NBER WORKING PAPER SERIES

A THEORY OF OPTIMAL CAPITAL TAXATION

Thomas Piketty
Emmanuel Saez

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

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

We are grateful to seminar participants at the Paris School of Economics, the London School of Economics, the National Bureau of Economic Research (Public Economics Program), the University of California at Berkeley, the Massachussetts Institute of Technology, and Stanford University for their comments. We thank Tony Atkinson, Alan Auerbach, Peter Diamond, Emmanuel Farhi, Mikhail Golosov, Louis Kaplow, Wojciech Kopczuk, Matt Weinzierl and Ivan Werning for helpful and stimulating discussions. We acknowledge financial support from the Center for Equitable Growth at UC Berkeley. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

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

A Theory of Optimal Capital Taxation
Thomas Piketty and Emmanuel Saez
NBER Working Paper No. 17989
April 2012
JEL No. H21

ABSTRACT

This paper develops a realistic, tractable theoretical model that can be used to investigate socially-optimal capital taxation. We present a dynamic model of savings and bequests with heterogeneous random tastes for bequests to children and for wealth per se. We derive formulas for optimal tax rates on capitalized inheritance expressed in terms of estimable parameters and social preferences. Under our model assumptions, the long-run optimal tax rate increases with the aggregate steady-state flow of inheritances to output, decreases with the elasticity of bequests to the net-of-tax rate, and decreases with the strength of preferences for leaving bequests. For realistic parameters of our model, the optimal tax rate on capitalized inheritance would be as high as 50%-60%–or even higher for top wealth holders–if the social objective is meritocratic (i.e., the social planner puts higher welfare weights on those receiving little inheritance) and if capital is highly concentrated (as it is in the real world). In contrast to the Atkinson-Stiglitz result, the optimal tax on bequest remains positive in our model even with optimal labor taxation because inequality is two-dimensional: with inheritances, labor income is no longer the unique determinant of lifetime resources. In contrast to Chamley-Judd, the optimal tax on capital is positive in our model because we have finite long run elasticities of inheritance to tax rates. Finally, we discuss how adding capital market imperfections and uninsurable shocks to rates of return to our optimal tax model leads to shifting one-off inheritance taxation toward lifetime capital taxation, and can account for the actual structure and mix of inheritance and capital taxation.

Thomas Piketty
Paris School of Economics
48 Boulevard Jourdan
75014 Paris, France
piketty@ens.fr

Emmanuel Saez
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

1 Introduction

According to the profession's most popular theoretical models, optimal tax rates on capital should be equal to zero in the long run–including from the viewpoint of those individuals or dynasties who own no capital at all. Taken literally, the policy implication of those theoretical results would be to eliminate all inheritance taxes, property taxes, corporate profits taxes, and individual taxes on capital income and recoup the resulting tax revenue loss with higher labor income or consumption or lump-sum taxes. Strikingly, even individuals with no capital or inheritance would benefit from such a change. E.g. according to these models it is in the interest of propertyless individuals to set property taxes to zero and replace them by poll taxes.

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

However, in the absence of an alternative tractable model, the zero capital tax result remains an important reference point in economics teaching and in policy discussions.[3] For instance, a number of economists and policy-makers support tax competition as a way to impose zero optimal capital taxes to reluctant governments.[4] We view the large gap between optimal capital tax theory and practice as one of the most important failures of modern public economics.

The objective of this paper is to develop a realistic, tractable, and robust theory of socially optimal capital taxation. By realistic, we mean a theory providing optimal tax conclusions that are not fully off-the-mark with respect to the real world (i.e., positive and significant capital tax rates–at least for some parameter values). By realistic, we also mean a theory offering such conclusions for reasons that are consistent with the reasons that are at play in the real world which–we feel–are related to the large concentration of inherited capital ownership.

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

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

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

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

By tractable, we mean that optimal tax formulas should be expressed in terms of estimable parameters and should quantify the various trade-offs in a simple and plausible way. By robust, we mean that our results should not be too sensitive to the exact primitives of the model nor depend on strong homogeneity assumptions for individual preferences. Ideally, formulas should be expressed in terms of estimable "sufficient statistics" such as distributional parameters and behavioral elasticities and hence be robust to changes in the underlying primitives of the model.[5]

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

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

We derive formulas for optimal tax rates τ_B on capitalized inheritance expressed in terms of estimable parameters and social preferences. The long run optimal tax rate τ_B increases with the aggregate steady-state flow of bequests to output b_y, decreases with the elasticity of bequests with respect to the net-of-tax rate e_B, and decreases with the strength of preferences

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

for bequests s_{b0}. Under the assumptions of our model, for realistic parameters, the optimal linear tax rate on capitalized inheritance would be as high as $50\% - 60\%$ under a meritocratic social objective preferences (i.e., those with little inheritance have high welfare weight in the social objective function). Because real world inherited wealth is highly concentrated–half of the population receives close to zero bequest, our results are robust to reasonable changes in the social welfare objective. For example, the optimal tax policy from the viewpoint of those receiving zero bequest is very close to the welfare optimum for bottom 50% bequest receivers. Interestingly, the optimal tax rate τ_B imposed on top wealth holders can be even larger (say, $70\% - 80\%$), especially if bequest flows are large, and if the probability of bottom receivers to leave a large bequest is small. Therefore our model can generate optimal tax rates as large as the top bequest tax rates observed in most advanced economies during the past 100 years, especially in Anglo-Saxon countries from the 1930s to the 1980s (see Figure 1). To our knowledge, this is the first time that a model of optimal inheritance taxation delivers tractable and estimable formulas that can be used to analyze such real world tax policies.

Our model also illustrates the importance of perceptions and beliefs systems about wealth inequality and mobility (i.e. individual most preferred tax rates are very sensitive to expectations about bequests received and left), and about the magnitude of aggregate bequest flows. When bequest flows are small, (e.g., 5% of national income, as was the case in Continental Europe during the 1950s-1970s), then optimal bequest taxes in our model would be moderate. When they are large (e.g., 15% of national income as in France currently or over 20% as in the 19th century France), then optimal bequest taxes in our model would be large–so as to reduce the tax burden falling on labor earners.[6]

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

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

higher than the optimal bequest tax rate $\widetilde{\tau}_B$, and labor tax rate τ_L. This is consistent with the fact that in modern tax systems the bulk of aggregate capital tax revenues comes from lifetime capital taxes (rather than from inheritance taxes). It is also interesting to note that the countries which experienced the highest top inheritance tax rates also applied the largest tax rates on top incomes, and particularly so on tax capital incomes (see Figures 2-3). To our knowledge this is the first time that a model of optimal capital taxation can provide a rational for why these various policy tools can indeed be complementary.

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

2 Relation to Existing Literature

There are two main results in the literature in support of zero capital income taxation: Atkinson-Stiglitz and Chamley-Judd. We discuss each in turn and then discuss the more recent literature.

Atkinson-Stiglitz. Atkinson and Stiglitz (1976) show that there is no need to supplement the optimal non-linear labor income tax with a capital income tax in a life-cycle model if leisure choice is (weakly) separable from consumption choices and preferences for consumption are homogeneous. In that model, the only source of lifetime income inequality is labor skill and hence there is no reason to redistribute from high savers to low savers (i.e. tax capital income) conditional on labor earnings.[7] This key assumption of the Atkinson-Stiglitz model breaks down in a model with inheritances where inequality in lifetime income comes from both differences in labor income and differences in inheritances received. In that context and conditional on labor earnings, a high level of bequests left is a signal of a high level of inheritances received, which

[7]Saez (2002) shows that this result extends to heterogeneous preferences as long as time preferences are orthogonal to labor skills. If time preferences are correlated with labor skills, then the optimal tax on saving is positive as it is an indirect way to tax ability. Golosov et al. (2011) calibrate a model where higher skills individuals have higher saving taste and show that the resulting optimal capital income tax rate depends significantly on the inter-temporal elasticity of substitution but that the implied welfare gains are relatively small in all cases.

provides a rationale for taxing bequests. To see this, consider a model with inelastic and uniform labor income but with differences in inheritances due to parental differences in preferences for bequests. In such a model, labor income taxation is useless for redistribution but taxing inheritances generates redistribution. This important point has been made by Cremer, Pestieau, and Rochet (2003) in a stylized partial equilibrium model with unobservable inherited wealth where the optimal tax on capital income becomes positive. Our model allows the government to directly observe (and hence tax) inherited wealth.

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

Chamley-Judd. Chamley (1986) and Judd (1985) show that the optimal capital income tax would be zero in the long-run. This zero long-run result holds for two reasons.

First, and as originally emphasized by Judd (1985), the zero rate results happens because social welfare is measured exclusively from the initial period (or dynasty). In that context, a constant tax rate on capital income creates a tax distortion growing exponentially overtime–which cannot be optimal (see Judd 1999 for a clear intuitive explanation). Such a welfare criterion can only make sense in a context with homogeneous discount rates. In the context of inheritance taxation where each period is a generation and where preferences for bequests are heterogeneous across the population, this does not seem like a valid social welfare objective as children of parents with no tastes for bequests would not be counted in the social welfare function. We will adopt instead a definition of social welfare based on long-run equilibrium steady-state utility.[9] We show in appendix C how the within generation and across generation

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

[9]In models with dynamic uncertainty, using the initial period social welfare criteria leads to optimal policies where inequality grows without bounds (see e.g. Atkeson and Lucas 1992). Obtaining "immiseration" as an optimal redistributive tax policy is not realistic and can be interpreted as a failure of the initial period social welfare criterion. Importantly, Farhi and Werning (2007) show that considering instead the long-run steady-state

redistribution problems can be disconnected using public debt so that there is essentially no loss of generality in focusing on steady state welfare.

Second, even adopting a long-run steady-state utility perspective, the optimal capital income tax rate is still zero in the standard Chamley-Judd model. This is because the supply side elasticity of capital with respect to the net-of-tax return is infinite in the standard infinite horizon dynastic model with constant discount rate.[10] The textbook model predicts enormous responses of aggregate capital accumulation to changes in capital tax rates, which just do not seem to be there in historical data. Capital-output ratios are relatively stable in the long run, in spite of large variations in tax rates (see e.g. Piketty, 2010, p.52). Our theory leaves this key elasticity as a free parameter to be estimated empirically. Our model naturally recovers the zero capital tax result of Chamley-Judd when the elasticity is infinite.

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

Methodologically, the new dynamic public finance solves for the fully optimal mechanism and hence obtains optimal tax systems that can be complex and history dependent, in contrast to actual practice. We instead limit ourselves to very simple (and more realistic) tax structures.

---

equilibrium as we do in this paper eliminates the immiseration results.

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

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

This allows us to consider richer heterogeneity in preferences which we believe is important in the case of bequests.[12] Therefore, we also view our methodological approach as complementary to this literature (Diamond and Saez, 2011 for a longer discussion of this methodological debate).

Capital market imperfections. A number of papers have shown that the optimal tax on capital income can become positive when capital market imperfections are introduced, even in models with no inheritance. Typically, the optimal capital income tax is positive because it is a way to redistribute from those with no credit constraints (the owners of capital) toward those with credit constraints (non-owners of capital). Aiyagari (1995) and Chamley (2001) make this point formally in a model with borrowing constrained infinitely lived agents facing labor income risk. They show that optimal capital income taxation is positive when consumption is positively correlated with savings[13] but do not attempt to compute numerical values for optimal capital tax rates. Farhi and Werning (2011) (cited above) also propose a quantitative calibration of an infinite horizon model with borrowing constraints but they find small welfare gains from capital taxation. In contrast, Conesa, Kitao, and Krueger (2009) calibrate an optimal tax OLG life-cycle model with uninsurable idiosyncratic labor productivity shocks and borrowing constraints, and find $\tau_K = 36\%$ and $\tau_L = 23\%$ in their preferred specification. The main effect seems to be that capital income tax is an indirect way to tax more the old and to tax less the young, so as to alleviate their borrowing constraints. While this is an interesting mechanism, we do not believe that this is the most important explanation for $\tau_K > 0$. There are other more direct ways to address the issue of taxing the young vs. the old (e.g. age-varying income taxes; some policies, e.g. pension schemes, do depend on age).[14] In contrast, the theory of capital taxation offered in the present paper is centered upon the interaction between inheritance and capital market imperfections.[15]

Government time-inconsistency and lack of commitment. Yet another way to explain real-world, positive capital taxes is to assume time inconsistency and lack of commitment.[16] Zero capital tax results are always long run results. In the short run, capital is on the table, and

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

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

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

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

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

it is always tempting for short-sighted governments to have $\tau_K > 0$, even though the optimal long run τ_K is equal to 0%. More generally, if governments cannot commit to long run policies, they will always be tempted to renege on their past commitments and to implement high capital tax rates, even though this is detrimental to long run welfare.

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

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

3 The Model

3.1 Notations and Definitions

We consider a small open economy facing an exogenous, instantaneous rate of return on capital $r \geq 0$. To keep notations minimal, we focus upon a simple model with a discrete set of generations $0, 1, .., t, ..$ Each generation has measure one, lives one period (which can be interpreted as H-year-long, where $H =$ generation length, realistically around 30 years), then dies and is replaced by the next generation. Total population is stationary and equal to $N_t = 1$, so aggregate variables Y_t, K_t,L_t, B_t, and per capita variables y_t, k_t, l_t, b_t, are identical (we use the latter).

Generation t receives average inheritance (pre-tax) b_t from generation $t-1$ at the beginning

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

of period t. Inheritances go into the capital stock and are invested either domestically or abroad for a "generational" rate of return $1+R=e^{rH}$. Production in generation t combines labor from generation t and capital to produce a single output good. The output produced by generation t is either consumed by generation t or left as bequest to generation $t+1$. We denote by y_{Lt} the average labor income received by generation t. We denote by c_t the average consumption of generation t and b_{t+1} the average bequest left by generation t to generation $t+1$. We assume that output, labor income, and capital income are realized at the end of period. Consumption c_t and bequest left b_{t+1} also take place at the end of the period. This condensed timing greatly simplifies the notations and exposition of the model but is unnecessary for our results.[18]

Individual i in generation t maximizes utility:

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

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

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

c_{ti} = consumption

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

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

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

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

In order to fix ideas, consider the special Cobb-Douglas (or log-log) case:

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

This simple form implies that individual i devotes a fraction s_i of his lifetime resources to end-of-life wealth, and a fraction $1-s_i$ to consumption. The parameters s_{wi} and s_{bi} measure the tastes for wealth per se and for bequest (more on this below).

In the general case with $V_i(c, w, \bar{b})$ homogeneous of degree one, the fraction s_i of lifetime resources saved depends on $(1-\tau_B)e^{rH}$, i.e., the relative price of bequests. Using the first order condition of the individual $V_{ic} = V_{iw} + (1-\tau_B)e^{rH}V_{i\bar{b}}$, we can then define $s_{bi} = s_i \cdot (1-$

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

$\tau_B)e^{rH}V_{i\bar{b}}/V_{ic}$ and $s_{wi} = s_i \cdot V_{iw}/V_{ic}$. Hence, s_i, s_{wi}, and s_{bi} are functions of $(1-\tau_B)e^{rH}$ instead of being constant as with Cobb-Douglas where income and substitution effects cancel out.

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

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

Taking as given the generational rate of return $R = e^{rH} - 1$, profit maximization implies that the domestic capital input k_t is chosen so that $F_K = R$, i.e. $k_t = \beta^{\frac{1}{1-\alpha}} l_t$ (with $\beta = \dfrac{k_t}{y_t} = \dfrac{\alpha}{R} =$ domestic generational capital-output ratio).[19] It is important to keep in mind that y_t is domestic output. In the open economy case we consider, y_t might differ from national income if the domestic capital stock k_t (used for domestic production) differs from the national wealth b_t.

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

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

3.2 Steady-state Inheritance Flows and Distributions

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

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

In our model, there are three independent factors explaining why different individuals receive different bequests b_{t+1i} within generation $t+1$: their parents received different bequests b_{ti}, earned different labor income y_{Lti}, or had different tastes for savings $s_{ti} = s_{wti} + s_{bti}$.[20]

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

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

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

In the real world, most individuals are at the same time life-cycle savers, wealth-lovers and bequest-lovers. But the exact magnitude of these various saving motives does vary a lot across individuals and over generations, just like other tastes.[21] We allow for any exogenous distribution for taste parameters $g(s_{wi}, s_{bi})$. For notational simplicity, we assume that tastes are drawn i.i.d. at each generation from the distribution $g(s_{wi}, s_{bi})$. Hence they are independent across individuals within a generation and independent across generations within a dynasty. In the Cobb-Douglas case, the parameters s_{wi}, s_{bi} are fixed independently of τ_B. In the general homogeneous of degree one case, the parameters s_{wi}, s_{bi} depend upon $(1-\tau_B)e^{rH}$ and hence are not strictly parameters. We adopt this slight abuse of notation for presentational simplicity.[22]

Assumption 1 *Taste parameters (s_{wi}, s_{bi}) are drawn i.i.d. at each generation from an exogenous distribution $g(s_{wi}, s_{bi})$ defined over a set of possible tastes $S \subset \overline{S}$ (where $\overline{S}$ is the set of all possible tastes: $\overline{S} = \{(s_{wi}, s_{bi})$ s.t. $s_{wi}, s_{bi} \geq 0$ and $s_i = s_{wi} + s_{bi} \leq 1\}$).*
S and $g(\cdot)$ can be discrete or continuous. We denote by $s_0 = \min\{s_i = s_{wi} + s_{bi} \in S\}$, $s_1 = \max\{s_i = s_{wi} + s_{bi} \in S\}$, with $0 \leq s_0 \leq s_1 \leq 1$, and $s = E(s_i)$ the average taste.
We assume that S includes zero saving tastes and at least one other taste: $s_0 = 0$, $s_1 > 0$.

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

Productivity Shocks. Labor productivity shocks are specified as follows. Individual i in generation t has a within-cohort normalized productivity parameter $\theta_{ti} = h_{ti}/h_t$. By definition,

[21] Kopczuk and Lupton (2007) and Kopczuk (2009, 2012) present evidence on heterogeneity in bequest motives.
[22] Rigorously, we would need to parametrize utility functions so that $s_{bi} = s_b(\sigma_{bi}, (1-\tau_B)e^{rH}), s_{wi} = s_w(\sigma_{wi}, (1-\tau_B)e^{rH})$ with $(\sigma_{bi}, \sigma_{wi})$ i.i.d parameters and $s_b(.)$ and $s_w(.)$ fixed functions.
[23] This could result from other types of shocks (see example below).

we have: $y_{Lti} = \theta_{ti} y_{Lt}$ (with $E(\theta_{ti}) = 1$). Productivity differentials θ_{ti} could come from innate abilities, acquired skills, individual occupational choices, or sheer luck–and most likely from a complex combination between the four. We assume that productivity shocks are drawn i.i.d. from the same distribution $h(\theta_i)$ at each generation and independently of savings tastes.

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

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

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

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

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

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

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

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

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

Finally, we denote by $z_{ti} = b_{ti}/b_t$ the within-cohort normalized bequest, and $\phi_t(z)$ the distribution of normalized bequest within cohort t. Given some initial distribution $\phi_0(z)$, the random processes for tastes and productivity $g(\cdot)$ and $h(\cdot)$ and the individual transition equation (1) entirely determine the low of motion for the distribution of inheritance $\phi_t(z)$ and the joint distribution of inheritance and labor productivity, which we denote by $\psi_t(z, \theta) = \phi_t(z) \cdot h(\theta)$.

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

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

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

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

Two points are worth noting. First, the aggregate magnitude of inheritance flows relative to output b_y grows with $r - g$. With high returns and low growth, wealth coming from the past is being capitalized at a faster rate than national income. Successors simply need to save a small fraction of their asset returns to ensure that their inherited wealth grows at least as fast as output. The multiplicative factor associated to intergenerational wealth transmission is large and leads to high inheritance flows. Conversely, with low returns and high growth, inheritance is dominated by new wealth, and the steady-state aggregate inheritance flow is a small fraction of output. As shown in Piketty (2011), this simple r-vs-g model is able to reproduce remarkably well the observed evolution of aggregate inheritance flows over the past two centuries. In particular, it can explain why inheritance flows were so large in the 19^{th} and early 20^{th} centuries (20%-25% of national income in 1820-1910), so low in the mid-20^{th} century (less than 5% around 1950-1960), and why they are becoming large again in the late 20^{th} and early 21^{st} centuries (about 15% in 2010 in France) (see Figures 4-5). With $r = 4\% - 5\%$ and $g = 1\% - 2\%$, simple calibrations of the above formula show that the annual inheritance flow b_y can indeed be as large as $20\% - 25\%$ of national income.[24] Available evidence suggests that the

[24] E.g. with $r - g = 3\%$, $H = 30$, $\alpha = 30\%$, $s = 10\%$, $\tau_B = \tau_L = 0\%$, then $b_y = 23\%$. With $r - g = 2\%$, then $b_y = 16\%$. With $r - g = 3\%$ and $\tau = 30\%$, then $b_y = 13\%$, but $b_y/(1-\tau) = 19\%$. The b_y formula given above relates to the generational, (capitalized bequest)/output ratio, while the empirical estimates depicted on Figures 4-5 refer to the cross-sectional, non-capitalized ratio. But one can show that both ratios are very

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

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

3.3 An Example with Binomial Random Tastes

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

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

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

...

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

That is, the steady-state distribution $\phi(z)$ is unbounded above and has the standard Pareto asymptotic upper tail found in empirical data and in wealth accumulation models with random multiplicative shocks (see Appendix A1 and Atkinson, Piketty and Saez (2011)). Inheritances

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

are obviously uncorrelated with labor income (since there is no inequality of labor income).

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

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

3.4 The Optimal Tax Problem

We now formally define our optimal tax problem. We assume that the government faces an exogenous revenue requirement: per capita public good spending must satisfy $g_t = \tau y_t$ where $\tau \geq 0$ is taken as given and y_t is exogenous per capita domestic output. We first assume that the government has only two tax instruments: a proportional tax on labor income at rate $\tau_L \geq 0$, and a proportional tax on capitalized inheritance at rate $\tau_B \geq 0$. We impose a period-

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

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

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

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

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

Assumption 4 $\tau < 1 - \alpha$

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

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

Example. Assume $r = 4\%, H = 30$, so that $e^{rH} = 1 + R = 3.32$, i.e. $R = 2.32$. *If* $\tau_B = 20\%$ *then* $\tau_K = 29\%$. *If* $\tau_B = 40\%$ *then* $\tau_K = 57\%$. *If* $\tau_B = 60\%$ *then* $\tau_K = 86\%$.

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

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

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

The question that we now ask is the following: what is the tax policy (τ_L, τ_B) maximizing long-run, steady-state social welfare? That is, we assume that the government can commit for ever to a tax policy $(\tau_{Lt} = \tau_L, \tau_{Bt} = \tau_B)_{t \geq 0}$ and cares only about the long-run steady-state distribution of welfare V_{ti}. Under assumptions 1-4, for any tax policy there exists a unique steady-state ratio b_y and distribution $\psi(z, \theta)$. The government chooses (τ_L, τ_B) so as to maximize the following, steady-state social welfare function:[31]

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

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

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

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

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

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

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

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

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

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

balance, see equation (4)):

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

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

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

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

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

4 Basic Optimal Capital Tax Formula

4.1 The Zero-Bequest-Receiver Social Optimum

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

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

Proposition 2 ***(zero-bequest-receiver optimum).*** *Under assumptions 1-4, linear social welfare ($\Gamma = 0$), and the welfare weights: $\omega_{p_z p_\theta} = 1$ if $p_z = 0$, and $\omega_{p_z p_\theta} = 0$ if $p_z > 0$, then:*

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

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

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

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

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

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

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

The first order condition in b_{t+1i} is $V_{ci} = V_{wi} + (1 - \tau_B)(1 + R) V_{\underline{b}i}$ This leads to $b_{t+1i} = s_i (1 - \tau_L) y_{Lti}$ (with $0 \leq s_i \leq 1$). Recall that $s_{bi} = s_i \cdot (1 - \tau_B)(1 + R) V_{\underline{b}i} / V_{ci}$.[35]

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

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

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

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

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

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

[35] In the Cobb-Douglas utility case, s_{bi} is simply the fixed exponent in the utility function. In the general homogeneous utility case, s_{bi} may depend on τ_B and $1 + R$.

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

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

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

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

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

Note 3. We also show in the appendix how to extend the optimal tax formula to the case $\Gamma > 0$. One simply needs to replace s_{b0} by: $s_{b0} = \dfrac{E(V_{ci}\theta_i V_i^{-\Gamma} s_{bi} \mid z_i = 0)}{E(V_{ci}\theta_i V_i^{-\Gamma} \mid z_i = 0)}$. I.e. the formula for s_{b0} needs to be reweighted in order to take into account the lower marginal social utility $V_i^{-\Gamma}$ of zero-receivers with high utility V_i (i.e. zero-receivers with high productivity θ_i).

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

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

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

4.2 Numerical calibrations

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

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

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

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

That is, with high bequest flow $b_y = 20\%$, zero receivers want to tax inherited wealth at a higher rate than labor income (73% vs. 22%); with low bequest flow $b_y = 5\%$, they want the opposite (18% vs. 42%). The intuition is the following. In societies with low b_y (typically because of high g), there is not much tax revenue to gain from taxing bequests. So even zero-receivers do not like bequest taxes too much: it hurts their children without bringing much benefit in exchange. High growth societies care about the future, not about the past. Conversely, in societies with high b_y (typically because of low g), it is worth taxing bequests, so as to reduce labor taxation and improve the welfare of those receiving no inheritance.

In our theory there is really no general reason why capitalized inheritance would be taxed more or less than labor income. Any situation can be optimal, depending on parameters. With the low b_y ratios observed in the 1950s-1960s, it is probably optimal to tax inheritance less than labor. But with the high b_y ratios observed in the 1900s-1910s or the 2000s-2010s, it is probably optimal to tax inheritance more than labor (see Figures 4-5).

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

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

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

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

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

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

This is probably the most important lesson of this paper: once one allows the elasticity of capital supply to be a free parameter and to take moderate (non-infinite) values, then one can naturally obtain fairly large levels for socially optimal capital tax rates. That is, if we take $b_y = 15\%$ (current French level), then we find that as long as the elasticity e_B is less than one the optimal inheritance tax rate is higher than the optimal labor tax rate. With a realistic value $e_B = 0.2$, we find $\tau_B = 56\%$ *and* $\tau_L = 31\%$.[37] In practice, this bequest elasticity effect e_B is also mitigated by the existence of a positive labor supply elasticity effect e_L , which makes low labor taxation and therefore high bequest taxation even more valuable (see section 6).

Finally, a higher bequest taste s_{b0} implies a lower τ_B. The key trade-off captured by our theory is that everybody is both a receiver and a giver of bequest (at least potentially). This is why zero receivers generally do not want to tax bequests at 100%. Of course if $s_{b0} = 0$ (zero receivers have no taste at all for leaving bequests), then we obtain $\tau_B = 1/(1 + e_B)$ as a special case: we are back to the classical revenue maximizing rule, and $\tau_B \to 100\%$ as $e_B \to 0$. But as long as $s_{b0} > 0$, we have interior solutions for τ_B, even if $e_B = 0$.

In fact, for very high values of s_{b0}, and very low values of b_y, one can even get a negative τ_B, i.e. a bequest subsidy. Intuitively, if b_y is sufficiently small (e.g. if g is sufficiently large), then the benefits of taxing bequests - in terms of tax revenue - become smaller than the utility costs (as measured by s_{b0}), so that even those who receive no bequest do not want to tax bequests. For plausible parameter values, however, the optimal bequest tax rate τ_B from the viewpoint of zero receivers is positive (we discuss bequest subsidies in detail in Appendix A2).

4.3 Alternative Social Welfare Weights

The main limitation of Proposition 2 is that it puts all the weight on the individuals who receive exactly zero bequest (possibly a very small group, depending upon the distributions of shocks). However because real world inheritance is highly concentrated (half of the population receives

[37] We leave a proper estimation of e_B to future research. Preliminary computations using time and cross section variations in French inheritance tax rates (e.g. in the French system childless individuals pay a lot more bequest taxes than individuals with children) suggest that e_B is relatively small (at most $e_B = 0.1 - 0.2$). Using U.S. time and cross-section variations, Kopczuk and Slemrod (2001) also find elasticities e_B around $0.1 - 0.2$.

negligible bequests), our optimal tax results are actually very robust to reasonable changes in the social welfare objective. We show this in two steps. First, the above formula can be extended to compute the optimal tax rate from the viewpoint of those individuals belonging to the percentile p_z of the distribution of inheritance:

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

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

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

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

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

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

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

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

Unsurprisingly, the optimal tax rate τ_B is a decreasing function of z. I.e. individuals who receive higher inheritance prefer lower bequest taxes. People above percentile p_{z^*} (i.e. above normalized inheritance z^*) do not want any bequest tax at all. If one cares mostly about the welfare of high receivers, then obviously one would not tax inheritance. Conversely, for individuals with very low z, the formula delivers optimal tax rates that are very close to the meritocratic Rawlsian optimum. Interestingly, $z^* < 1$, i.e. agents with average bequest prefer bequest subsidies (if $z = 1$, then $\tau_B < 0$).[38] The intuition is the following. In terms of after-tax total resources, agents receiving average bequest have nothing to gain by (linearly) taxing successors from their own cohort. So since taxing bequests reduces the utility from leaving wealth to the next generation, there is really no point having a positive τ_B.

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

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

One simple–yet plausible–way to calibrate the formula is the following. Assume that we are trying to maximize the average welfare of bottom 50% bequest receivers ($p_z \leq 0.5$). In every country for which we have data, the bottom 50% share in aggregate inherited wealth is typically about 5% or less (see Piketty, 2011, p.1076), which means that their average z is about 10%. The average labor productivity θ_z within this group is below 100% (bottom 50% inheritors also earn less than average), but generally not that much below, say at least 50% (which would imply that they are all fairly close to the minimum wage, i.e. that they almost perfectly coincide with the bottom 50% labor earners) and more realistically around 70%. As one can see, given that

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

z/θ_z is very small anyway, this θ_z effect has a limited impact on optimal tax rates. I.e. in the benchmark case with $b_y = 15\%$, $e_B = 0.2$, $z = 10\%$, the optimal bequest tax rate is equal to $\tau_B = 49\%$ with $\theta_z = 70\%$, vs. $\tau_B = 46\%$ with $\theta_z = 50\%$, *(vs.* $\tau_B = 56\%$ if $z = 0\%$). That is, inheritance is so concentrated that bottom 50% bequest receivers and zero bequest receivers have welfare maximizing bequest tax rates which are in any case relatively close.

Example 3. *Assume* $\tau = 30\%, \alpha = 30\%, b_y = 15\%, e_B = 0.2, s_{bz} = 10\%$.
If $z = 0\%$, *then* $\tau_B = 56\%$ *and* $\tau_L = 31\%$.
If $z = 10\%$ *and* $\theta_z = 70\%$, *then* $\tau_B = 49\%$ *and* $\tau_L = 32\%$.
If $z = 10\%$ *and* $\theta_z = 50\%$, *then* $\tau_B = 46\%$ *and* $\tau_L = 33\%$.

Our optimal tax formulas show the importance of distributional parameters for the analysis of socially efficient capital taxation. They also illuminate the potentially crucial role of perceptions about distributions. If individuals have wrong perceptions about their position in the various distributions, this can have large impacts on their most preferred tax rate. E.g. with full information all individuals with inheritance percentile below p_{z^*} would prefer a positive bequest tax. In actual fact, the distribution is so skewed that less than 20% of the population has inherited wealth above average (i.e. the true p_{z^*} is typically above 0.8).[39] But to the extent that many more people believe to be above average, either in terms of received or left bequest, this might explain why (proportional) bequest taxes can have majorities against them.

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

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

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

with $x_z = E(z_{t+1i}|z_{ti} = z) =$ *average normalized bequest left by* p_z*-receivers*
$\nu_z = s_{bz}/s_z =$ *share of* p_z*-receivers wealth accumulation due to bequest motive*
$z =$ *normalized inheritance of* p_z*-receivers.*
(b) *If* $x_z \to 0$ *as* $z \to 0$, *then* $\tau_B \to 1/(1+e_B)$ *as* $z \to 0$ *(revenue maximizing tax rate)*

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

Proof. One simply needs to substitute $(1-\alpha-\tau)s_{bz}/b_y$ by $e^{-(r-g)H}\nu_z x_z/\theta_z - s_{bz}[\tau_B + (1-\tau_B)z/\theta_z]$ in the original formula. See Appendix A3. **QED.**

By construction, both formulas are equivalent. Whether one should use one or the other depends on which empirical parameters are available. The original formula uses the aggregate inheritance flow b_y (a parameter that is relatively easy to estimate, since it relies mostly on aggregate data) and the bequest taste s_{bz} (a preference parameter that is relatively difficult to estimate).[40] The alternative formula is based almost entirely on distributional parameters which in principle can be estimated empirically - but require comprehensive microeconomic data (such as wealth data spanning over two generations).[41] Its main advantage is that it illuminates the key role played by distribution for optimal capital taxation.

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

We leave a proper empirical calibration of our optimal tax formula to future research. Here we simply illustrate the crucial role played by the distribution of x_z. If $x_z = 10\%$, i.e. if the children of bottom 50% successors receive as little as what their parents received (relative to the average), then the optimal bequest tax rate is $\tau_B = 77\%$ for an elasticity $e_B = 0.2$ (it would be 95% with a zero elasticity). But if $x_z = 100\%$, i.e. if on average they receive as much as other children, then the optimal bequest tax rate is only $\tau_B = 45\%$. Presumably the real world is in between, say around $x_z = 50\%$, in which case $\tau_B = 61\%$.

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

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

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

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

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

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

Note that our framework implicitly double counts welfare arising from bequest planning as bequests enter the utility of donors and enter the budget constraint of donees. As discussed in the literature (e.g., Cremer and Pestieau, 2004, Diamond, 2006 and Kaplow, 2001 and 2008), double counting raises issues as it can generate "free utility" devices by subsidizing giving and taxing back proceeds. This issue arises in our setting when social welfare weights are heavily tilted toward high $z\%$ receivers. Indeed, if $z \geq \theta_z$, then τ_B is no longer well defined as the government would want an infinite subsidy to bequest: it is always desirable for very high bequest receivers to decrease τ_B and increase τ_L.

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

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

4.4 Nonlinear Bequest Taxes

Our basic optimal tax formula can also be extended to deal with non-linear bequest taxes. We now assume that the tax rate τ_B applies only above an exemption $b_t^* > 0$. Most estate or inheritance tax systems adopt such exemptions. The exemption is sometimes very high relative to average in countries such as the United States where less than 1% of estates are taxable, or lower as in France where a significant fraction of estates are taxable (typically 10%-20%).[42] Naturally $b_t^* = b^* e^{gHt}$ grows at rate g to ensure a steady state equilibrium. Denoting by B_t^* aggregate taxable bequests (i.e., the sum of $b_t - b_t^*$ across all bequests above b_t^*), the government budget constraint becomes

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

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

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

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

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

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

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

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

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

with $s_{b0}^* = E[(s_{bi}/s_i)(b_{t+1i} - b_{t+1}^*)^+ | z_i = 0]/E(\tilde{y}_{ti}|z_i = 0] =$ strength and likelihood that non-receivers will leave taxable bequests (weighted by marginal utility$\times$labor income).

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

Four points are worth noting. First, if zero-receivers never accumulate a bequest large enough to be taxable, then $s_{b0}^* = 0$, and the formula reverts to the revenue maximizing tax rate

[42] In any case the fraction of the population paying bequest taxes is generally much less than 50% - a fact that must naturally be related to the high concentration of inherited wealth: bottom 50% successors always receive barely 5% of aggregate inheritance (while the top 10% receives over 60% in Europe and over 70% in the U.S.), so there is little point taxing them. See e.g. Piketty (2011, p.1076).

$\tau_B = 1/(1+e^*_B) = 1/(1+a \cdot \bar{e}_B)$.[43] More generally, if zero-receivers have a very small probability to leave a taxable bequest (say, if b^* is sufficiently large), then s^*_{b0} is close to 0, and τ_B is close the revenue maximizing tax rate. This can be easily generalized to small p_z-receivers (say, bottom 50% receivers). If the elasticity is moderate (say, $e^*_B = 0.2$), then this implies the socially optimal inheritance tax rate on large bequests will be extremely high (say, $\tau_B = 70\% - 80\%$).

This model can help explain why very large top inheritance tax rates were applied in countries like the U.S. and the U.K. between the 1930s and the 1980s (typically around 70%-80%; see Figure 1 above). In particular, the fact that the rise of top inheritance tax rates was less dramatic in Continental Europe (French and German top rates generally did not exceed 30%-40%) seems qualitatively consistent with the fact these countries probably suffered a larger loss in aggregate inheritance flow ratios b_y and b^*_y following the world wars capital shocks.[44]

Second, as b^* grows, there are two options: either s^*_{b0}/b^*_y converges to zero or converges to a positive level. The first case corresponds to an aristocratic society where top bequests always come from past inheritances and never solely from self-made wealth. In that case again, the optimum τ_B would be the revenue maximizing rate. The second case corresponds to a partly meritocratic society where some of the top fortunes are self-made. In that case, even for very large b^*, non-receivers want a tax rate on bequests strictly lower than the revenue maximizing rate. In reality, it is probable that s^*_{b0}/b^*_y declines with b^* as the fraction of self-made wealth likely declines with the size of wealth accumulated. If the elasticity $\bar{e}_B$ and a are constant, then this suggests that the optimum τ_B increases with b^*. The countervailing force is that aristocratic wealth is more elastic as the bequest tax hits those fortunes several times across several generations, implying that $\bar{e}_B$ might actually grow with b^*.[45]

Third, one can also ask the question of what is the optimal b^* from the point of view of zero-receivers. Solving for the optimal b^* is difficult mathematically. If the optimal τ_B is zero when $b^* = 0$ (because zero-receivers care a lot of leaving bequests), then it is likely that τ_B will become positive when b^* grows (if society is relatively aristocratic). Then a combination $\tau_B > 0$ and $b^* > 0$ will be better that $\tau_B = 0$ and $b^* = 0$. The trade-off is the following: increasing b^*

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

[44] The German top rate reached 60% in 1946-1948 when it was set by the Allied Control Council, and was soon reduced to 38% in 1949 when the Federal Republic of Germany regained sovereignty over its tax policy. One often stated argument in the German public debate was the need to favor reconstruction and new capital accumulation. See e.g. Beckert (2008). In contrast, according to the "war mobilization" theory (see Scheve and Stasavadge 2011), inheritance taxes would have increased at least as much in Germany and France as in the UK and the US.

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

reduces the tax base b_y^* and hence estate tax revenue so this is a negative. The positive is that it reduces s_{b0}^* (probably at a faster rate than b_y^*, allowing for a greater optimal τ_B.

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

5 Inheritance Taxation vs. Lifetime Capital Taxation

So far we have focused upon optimal taxation of capitalized inheritance and derived optimal tax formulas that can justify relatively large tax rates when the aggregate inheritance flow is large. With inheritance flows b_y around 10%-15% of national income (as observed in today's developed economies, with a gradual upward trend), our formulas suggest that socially optimal tax rates τ_B in our model would be around 40%-60%, or even higher, thereby raising as much as 5%-8% of national income in annual tax revenues. As mentioned in introduction, actual tax revenues from capital taxes are even slightly higher, around 8-9% in the European Union and the United States. However only a small part comes from inheritance taxes–generally less than 1% of GDP as bequest tax rates are usually relatively small, except for very top (taxable) estates. Most revenue comes from "lifetime capital taxes", falling either on the capital stock (annual property and wealth taxes, typically about 1-2% of GDP) or on the capital income flow (annual taxes on corporate profits, rental income, interest, dividend and capital gains, typically about 4%-5% of GDP).[46] Why do we observe small inheritance taxes and large lifetime capital taxes? Our basic model cannot tackle this question, since all forms of capital taxes are equivalent (Section 3).

Clearly the conclusion would be different in a full-fledged, multi-period model with life-cycle savings.[47] Positive capital income taxes $\tau_K > 0$ would then impose additional distortions on inter-temporal consumption decisions within a given lifetime. Following the Atkinson-Stiglitz logic, it would generally be preferable to have $\tau_K = 0$ and to raise 100% of the capital tax

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

[47] See Section 6 below for such an extension.

revenue via a bequest tax $\tau_B > 0$. Naturally, if the inter-temporal elasticity of substitution is fairly small, then this extra distortion would also be small, and both tax policies would be relatively close to one another. In the real world however we do observe a collective preference in favor of lifetime capital taxes (either stock-based or flow-based) over one-off bequest taxes, so there must be some substantial reasons for this fact. What can account for this?

In this section, we explore two mechanisms explaining why lifetime capital taxes are more heavily used than one-off inheritance taxes: the existence of a fuzzy frontier between capital income and labor income flows; and the existence of uninsurable idiosyncratic shocks to rates of return. Each mechanism allows us to explore different aspects of the optimal capital tax mix. We certainly do not pretend that these are the only important factors. For example, individuals may be subject to various forms of tax illusion whereby smaller annual capital taxes are less visible than one big bequest tax per generation.[48] Other forms of capital market imperfections, such as borrowing constraints, might also play an important role. For example, large inheritances taxes may force successors to quickly and inefficient sell their property.[49]

5.1 Fuzzy Frontier Between Capital and Labor Income Flows

The simplest rationale for taxing capital income is the existence of a fuzzy frontier between capital and labor income flows. Any gap between the labor income tax rate τ_L and the capital income tax rate τ_K may induce tax avoidance. E.g., self-employed individuals can largely decide which part of their total compensation takes the form of wage income, and which part takes the form of dividends or capital gains. Opportunities for income shifting also exist for a large number of top executives (e.g. via stock options and capital gains). There is extensive empirical evidence that income shifting is a significant issue, and accounts for a large fraction of observed behavioral responses to tax changes.[50] At some level, this fuzzy-frontier problem can be viewed as the consequence of capital markets imperfections. With first-best markets, full financial intermediation and complete separation of ownership and control, distinguishing the returns to capital services from the returns to labor services would be easily feasible.

For simplicity, we assume "full fuziness". Individuals can shift their labor income flows into

[48] This could contribute to explain why most individuals seem to prefer to pay an annual property tax equal to 1% of their property value (or 25% of their 4% annual return) during 30 years rather than to pay 30% of the property value all at once at the time they inherit the asset.

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

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

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

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

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

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

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

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

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

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

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

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

For large bequest flows $b_y \simeq 10 - 20\%$, a comprehensive income tax system only reduces slightly the need for inheritance taxation. In contrast, for bequest flows $b_y \simeq 5\%$, the reduction can be very large. This might explain the large number of exemptions for capital income that were created during the reconstruction period, particularly in countries like France or Germany.

In practice, only a fraction of the population can easily shift capital into labor income (and conversely). This has to be weighted against costs of capital taxation in a model with life-cycle

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

savings. Therefore the resulting optimal tax gap $\Delta\tau = \tau_L - \tau_K \geq 0$ would depend negatively on the fraction of income shifters and positively on the intertemporal elasticity of substitution.[52]

5.2 Uninsurable Idiosyncratic Shocks to Rates of Return

Let us assume away the fuzzy-frontier problem and consider the implications of uninsurable idiosyncratic shocks to rates of return for the optimal tax mix. The basic intuition is straightforward. From a welfare viewpoint, as well as from an optimal tax viewpoint, what matters is capitalized bequest $\tilde{b}_{ti} = b_{ti} e^{r_{ti}H}$, not raw bequest b_{ti}. The problem of a bequest tax is that it depends only on b_{ti}, not on the idiosyncratic variations in $e^{r_{ti}H}$. So it makes more sense to charge part of the tax burden via bequest taxation $\widetilde{\tau}_B$, and part of the tax burden via lifetime capital taxation τ_K–possibly a much larger part–in case the uncertainty about future returns is very large. In practice there is also a difference in timing. At the time of setting the bequest tax rate τ_B, the future rate of return $e^{r_{ti}H}$ on a given asset over one generation is unknown. Rates of return are notoriously difficult to predict, and they vary enormously over assets and across individuals. In that context, it is preferable to impose a moderate bequest tax at time of receipt combined with an annual capital income tax on the returns.[53]

Formally, let us assume that individual life-time rates of returns $R_{ti} = e^{r_{ti}H} - 1$ vary across individuals. Let us denote by R the aggregate rate of return across all individuals. We assume that shocks R_{ti} are idiosyncratic so that there is no risk in aggregate.

If R_{ti} is exogenous to the behavior of individuals, then it is clearly optimal for the government to set $\tau_K = 100\%$ to insure individuals against risky returns. In effect, the government is replacing risky individual returns R_{ti} by the aggregate return R , thereby providing social insurance. Standard financial models assume that individuals can insure themselves by diversifying their portfolios but in practice self-insurance is far from complete, implying that taxes have a role to play in order to reduce uncertainty.[54]

[52]Alternatively if one assumes a finite elasticity of income shifting with respect to the gap in tax rates, then the optimal tax gap will depend negatively on this elasticity (see Piketty, Saez, Stantcheva (2011)). Here we implicitly assumed an infinite elasticity, so that tax rates have to be exactly equal. Note also that the administrative capability to distinguish between capital and labor income flows and to impose separate tax rates is to some extent endogenous. E.g. it is easier if for the tax administration to observe or estimate capital income if taxpayers file annual wealth declarations in addition to annual income declarations.

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

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

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

In the appendix we also provide examples with numerical values. These simulations rely on simplifying assumptions, and are only illustrative and exploratory. In particular, we know very little about the elasticity e_R of the aggregate rate of return R. Available macroeconomic evidence shows that aggregate rates of return, factor shares and wealth-income ratios are relatively stable over time and across countries, which–given large variations in taxes–would tend to suggest relatively low elasticities e_R (perhaps around $0.1-0.2$).[55] This would seem to imply that the optimal capital income tax rate is much larger than the optimal labor income tax rate. E.g. if $e_R = 0.1$ then in our simulations we obtain $\tau_K = 78\%$ and $\tau_L = 35\%$. However the simulations also show that the results are very sensitive to the exact value of e_R . E.g. if $e_R = 0.5$ then capital income would be taxed much less than labor income: $\tau_K = 17\%$, and $\tau_L = 37\%$. This is because in the model a lower return R not only reduces the capital income tax base but also has a negative impact on the aggregate steady-state bequest flow b_y.[56]

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

[55]Conceivably, higher individual effort e_{ti} translates into higher individual return R_{ti} mostly at the expense of others (e.g., traders obtaining advance information about when to sell a given financial asset), i.e. the aggregate R is very little affected. In the extreme case where this is a pure zero-sum game (R fixed), then the relevant elasticity is $e_R = 0$, and the optimal tax rate is $\tau_K = 100\%$. For an optimal tax model based upon pure rent-seeking elasticities, see Piketty, Saez and Stantcheva (2011).

[56]In addition, these simulations do not take into account the distortionary impact of τ_K on inter-temporal consumption allocation along the life-cycle.

More generally, $\tau_K > \tau_L$ was actually the norm in most income tax systems when the latter were instituted in the early 20th century (generally around 1910-1920). At that time income tax systems typically involved a progressive surtax on all forms of labor and capital income (including imputed rent), and a set of schedular taxes taxing wage income less heavily than capital income. It has now become more common to have $\tau_K < \tau_L$, via special tax exemptions for various categories of capital income. But we feel that this mostly reflects a rising concern for international tax competition and tax evasion and the persistent lack of tax coordination,[57] rather than considerations about the global welfare optimum.

6 Extensions

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

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

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

This formula is similar to the inelastic case except that e_L appears both in the numerator and denominator. τ_B increases with e_L if $\tau(1 + e_B) + s_{b0}(1 - \alpha) \geq b_y$ which is satisfied empirically. Hence, a higher e_L implies a higher τ_B and a lower τ_L. Intuitively, a higher labor supply elasticity makes high labor taxation less desirable and tilts the optimal tax mix tilt more towards bequest taxes. Numerical examples presented in appendix show that, for realistic parameters, very large bequest elasticities and very small labor supply elasticities are needed to obtain $\tau_B < \tau_L$.

Closed Economy. Our optimal tax results can easily be extended to the closed economy case where the capital stock K_t is equal to domestic inheritance (i.e. $K_t = B_t$). The factor prices (wage rate and rate of return) are now endogenous and given by the marginal product of labor

[57] The view is that it is easier to reallocate one's financial portfolio abroad than one's labor income, and that it is harder to apply the residence principle of taxation for capital income; or at least this is a view that became very influential in a number of small open economies, typically in Nordic countries.

and capital. As in standard optimal tax theory (Diamond and Mirrlees, 1971), optimal tax formulas are independent of the production side and hence remain the same with endogenous factor prices. The important point is that the elasticity e_B (and e_L with elastic labor supply) entering the formula is the pure supply elasticity, i.e. keeping factor prices constant.

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

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

More precisely, we prove the following. In the closed economy case, the government will accumulate sufficient assets or liabilities to ensure that the Modified Golden Rule holds whereby $r = r^* = \delta + \Gamma g$ with δ = social rate of time preference and Γ = concavity of social welfare function.[59] The government will then apply the same optimal bequest and labor tax rates as in the case with a period-by-period budget constraint with two minor modifications (appendix C, proposition C3). First, s_{b0} is replaced by $s_{b0}e^{\delta' H}$ in the optimal τ_B formula with $\delta' = \delta+(\Gamma-1)g$. This correction appears because τ_{Bt} hurts bequests leavers from generation $t-1$ while revenue accrues in generation t. Note that with no social discounting $\delta = 0$ and log-utility $\Gamma = 1$, there is no correction. Second, the formula for τ_L has to be adjusted for the interest receipt or payment term if the government has assets or debts at the optimum.

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

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

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

models the impact of a consumption tax on private utility and government finances.

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

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

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

If we further assume perfect correlation of productivity shocks across generations ($\theta_{it} = \theta_{i0}$ for all t), we lose our key ergodicity assumption. In the long run, the distribution of inheritance $\phi(z)$ would then be perfectly correlated with the distribution of labor productivity $h(\theta)$. Hence, the labor income tax τ_L and the bequest tax τ_B would have the same distributional impact. Since the latter imposes an extra utility cost - via the usual joy-of-giving externality -, there is no point having a positive τ_B.[61] But as long as inequality is two-dimensional there is room for a two-dimensional tax policy tool.

Overlapping Generations and Life-cycle Savings. Our results and optimal tax formulas can also be extended to a full-fledged continuous time model with overlapping generations and life-cycle savings. We keep the same closed-form formulas for optimal inheritance tax rates.

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

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

Regarding optimal lifetime capital taxation, we keep the same general, qualitative intuitions, but numerical methods are needed to compute the full optimum.

In that model, b_{yt} is now defined as the cross-sectional, macroeconomic ratio between the aggregate inheritance flow B_t transmitted at a given time t and domestic output Y_t produced at this same time t (as plotted on Figures 4-5). If inheritances are received around mid-life (relative to earnings), then the cross-sectional macroeconomic ratio is close to the share of capitalized inheritance in total lifetime resources of the cohort inheriting at time t of our basic model (there is small correction factor in the b_y formula, see Appendix B). In any case, the optimal tax formulas of Proposition 2 continue to apply in this model.

For optimal lifetime capital taxation, life-cycle savings now generate an extra distortion. That is, positive tax rates on capital income $\tau_K > 0$ distort the intertemporal allocation of consumption within a lifetime. The magnitude of the associated welfare cost depends on the intertemporal elasticity of substitution $\sigma = 1/\gamma$ (which might well vary across individuals). As long σ is relatively small, the impact on our optimal capital tax results is moderate.

7 Conclusion

This paper has developed a tractable theory of optimal capital taxation. The results coming out of our model challenge the conventional zero capital tax results, which in our view rely on ad hoc modeling assumptions which are often left implicit. If one assumes from the beginning that there is little or egalitarian inheritance, then it is perhaps not too surprising if one concludes that inheritance taxation is a secondary issue. If one assumes from the beginning that the long run elasticity of saving and capital supply is infinite, then it is maybe not too surprising if one concludes that the optimal capital income tax is zero in the long run. Our model relaxes these assumptions, and shows that the optimal tax mix between labor and capital depends on the various elasticities at play and on critical distributional parameters. We hope our results will contribute to the emergence of more pragmatic debates about capital taxation, based more upon relevant empirical parameters than abstract theoretical results relying on strong assumptions.

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

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

References

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

Atkeson, Andrew and Robert E. Lucas 1992. "On Efficient Distribution with Private Information," *Review of Economic Studies* 59, 427-453.

Atkinson, Anthony 1980. "Inheritance and the Redistribution of Wealth", in G.M. Heal and G.A. Hughes, eds., *Public Policy and the Tax System*, 36-66, London: Allen and Unwin.

Atkinson, Anthony, Thomas Piketty and Emmanuel Saez 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49(1): 3–71.

Atkinson, Anthony, and Agnar Sandmo 1980. "Welfare Implications of the Taxation of Savings", *The Economic Journal*, 90, 529-549.

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

Atkinson, A.B., and Joseph Stiglitz 1980. *Lectures on Public Economics*, McGraw Hill.

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

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

Cai, Hongbin and Daniel Treisman 2005. "Does Competition for Capital Discipline Governments? Decentralization, Globalization, and Public Policy." *American Economic Review* 95(3), 817-830.

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

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

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger 2009. "Taxing Capital? Not a Bad Idea after All!." *American Economic Review*, 99(1), 25–48.

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

Cremer, Helmuth, Pierre Pestieau, and Jean-Charles Rochet 2003. "Capital Income Taxation when Inherited Wealth is not Observable", *Journal of Public Economics* 87, 2475–90.

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

Diamond, Peter, and James A. Mirrlees 1971. "Optimal Taxation and Public Production: Production Efficiency." *American Economic Review,* 61(1), 8–27.

Diamond, Peter, and Emmanuel Saez 2011. "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25(4), 165-190.

Edwards, Chris and Daniel J. Mitchell 2008. *Global Tax Revolution: The Rise of Tax Competition and the Battle to Defend It* , Cato Institute, 250p.

European Commission 2011. *Taxation trends in the European Union,* 2011 Edition, Eurostat.

Farhi, Emmanuel, Christopher Sleet, Iván Werning and Sevin Yeltekin 2011. "Non-linear Capital Taxation Without Commitment", Harvard Working Paper.

Farhi, Emmanuel and Iván Werning 2007. "Inequality and Social Discounting", *Journal of Political Economy*, 115(3), 365-402.

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

Farhi, Emmanuel and Iván Werning 2011. "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," MIT Working Paper.

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

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

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski 2011. "Optimal Dynamic Taxes." Princeton Working Paper.

Golosov, Mikhail, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl 2011. "Preference Heterogeneity and Optimal Capital Income Taxation", Yale Working Paper.

Golosov, Michael, Tsyvinski, Aleh, and Ivan Werning 2006. "New Dynamic Public Finance: A User's Guide," NBER Macroeconomics Annual.

Gordon, Roger H. 1985. "Taxation of Corporate Capital Income: Tax Revenues Versus Tax Distortions", *Quarterly Journal of Economics*, 100, 1-27.

Graetz, Michael J. and Ian Shapiro 2005. *Death by a Thousand Cuts: The Fight over Taxing Inherited Wealth*, Princeton, NJ: Princeton University Press.

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

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

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

Kaplow, Louis 2001. "A Framework for Assessing Estate and Gift Taxation", in Gale, William G., James R. Hines Jr., and Joel Slemrod (eds.) *Rethinking estate and gift taxation* Washington, D.C.: Brookings Institution Press.

Kaplow, L. 2008. *The Theory of Taxation and Public Economics,* Princeton University Press.

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

Kocherlakota, Narayana 2010. *The New Dynamic Public Finance* (Princeton: Princeton University Press).

Kopczuk, Wojciech 2001. "Optimal Estate Taxation in the Steady-State", Working Paper, University of British Columbia.

Kopczuk, Wojciech 2009. "Economics of Estate Taxation: Review of Theory and Evidence", *Tax Law Review* 63(1), 139–157.

Kopczuk, Wojciech 2012. "Taxation of Transfers and Wealth", in preparation *Handbook of Public Economics*, Volume 5, Elsevier.

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

Kopczuk, Wojciech and Joel Slemrod 2001. "The Impact of the Estate Tax on Wealth Accumulation and Avoidance Behavior", in *Rethinking Estate and Gift Taxation*, W. Gale, J.R. Hines and J. Slemrod eds., The Brookings Institution, 299–349.

Lucas, Robert 1990. "Supply-Side Economics: An Analytical Review", Oxford Economic Papers, New Series, 42(2), 293–316.

OECD 2011. *Revenue Statistics, 1965-2010*, OECD, Paris, 2011 Edition.

Piketty, Thomas 2010. "On the Long-Run Evolution of Inheritance: France 1820-2050", Working Paper, Paris School of Economics, 424p.

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

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

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

Piketty, Thomas, and Emmanuel Saez 2012. "Optimal Labor Income Taxation", in preparation *Handbook of Public Economics*, Volume 5, Elsevier.

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva 2011. "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities", NBER Working Paper No. 17616.

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

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

Saez, Emmanuel 2001. "Using Elasticities to Derive Optimal Tax Rates", *Review of Economic Studies*, 68, 205-229.

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

Saez, Emmanuel, Joel Slemrod, and Seth Giertz 2012. "The Elasticity of Taxable Income with respect to Marginal Tax Rates: A Critical Review", *Journal of Economic Literature*, 50(1).

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

Stiglitz, Joseph E. 1985. "Inequality and Capital Taxation", Working Paper, Stanford.

Vickrey, William 1947. *Agenda for Progressive Taxation*, New York: Ronald, 496p.

Weinzierl, Matthew 2011. "The Surprising Power of Age-Dependent Taxes", *Review of Economic Studies* 78(4), 1490-1518.

Wilhelm, Mark O. 1996. "Bequest Behavior and the Effect of Heirs' Earnings: Testing the Altruistic Model of Bequests", *American Economic Review* 86(4), 874-892.

Figure 1: Top Inheritance Tax Rates 1900-2011

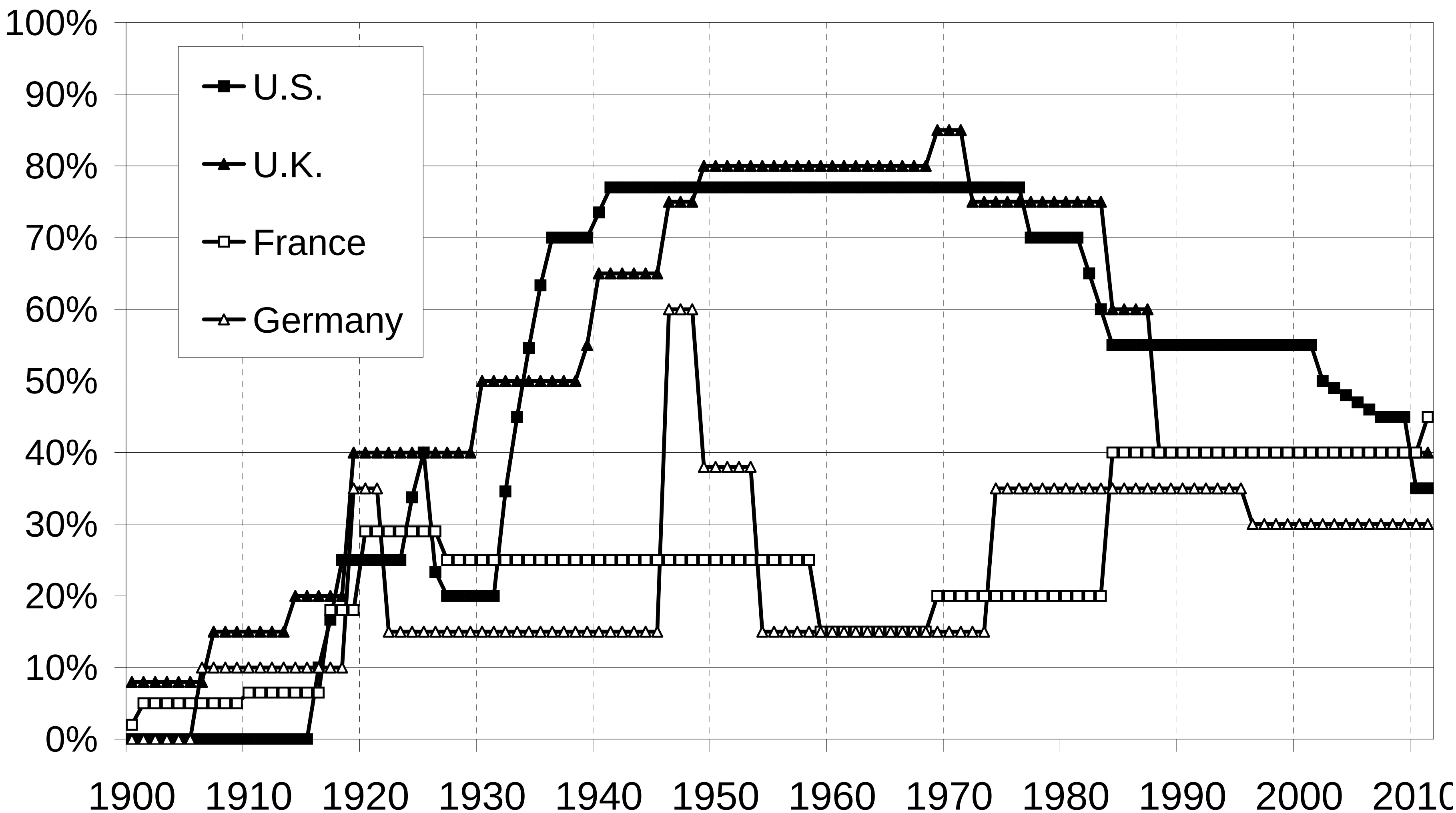


Figure 2: Top Income Tax Rates 1900-2011

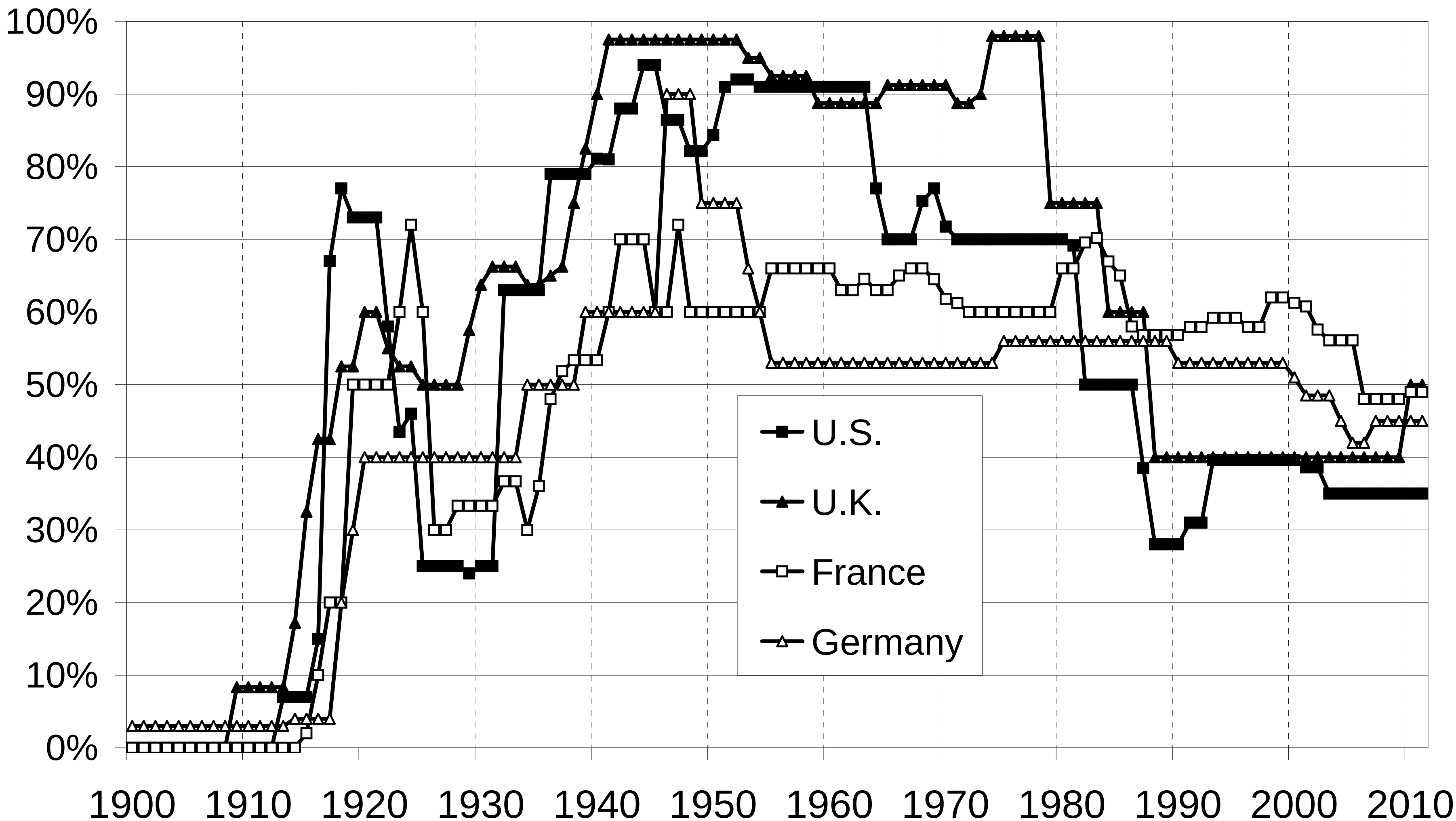


Figure 3: Top Income Tax Rates: Earned (Labor) vs Unearned (Capital)

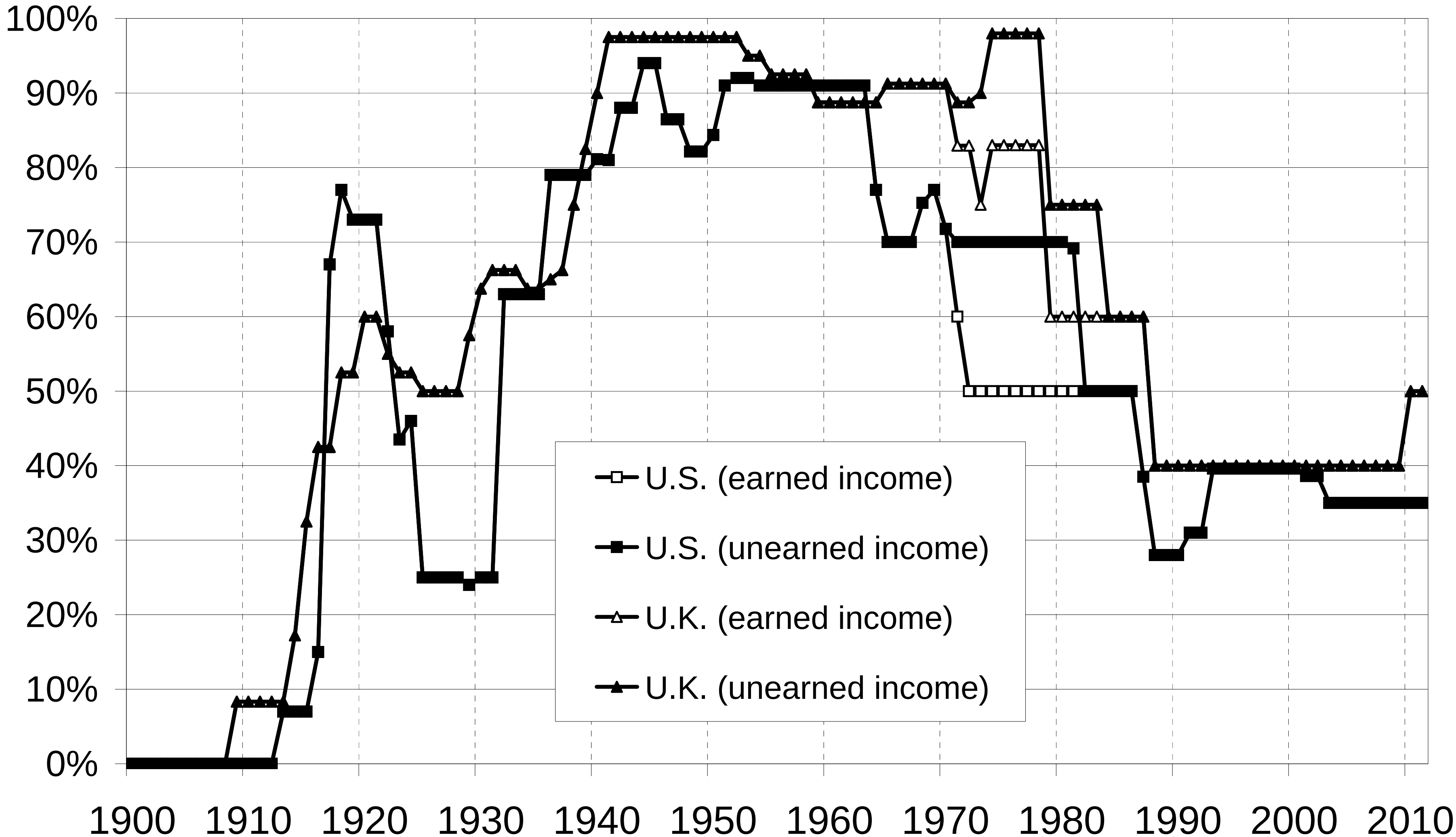
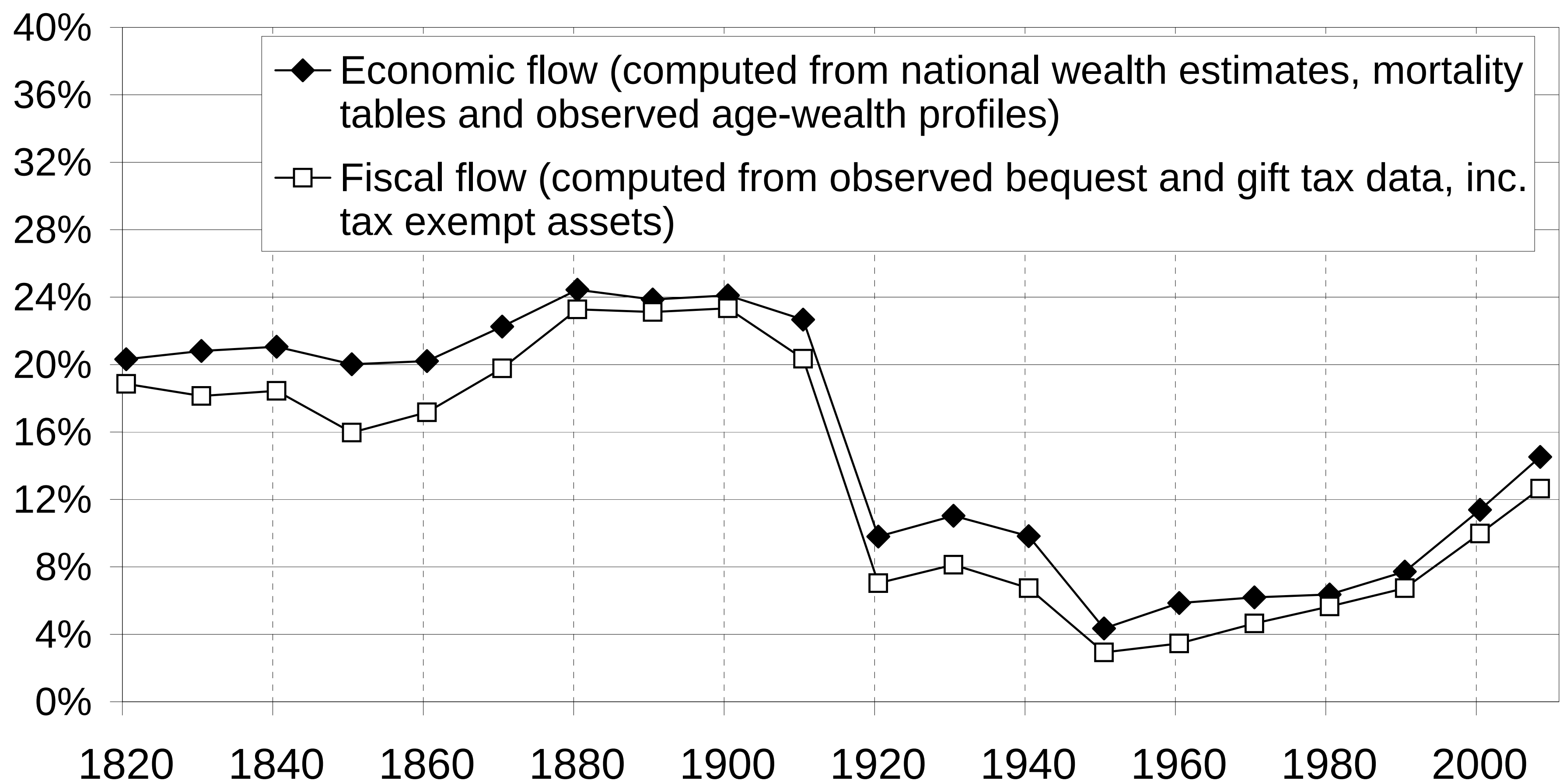
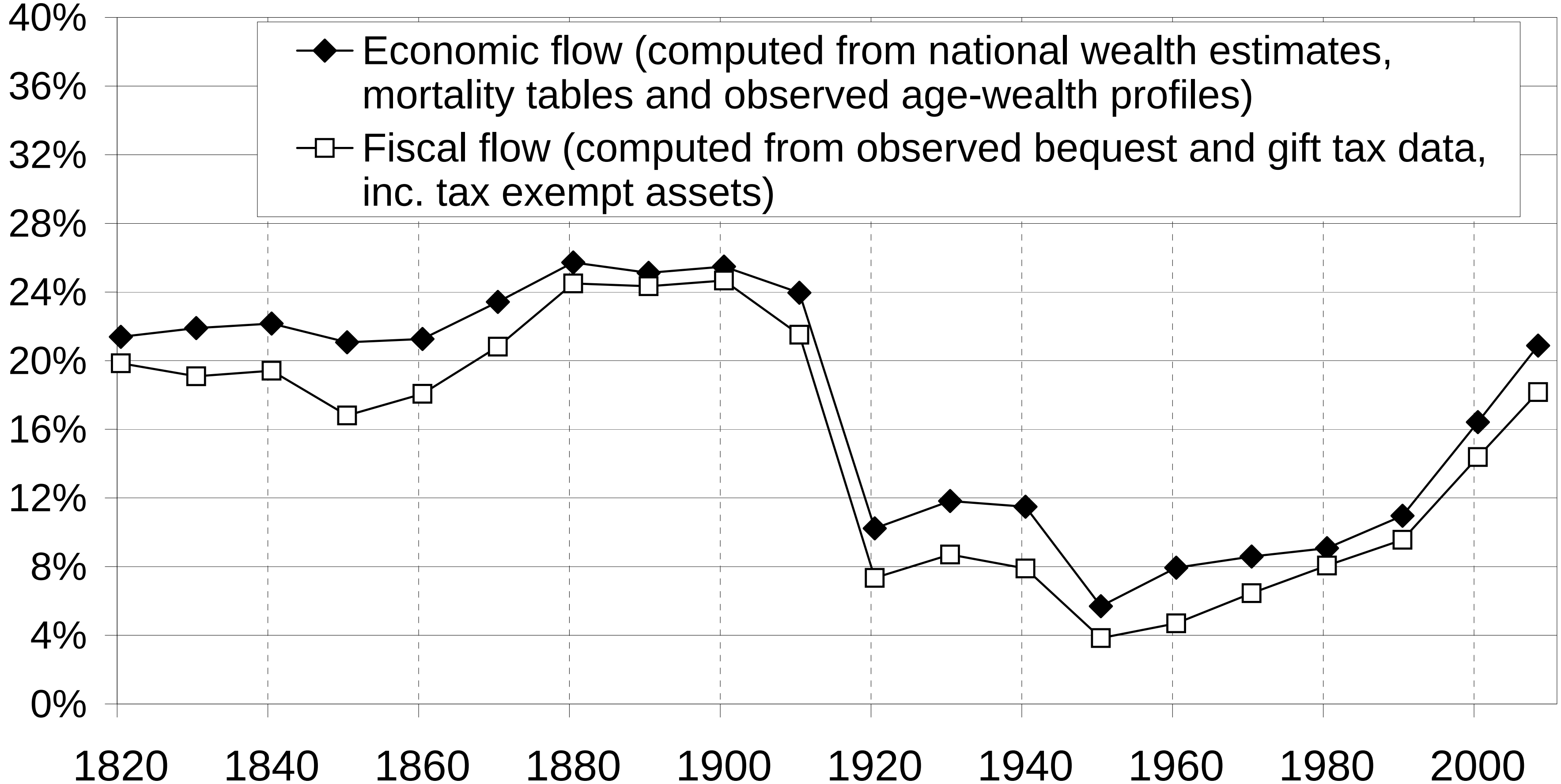


Figure 4: Annual inheritance flow as a fraction of national income, France 1820-2008



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

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



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