold of 1m ($ or €) for each percentile of bequest received using as parameters: $e_B = 0.2$, $e_L = 0.2$, $\tau_L = 30\%$, $\nu = 1$ (pure bequest motives), $R/G = 1.8$, y_L, $b^{received}$ and b^{left} estimated from micro-data for each percentile (SCF 2010 for the U.S., Enquête Patrimoine 2010 for France). [In contrast to Figure 1, inheritances are taxed only above an exemption threshold in this figure.]

tax rates—around 60%—are very much in line with historical experience, especially in Anglo-Saxon countries from the 1930s to the 1980s, when top estate tax rates were systematically above 60% (Figure 3). The decline of U.S. top rates since the 1980s could be due to a shift in political power away from the bottom 80% and toward the top 10%. Finally, comparing Figure 2 to Figure 1 shows that a smaller minority at the top opposes top bequest taxes than linear bequest taxes, explaining perhaps why actual bequest taxes often have large exemption levels.

5. CONCLUSION AND EXTENSIONS

This paper has derived robust optimal inheritance tax formulas expressed in terms of sufficient statistics. This approach casts fruitful light on the problem and unifies previous seemingly disparate results. In accordance with the public debate, the optimal tax rate trades off equity and efficiency. This trade-off is nondegenerate if the elasticity of bequests with respect to taxation is not infinite and inheritances matter for lifetime resources and social preferences. If the elasticity is low, the concentration of inheritances is high, and society favors
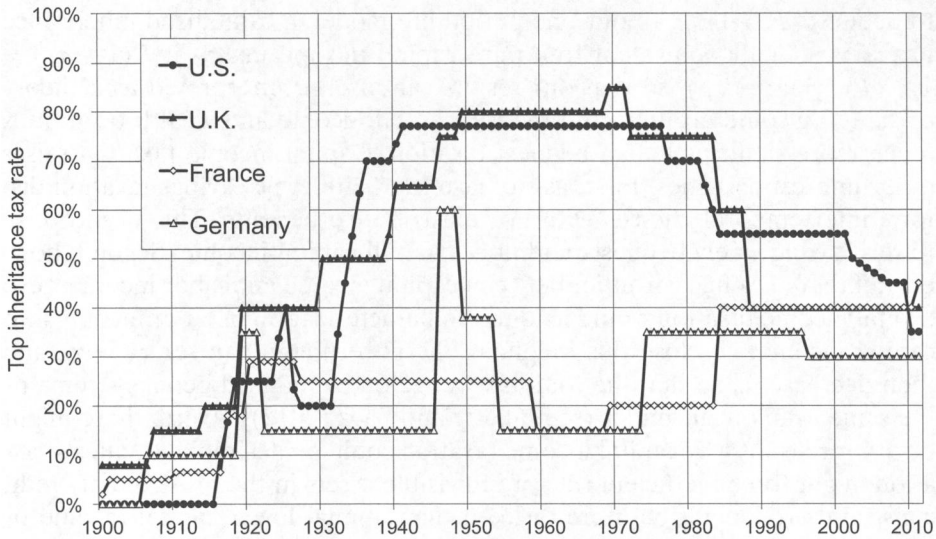
FIGURE 3.—Observed top inheritance tax rates 1900–2011.

those with little inheritance, the optimal tax rate is high. Our analysis could be extended in various ways.

First, solving the full nonlinear optimum (instead of only the two-bracket case) would be valuable. This complicates the analysis but does not radically change the optimal tax problem.

Second, if the government can use debt, labor taxation τ_L is exactly equivalent to a consumption tax τ_C even in the presence of bequests, provided the government compensates individuals for initial wealth implicitly taxed when switching from labor to consumption tax. Hence, the same formulas for τ_B apply when considering the trade-off between bequest taxation and consumption taxation (instead of labor taxation). The view that consumption taxation can successfully tax wealthy idle heirs is illusory because, with labor income taxation, wealthy heirs would have received smaller inheritances to start with. With nonlinear taxation, the full equivalence between labor and consumption tax naturally breaks down. But it is still the case that consumption taxation is a poor instrument to target inheritors, unless inheritance taxes are not available.[21]

Third, our analysis was limited to capitalized inheritance taxation. That is, the same tax rate τ_B is used to tax bequest received b_{ti} and lifetime return

---

[21]This simple point (i.e., with ill functioning inheritance and capital taxes, one can use progressive consumption taxes to tax wealthy successors) was first made by Kaldor (1955). See Piketty and Saez (2012, Appendix B.4).

to bequest $(R - 1)b_{ti}$. In our one-period life model, a capitalized inheritance tax τ_B is actually equivalent to a pure capital income tax τ_K if $R(1 - \tau_B) = 1 + (R - 1)(1 - \tau_K)$, so that our results can also be interpreted as a theory of capital income taxation. In practice, capital income and wealth taxation is much more significant than bequest taxation. Capital income taxation raises other interesting issues. First, as we have seen, life-cycle savings taxation distorts intertemporal choices with no redistributive benefits. This would push toward taxing solely bequests and not tax at all capital income. Second, however, if there is a fuzzy frontier between capital income and labor income, zero capital income taxation would lead to re-characterization of labor income into capital income. To close this loophole, the government can set $\tau_K = \tau_L$ and then decrease τ_B so that the total tax wedge on capitalized bequests remains the same as in our formulas (see Piketty and Saez (2012)). Third, there might be other reasons why capital income taxation could be desirable. Bequests taxation might force inefficient sale of indivisible assets in the presence of credit constraints (or might be more disliked than annual lower capital income or wealth taxes due to fiscal illusion). More importantly, rates of return on capital vary widely across individuals. To the extent that such risk is not optimally diversified, capital income taxation could be desirable for rate of return insurance reasons. That is, with capital market imperfections, lifetime capital income and wealth taxation might be the efficient way to implement optimal inheritance taxes (Piketty and Saez (2012) presented a basic model along those lines).

APPENDIX

A.1. *Proof of Formula (9) With Social Discounting* Δ

We define e_{Bt} as the elastic response of b_t to the tax reform $d\tau = (d\tau_{Bt} = d\tau_B, d\tau_{Lt})_{t \geq T}$, so that $\frac{db_t}{b_t} = -e_{Bt}\frac{d\tau_B}{1-\tau_B}$, where db_t is the aggregate bequest response to the full reform $d\tau$. Note that the response of b_t starts only in period T (as bequest leavers care only about the net-of-tax bequests they leave). The response builds over generations and eventually converges to the long-run elasticity e_B, as defined in (3). We define the elasticity e_{Lt} so that $\frac{dy_{Lt}}{y_{Lt}} = -e_{Lt}\frac{d\tau_{Lt}}{1-\tau_L}$, where dy_{Lt} is the response to the full reform $d\tau$. Period-by-period budget balance requires:

$$(A.1) \quad Rb_t\, d\tau_B\left(1 - e_{Bt}\frac{\tau_B}{1 - \tau_B}\right) = -d\tau_{Lt}y_{Lt}\left(1 - e_{Lt}\frac{\tau_L}{1 - \tau_L}\right).$$

Using the individual FOC $V_c^{ti} = R(1 - \tau_B)V_{\underline{b}}^{ti}$ when $b_{t+1i} > 0$, along with the budget balance equation (A.1), allows us to rewrite the first order condition

$dSWF = 0$ from the text as

$$0 = \sum_{t \geq T} \Delta^t \int_i g_{ti} \left[-d\tau_B R b_{ti}(1 + e_{Bti}) \right.$$

$$\left. + \frac{1 - e_{Bt}\tau_B/(1 - \tau_B)}{1 - e_{Lt}\tau_L/(1 - \tau_L)} \frac{y_{Lti}}{y_{Lt}} R b_t \, d\tau_B \right] - \sum_{t \geq T-1} \Delta^t \int_i g_{ti} \, d\tau_B \frac{b_{t+1i}}{1 - \tau_B}.$$

The third term is a sum starting at $T - 1$ (instead of T), as the reform hurts bequest leavers starting in generation $T - 1$. As everything has converged for $t \geq T$, dividing by $Rb_t \, d\tau_B$ and using $\bar{y}_L$, $\bar{b}^{\text{received}}$, $\bar{b}^{\text{left}}$ defined in (4) and $\hat{e}_{Bt} = \int_i g_{ti} b_{ti} e_{Bti} / \int_i g_{ti} b_{ti}$, the FOC is rewritten as

$$0 = -\sum_{t \geq T} \Delta^t \bar{b}^{\text{received}}(1 + \hat{e}_{Bt}) + \sum_{t \geq T} \Delta^t \frac{1 - e_{Bt}\tau_B/(1 - \tau_B)}{1 - e_{Lt}\tau_L/(1 - \tau_L)} \bar{y}_L$$

$$- \sum_{t \geq T-1} \Delta^t \frac{\bar{b}^{\text{left}}}{R(1 - \tau_B)}.$$

To parallel the analysis of Section 2.2, we define the *discounted elasticities* e_B, $\hat{e}_B$, e_L as follows:

(A.2)   *Discounted bequest elasticities*: $e_B = (1 - \Delta) \sum_{t \geq T} \Delta^{t-T} e_{Bt}$,

$$\hat{e}_B = (1 - \Delta) \sum_{t \geq T} \Delta^{t-T} \hat{e}_{Bt},$$

(A.3)   *Discounted e_L elasticity*: $\dfrac{1 - e_B\tau_B/(1 - \tau_B)}{1 - e_L\tau_L/(1 - \tau_L)}$

$$= (1 - \Delta) \sum_{t \geq T} \Delta^{t-T} \frac{1 - e_{Bt}\tau_B/(1 - \tau_B)}{1 - e_{Lt}\tau_L/(1 - \tau_L)}.$$

Naturally, in the case e_{Lt} constant in t, then we have $e_{Lt} \equiv e_L$. This is the case with isoelastic utilities $U^{ti}(c - l^{1+1/e_L}, \underline{b})$. Using those definitions, we can rewrite the first order condition as

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

where the Δ in the denominator of the third term appears because the sum for the third term starts at $T - 1$ instead of T. Rearranging this expression leads immediately to formula (9).

A.2. *Optimal Taxation in the Farhi–Werning Model*

We can consider steady-state social welfare maximization subject to the generational budget constraint $E_t = \tau_{Bt} b_{t+1} + \tau_L y_{Lt}$ (see Section S.2 of the Supplemental Material) and hence drop any t subscripts. We specialize utility functions to such that $V^i(c, \underline{b}, l) = U^i(u(c, \underline{b}), l)$ with the sub-utility of consumption $u(c, \underline{b})$ homogeneous of degree 1 and homogeneous across individuals in the population.

Let us prove that any budget neutral tax system (τ_B, τ_L, E) can be replaced by an alternative tax system $(\tau_B = 0, \tau'_L, E')$ that leaves all parents' utilities unchanged and raises at least as much revenue. We adapt the Kaplow–Laroque recent and elegant proof of the Atkinson–Stiglitz theorem to the linear earnings tax case. Let us denote by $p = \frac{1}{R}$ and $q = \frac{1}{R(1-\tau_B)}$ the pre-tax and post-tax prices of bequests left $\underline{b}$. Let $v(y, q) = \max_{c, \underline{b} \geq 0} u(c, \underline{b})$ s.t. $c + q\underline{b} \leq y$ be the indirect utility of consumption. Because $u(c, \underline{b})$ is homogeneous of degree 1, $v(y, q)$ is linear in y to that $v(y, q) = y \cdot \phi(q)$, as we show in the case with economic growth in Section S.1.2.

Starting from the initial tax system (τ_B, τ_L, E), let us consider the alternative tax system $(\tau_B = 0, \tau'_L, E')$ such that $\phi(p)(1 - \tau'_L) = \phi(q)(1 - \tau_L)$ and $\phi(p)E' = \phi(q)E$. This alternative system is precisely designed so that $v(y_{Li}(1 - \tau'_L) + E', p) = v(y_{Li}(1 - \tau_L) + E, q)$ for all y_{Li}. Hence, it leaves all parents' utilities and labor supply choices unchanged.

This alternative tax system raises as much revenue as the initial tax system. To see this, suppose individual i chooses $(l_i, c_i, \underline{b}_i)$ under the initial tax system so that $c_i + q\underline{b}_i = w_i l_i(1 - \tau_L) + E$. Attaining utility $v(y_{Li}(1 - \tau'_L) + E', p)$ with $\tau'_B = 0$ costs $y_{Li}(1 - \tau'_L) + E'$. As $v(y_{Li}(1 - \tau_L) + E, q) = v(y_{Li}(1 - \tau'_L) + E', p)$, this utility is also attained under the initial choice $(l_i, c_i, \underline{b}_i)$, which must hence cost at least $y_{Li}(1 - \tau'_L) + E'$ under price p, so that $y_{Li}(1 - \tau'_L) + E' \leq c_i + p\underline{b}_i = y_{Li}(1 - \tau_L) + E - \underline{b}_i(q - p)$. Hence, $\tau_L y_{Li} + (q - p)\underline{b}_i - E \leq \tau'_L y_{Li} - E'$, that is, $\tau_L y_{Li} + \tau_B b_i - E \leq \tau'_L y_{Li} - E'$, so that the new tax system raises at least as much revenue individual by individual.

The alternative system leaves all parents' utilities unchanged and raises more revenue. This implies that $\tau_B = 0$ is optimal when social welfare depends solely on parents' utilities. Adding children's utilities in social welfare makes τ_B less desirable, so that $\tau_B < 0$ is optimal in that case.

A.3. *Anticipated and Long-Run Elasticities in the Dynastic Model*

We provide here detailed intuitions for why the anticipatory elasticity $e_B^{\text{anticip.}}$, the post-reform elasticity e_B^{post}, and the long-run steady-state elasticity e_B are all finite in the ergodic model with stochastic wages (the Aiyagari model) and why they become infinite when stochastic shocks vanish (the Chamley–Judd model). Last, we show that, in the Chamley–Judd model with endogenous discount rate, the long-run elasticity e_B may be finite but the anticipatory elasticity is still infinite. We only provide intuitions rather than fully rigorous detailed

proofs because the formal proof can be immediately obtained by combining the Chamley–Judd and Aiyagari results with our optimal tax formulas. Namely, the fact that Chamley and Judd obtained a zero optimal long-run tax rate implies that the elasticity e_B^{pdv} is infinite. The fact that Aiyagari (1995) obtained a positive optimal long-run tax rate implies that the elasticity e_B^{pdv} is finite.

Nonstochastic Wages (Chamley–Judd)

Let us consider first the standard case with uniform and constant discount rate δ as in the main text. Let us further assume that $\delta R = 1$ and that $\tau_{Bt} \equiv 0$, so that we start from an initial situation with a well-defined steady state.

In the Chamley–Judd model, future wages y_{Lti} are fully known as of period zero. In that case, the natural assumption is that there are no credit constraints and hence the individual first order condition $u'(c_{ti}) = \delta R(1 - \tau_{Bt+1})u'(c_{t+1i})$ always holds.[22]

In that case, with $\delta R = 1$ and $\tau_{Bt} \equiv 0$, the individual fully smoothes consumption $c_{ti} = c_{0i}$ for all t with $c_{0i} = [1 - 1/R](b_{0i} + \sum_{t \geq 0} y_{Lti}/R^t)$ to satisfy the intertemporal budget.

The future tax reform from the main text ($d\tau_{Bt} = d\tau_B$ for $t \geq T$) leads to a decreasing consumption path after the reform and a flat shift of the pre-reform consumption path proportional to $R^{-T}d\tau_B$ (as it affects the PDV of resources by a factor proportional to $R^{-T}d\tau_B$). The aggregated budget constraint implies that $b_{t+1} = Rb_t + y_{Lt} - c_t$ for $t < T$ so that $b_t = R^t b_0 - c_0[1 + R + \cdots + R^{t-1}] + y_{Lt-1} + \cdots + R^{t-1}y_{L0}$ and hence $db_t = -dc_0(R^t - 1)/(R - 1) \sim R^{t-T}d\tau_B$. This implies that $e_{Bt} \sim R^{t-T}$. Therefore, $e_B^{\text{anticip.}} = (1 - \delta)\sum_{t<T}\delta^{t-T}e_{Bt} \sim (1 - \delta)\sum_{t<T}(\delta R)^{t-T} \sim (1 - \delta)T$ is infinite for large T. As is well known, the long-run elasticity e_B is also infinite, as any long-run tax starting from a $\delta R = 1$ steady state leads to an exponentially decreasing path of consumption and hence as much individual debt as possible.

Let us now consider the case with endogenous discount factor $\delta^i(c)$, decreasing in c. In that case, there is a steady state such that $\delta^i(c_i)R = 1$ for all i. Intuitively, if $\delta^i R > 1$, individual i accumulates more wealth, eventually allowing him to consume more, so that $\delta^i R$ is driven down to 1 (and conversely). In steady state (when all variables have converged), $c_i = (R - 1)b_i + y_{Li}$ and b_i is an implicit function of R through the equation $\delta^i((R - 1)b_i + y_{Li})R = 1$. Hence, the individual supply b_i is a smooth function of R. Hence, aggregate long-run bequests b are also a smooth function of R and the long-run elasticity e_B is therefore finite.

It is still the case, however, that a future reform shifts the entire (pre-reform) consumption path so that $dc_{0i} \sim R^{-T} d\tau_B$, which implies $db_t \sim R^{t-T} d\tau_B$, $e_{Bt} \sim$

[22]This will be true for large t without any assumption if y_{Lti} converges to a constant y_{Li} for large t, the natural assumption for steady-state reasoning.

R^{t-T}. Therefore, $e_B^{\text{anticip.}} \sim T$ is infinite for large T. This implies that the optimal long-run tax rate τ_B is zero in spite of a finite long-run elasticity e_B, an important point made by Judd (1985, Theorem 5).

Stochastic Wages (Aiyagari)

In the stochastic model (where we consider only wage shocks and no preference shocks for simplicity of exposition), individual ti solves the problem

$$\max_{b_{t+1i}} u\big(Rb_{ti}(1 - \tau_{Bt}) + \tau_{Bt}Rb_t + y_{Lti} - b_{t+1i}\big)$$

$$+ \sum_{s>t} \delta^{s-t} E_t u\big(Rb_{si}(1 - \tau_{Bs}) + \tau_{Bs}Rb_s + y_{Lsi} - b_{s+1i}\big).$$

Consider again the future tax reform $(d\tau_{Bt'} = d\tau_B$ for $t' \geq T)$. In any path where $b_{si} = 0$ for some s such that $t < s \leq T$, any current marginal bequest change db_{t+1i} has zero impact on post s generations, and hence the future tax rate change is irrelevant for the current decision b_{t+1i}. Concretely, in the scenario where my child fully consumes my bequest and leaves nothing to my grandchildren, a marginal increase in bequest taxes for my grandchildren does not affect my bequest decision. Hence, the behavioral response db_{t+1i} to the future tax increase is discounted relative to the Chamley–Judd model with no uncertainty by a factor $\kappa^{ti}(T - (t + 1))$, which is the probability that all my descendants from $s = t + 1$ to $s = T$ all leave positive bequests $b_{si} > 0$.

By ergodicity, as initial wealth is irrelevant in the distant future, for $T - s$ large, $\kappa^{ti}(T - (s + 1))/\kappa^{ti}(T - s)$ converges to a constant $\kappa < 1$ that depends on the structure of shocks, the tax system, etc. but is uniform across individuals. This constant κ is equal to the fraction of individuals with positive bequests in the ergodic cross-section. Hence, κ is necessarily less than 1 as long as the fraction of individuals with zero bequests is strictly positive in steady state. Naturally, when uncertainty in future labor shocks vanishes, κ converges to 1.

Hence, at the aggregate level, the response db_t to the future tax increase starting at date T is reduced by an exponential factor proportional to κ^{T-t}.

As we have seen, in the Chamley–Judd model with no uncertainty and $\delta R = 1$, we have $db_t \sim \delta^{T-t} d\tau_B$. Hence, with stochastic shocks, $db_t \sim \delta^{T-t}\kappa^{T-t} d\tau_B$, so that $e_{Bt} \sim \delta^{T-t}\kappa^{T-t}$. This implies that $e_B^{\text{anticip.}} = (1 - \delta)\sum_{t<T} \delta^{t-T} e_{Bt} \sim (1 - \delta)\sum_{t<T} \kappa^{T-t} = (1 - \delta)/(1 - \kappa)$ is finite.

In the ergodic long run, with stochastic shocks, aggregate bequests b_t converge, so naturally b_t will be a smooth function of τ_B.[23] In that case, the long-run elasticity e_B is finite. This also implies that the post-elasticity $e_B^{\text{post}} = (1 - \delta)\sum_{t\geq T} \delta^{t-T} e_{Bt}$ is finite, which establishes that $e_B^{\text{pdv}} = e_B^{\text{anticip.}} + e_B^{\text{post}}$ is finite and delivers a nonzero optimal τ_B, as in Aiyagari (1995).

---

[23]For $\tau_B = 1$, there is no incentive to leave bequests and $b_t = 0$. Conversely, for sufficiently large subsidies, if $\delta R(1 - \tau_B) > 1$, then bequests b_t would explode. In between, b_t is a smooth function of $1 - \tau_B$.

REFERENCES

AIYAGARI, S. R. (1995): "Optimal Capital Income Taxation With Incomplete Markets, Borrowing Constraints, and Constant Discounting," *Journal of Political Economy*, 103 (6), 1158–1175. [1871,1883,1884]

ATKINSON, A., AND A. SANDMO (1980): "Welfare Implications of the Taxation of Savings," *Economic Journal*, 90, 529–549. [1862]

ATKINSON, A., AND J. E. STIGLITZ (1976): "The Design of Tax Structure: Direct Versus Indirect Taxation," *Journal of Public Economics*, 6 (1–2), 55–75. [1852,1853]

CHAMLEY, C. (1986): "Optimal Taxation of Capital Income in General Equilibrium With Infinite Lives," *Econometrica*, 54 (3), 607–622. [1851]

——— (2001): "Capital Income Taxation, Wealth Distribution and Borrowing Constraints," *Journal of Public Economics*, 79, 55–69. [1872]

CREMER, H., AND P. PESTIEAU (2004): "Wealth Transfer Taxation: A Survey of the Theoretical Literature," in *Handbook on Altruism, Giving and Reciprocity*. Amsterdam: North-Holland. [1852]

DEATON, A. (1979): "Optimally Uniform Commodity Taxes," *Economic Letters*, 2, 357–361. [1865]

DIAMOND, P., AND J. A. MIRRLEES (1971): "Optimal Taxation and Public Production: Production Efficiency," *American Economic Review*, 61 (1), 8–27. [1861]

FARHI, E., AND I. WERNING (2010): "Progressive Estate Taxation," *Quarterly Journal of Economics*, 125 (2), 635–673. [1852,1853,1864-1866]

FLEURBAEY, M. (2008): *Fairness, Responsibility and Welfare*. Oxford: Oxford University Press. [1858]

HENDREN, N. (2013): "The Policy Elasticity," Working Paper, Harvard. [1855]

JUDD, K. (1985): "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28 (1), 59–83. [1851,1871,1884]

KALDOR, N. (1955): *An Expenditure Tax*. London: Unwin University Books. [1879]

KAPLOW, L. (2001): "A Framework for Assessing Estate and Gift Taxation," in *Rethinking Estate and Gift Taxation*, ed. by W. Gale, J. Hines, and J. Slemrod. Washington, D.C.: Brookings Institution Press. [1852]

KING, M. (1980): "Savings and Taxation," in *Public Policy and the Tax System*, ed. by G. Hughes and G. Heal. London: George Allen Unwin, 1–36. [1862]

KOCHERLAKOTA, N. (2010): *The New Dynamic Public Finance*. Princeton: Princeton University Press. [1853]

KOPCZUK, W. (2013): "Taxation of Transfers and Wealth," in *Handbook of Public Economics*, Vol. 5. Amsterdam: North-Holland. [1852]

KOPCZUK, W., AND J. LUPTON (2007): "To Leave or Not to Leave: The Distribution of Bequest Motives," *Review of Economic Studies*, 74 (1), 207–235. [1866,1873,1877]

KOPCZUK, W., AND J. SLEMROD (2001): "The Impact of the Estate Tax on Wealth Accumulation and Avoidance Behavior," in *Rethinking Estate and Gift Taxation*, ed. by W. Gale, J. R. Hines, and J. Slemrod. Washington, D.C.: Brookings Institution Press, 299–349. [1876]

PIKETTY, T. (2000): "Theories of Persistent Inequality and Intergenerational Mobility," in *Handbook of Income Distribution*, ed. by A. B. Atkinson and F. Bourguignon. Amsterdam: North-Holland. [1869]

——— (2011): "On the Long-Run Evolution of Inheritance: France 1820–2050," *Quarterly Journal of Economics*, 126 (3), 1071–1131. [1858,1864,1876]

PIKETTY, T., AND E. SAEZ (2012): "A Theory of Optimal Capital Taxation," Working Paper 17989, NBER. [1854,1862,1879,1880]

——— (2013a): "Optimal Labor Income Taxation," in *Handbook of Public Economics*, Vol. 5. Amsterdam: North-Holland. [1852]

——— (2013b): "Supplement to 'A Theory of Optimal Inheritance 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]

PIKETTY, T., G. POSTEL-VINAY, AND J.-L. ROSENTHAL (2011): "Inherited vs Self-Made Wealth: Theory and Evidence From a Rentier Society (1872–1937)," Working Paper, PSE. [1874,1877]
SAEZ, E. (2002): "The Desirability of Commodity Taxation Under Nonlinear Income Taxation and Heterogeneous Tastes," *Journal of Public Economics*, 83, 217–230. [1866]
SAEZ, E., AND S. STANTCHEVA (2013): "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," Working Paper 18835, NBER. [1858]

Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France; piketty@pse.ens.fr

and

Dept. of Economics, University of California at Berkeley, 530 Evans Hall, Berkeley, CA 94720, U.S.A., and NBER; saez@econ.berkeley.edu.