Journal of Public Economics 97 (2013) 61–74

Contents lists available at SciVerse ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Optimal progressive capital income taxes in the infinite horizon model☆

Emmanuel Saez

University of California, Department of Economics, 530 Evans Hall #3880, Berkeley, CA 94720, United States
NBER, United States

ARTICLE INFO

Article history:
Received 26 March 2011
Received in revised form 18 August 2012
Accepted 3 September 2012
Available online 18 September 2012

JEL classification:
H21
H62

Keywords:
Optimal capital taxation
Wealth distribution

ABSTRACT

This paper analyzes optimal progressive capital income taxation in an infinite horizon model where individuals differ only through their initial wealth. We consider progressive capital income tax schedules taking a simple two-bracket form with an exemption bracket at the bottom and a single marginal tax rate above a time varying exemption threshold. Individuals are taxed until their wealth is reduced down to the exemption threshold. The fraction of individuals subject to capital income taxation vanishes to zero in the long-run in analogy to the zero long-run capital tax result of Chamley and Judd with optimal linear taxes. However, in contrast to linear taxation, optimal nonlinear capital taxation can have a drastic impact on the long-run wealth distribution. When the intertemporal elasticity of substitution is not too large and the top tail of the initial wealth distribution is infinite and thick enough, the optimal exemption threshold converges to a finite limit. As a result, the optimal tax system drives all the large fortunes down a finite level and produces a truncated long-run wealth distribution.

1. Introduction

Most developed countries have adopted comprehensive individual income tax systems with graduated marginal tax rates in the course of their economic development process. The United States introduced the modern individual income tax in 1913, France in 1914, Japan in 1887, and the German states such as Prussia and Saxony, during the second half of the 19th century, the United Kingdom introduced a progressive super-tax on comprehensive individual income in 1909. Because of large exemption levels, these early income tax systems hit only the top of the income distribution. While tax rates were initially set at low levels, during the first half of the twentieth century, the degree of progressivity of the income tax was sharply increased and top marginal tax rates reached very high levels. In most cases, the very top rates applied only to an extremely small fraction of taxpayers. Therefore, the income tax was devised to have its strongest impact on the very top income earners. As documented by the top income studies surveyed by Atkinson et al. (2011), these top income earners derived the vast majority of their income in the form of capital income. Therefore, the very progressive schedules set in place during the inter-war period can be seen as a progressive capital income tax precisely designed to hit the largest wealth holders, and redistribute the fortunes accumulated during the industrial revolutions of the 19th century — a time with very modest taxation of capital income. Most countries have also introduced graduated forms of estate or inheritance taxation that further increase the degree of progressivity of taxation. Such a progressive income and estate tax structure should have a strong wealth equalizing effect.[1]

An important question in tax policy analysis is whether using capital income taxation to redistribute accumulated fortunes is desirable. As in most tax policy problems, there is a classical equity and efficiency trade-off: progressive capital income taxation can redistribute from the wealthy to the non-wealthy but might distort savings and consumption behavior and hence reduce wealth accumulation.[2] A number of studies on optimal dynamic taxation have suggested that capital taxation might have very large efficiency costs (see e.g., Lucas, 1990; Atkeson et al., 1999). As is well known, in the infinite horizon model, linear capital income taxes generate distortions increasing exponentially with

☆ Financial support from NSF grant SES-0134946 is gratefully acknowledged. This paper owes much to stimulating discussions with Thomas Piketty. I thank co-editor Pierre Pestieau, Alan Auerbach, V.V. Chari, Juan-Carlos Conesa, Narayana Kocherlakota, Antonio Rangel, numerous seminar participants, two anonymous referees, and especially Kenneth Judd for very helpful and detailed comments.
E-mail address: saez@econ.berkeley.edu.

[1] Indeed Piketty (2001b) and Piketty and Saez (2003) argue that the development of progressive taxation was one of the major causes of the decline of top capital incomes over the 20th century in France and in the United States.

[2] This trade-off that was at the center of the political debate on the introduction of progressive taxation in western countries. See Piketty (2001b) for a detailed account on the French case, and Brownlee (2000) for the United States.

0047-2727/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jpubeco.2012.09.002

time. The influential studies by Chamley (1986) and Judd (1985) show that, in the long-run, optimal linear capital income tax should be zero.[3]

This paper considers a simple departure from the standard infinite horizon model with no uncertainty of Chamley (1986) and Judd (1985). Instead of considering only time varying linear capital income taxation, we introduce a very simple form of nonlinear taxation. We consider progressive capital income tax schedules taking a two-bracket form with an exemption bracket at the bottom and a single marginal tax rate above a time varying exemption threshold. Such a tax system can be seen as a crude approximation of the actual early progressive income tax systems discussed above. To keep the model and the derivations tractable and transparent, we consider the simplest possible model with heterogeneity in initial wealth only, CRRA utility functions, exogenous and constant rate of return on capital equal to the individual discount rate, inelastic labor supply with uniform wages.[4] Importantly, we consider only a very limited set of tax policy tools that we believe captures an important and realistic trade-off. In the very simple model we use, we rule out complete redistribution of initial wealth which would naturally be first-best as initial wealth is exogenous in our model.[5] We obtain two main results.

First, our model retains the central long-run vanishing capital tax result as in Chamley–Judd. In our model, the fraction of individuals subject to capital income taxation vanishes to zero in the long-run so that capital income tax revenue does converge to zero in the long-run. Second, however, under a broad set of parametric assumptions, we find that progressive capital income taxation is much more effective than linear taxation to redistribute wealth. If the intertemporal elasticity of substitution (which measures the efficiency costs of capital income taxation) is small enough, even if the initial wealth distribution is unbounded, the optimal nonlinear capital income tax produces a wealth distribution that is truncated above in the long-run. Namely, no fortunes above a given threshold are left in the long-run. Therefore, large wealth owners continue to be taxed until their wealth level is reduced down to a given threshold.

The mechanism explaining why progressive taxation is desirable can be understood as follows. In the infinite horizon model, linear taxation of capital income is undesirable because it introduces a price distortion exponentially increasing with time. That is why optimal linear capital income taxation must be zero in the long-run. However, with a simple progressive tax structure with a single marginal tax rate above an exemption threshold, large wealth holders will be in the tax bracket and therefore will face a lower net-of-tax rate of return than modest wealth holders who are in the exempted bracket. As a result, the infinite horizon model predicts that large fortunes will decline until they reach the exemption level where taxation stops. Thus, this simple tax structure reduces all large fortunes down to the exemption level and thus effectively imposes a positive marginal tax rate only for a *finite* time period for any individual (namely until his wealth reaches the exemption threshold) and thus avoids the infinite distortion problem of the linear tax system with no exemption.[6] The second virtue of this progressive tax structure is that the time of taxation is increasing with the initial wealth level because it takes more time to reduce a large fortune down to the exemption threshold than a more modest one. This turns out to be desirable in general for the following reason. Large wealth holders consume mostly out of their initial wealth rather than their annual stream of labor income. Therefore, the positive human wealth effect created by capital taxation on initial consumption is small relative to the income effect for large wealth holders. As a result, capital taxation leads to a lower pace of wealth decumulation for the rich, and thus they can be taxed longer at a lower efficiency cost than the poor. It is important to recognize however, that the size of behavioral responses to capital income taxation, measured by the intertemporal elasticity of substitution, matters. When this elasticity is large, it is inefficient to tax any individual, however rich, for a very long time and thus, it is preferable to let the exemption level grow without bounds as time elapses producing an unbounded long-run wealth distribution.

Naturally, the parsimonious model developed here does not capture all the relevant issues arising with capital income taxation. The present model takes as given the initial unequal wealth distribution, and ignores completely the issue of creation of new wealth. This contribution can be seen as a theory of the taxation of rentiers where the central trade-off is the following: using capital income taxation is desirable to redistribute from the rich to the poor but capital income taxation induces individuals to over-consume initially and run down their wealth levels, hence reducing the capital income tax base down the road. This basic model therefore ignores completely the issue of creation of new fortunes. New fortunes are created in general by successful entrepreneurs or spells of very high labor income. Those fortunes can then be passed down to future generations through bequests. Taxation of capital income reduces the (long-term) benefits of creating a fortune, and may thus reduce entrepreneurial effort or labor supply as well.[7]

Conversely, in models with credit constraints, Aiyagari (1995) and Chamley (2001) have shown that capital income taxation may be desirable, even in the long run as capital income taxes can redistribute from the rich who are not credit constrained toward to poor who are credit constrained. Similarly, the recent and fast growing New Dynamic Public Finance literature (see Golosov et al., 2006; Kocherlakota, 2010 for valuable recent surveys) shows that dynamic labor productivity risk leads to non-zero capital income taxes. Therefore, it is not immediately clear in which direction would the introduction of entrepreneurs tilt the results presented here. We expect, however, that the economic forces regarding the taxation of rentiers described here would still be present in this more general model.

The paper is organized as follows. Section 2 presents the model and the government objective. Section 3 considers linear taxation and provides useful preliminary results on the desirability of taxing

[3] Another strand of the literature has used overlapping generations (OLG) models to study optimal capital income taxes. In general capital taxes are expected to be positive but quantitatively small in the long-run (see e.g., Feldstein, 1978; Atkinson and Sandmo, 1980; King, 1980). However, when non-linear labor income tax is allowed, under some conditions, optimal capital taxes should be zero (see Atkinson and Stiglitz, 1976; Ordover and Phelps, 1979). More importantly, in the OLG model, capital accumulation is due uniquely to life-cycle saving for retirement. This contrasts with the actual situation where an important share of wealth, especially for the rich, is due to bequests (Kotlikoff and Summers, 1981). The OLG model with no bequests therefore is not well suited to the analysis of the taxation of large fortunes. Cremer and Pestieau (2004) survey this large literature. I come back to this issue in conclusion.

[4] We discuss how some of those assumptions affect our results and consider various extensions in Section 5.

[5] This approach is not an exception in optimal tax theory. For example, the famous Ramsey model of commodity taxation (as well as the basic Chamley–Judd model with a representative agent) rules out lump sum taxation that would be first-best efficient. The recent New Dynamic Public Finance literature (see e.g., Kocherlakota, 2010) carefully grounds optimal dynamic taxation upon informational assumptions using the mechanism design approach. The drawback is that optimal tax structures are very complex and history dependent (see Diamond and Saez, 2011 for a discussion of the pros and cons of the mechanism design approach vs. the limited government tool set we adopt here).

[6] Piketty (2001a) (in the unpublished appendix of the working paper version) made the important and closely related point that, in the infinite horizon model, a constant capital income tax above a high threshold does not affect negatively the long-run capital stock in the economy because the reduction of large fortunes is compensated by an increase of smaller wealth holdings. This, of course, is not true with linear capital income taxation. We come back to this important point in Section 5 when we consider extensions.

[7] Cagetti and De Nardi (2006) propose a positive analysis of capital income taxation and the wealth distribution in a dynamic and stochastic model with entrepreneurs. They do not, however, tackle the normative issue of optimal capital income taxation. Piketty and Saez (2012) propose a theory of optimal capital taxation in a model with heterogeneous tastes for bequests and hence endogenous and heterogeneous inherited wealth.

richer individuals longer. Section 4 introduces progressive capital income taxation and derives the key theoretical results. Section 5 analyzes how relaxing the simplifying assumptions of the basic model affects the results. Section 6 offers some concluding comments.

2. The general model

2.1. Individual program

We consider a simple infinite horizon model with no uncertainty and perfectly competitive markets. All individuals have the same instantaneous CRRA utility functions with constant intertemporal elasticity of substitution σ: $u(c)=[c^{1-1/\sigma}-1]/[1-1/\sigma]$. The elasticity σ is the key parameter measuring the behavioral response to capital income taxation (see below). When $\sigma=1$, we have of course $u(c)=\log c$. As is well known, the CRRA assumption simplifies greatly the tractability of the model and allows us to derive formal results transparently.[8] We assume that labor supply is inelastically supplied and all individual earn the same wage w. This strong assumption is made for simplicity and in order to focus solely on the problem of redistribution of initial wealth through capital income taxation. The important extension with elastic labor supply is left for future work. Note also that our analysis carries over unchanged to the case with no wages, i.e., when $w=0$ implying that our key insight is largely orthogonal to the issue of endogenous labor supply, labor or consumption taxes (see our discussion below). All individuals discount the future at rate $\rho>0$ and maximize the intertemporal utility $U=\int_0^\infty u(c_t)e^{-\rho t}dt$. We make the following simplifying assumption:

Assumption 1. The real interest rate is exogenous and constantly equal to the discount rate ρ, wages are exogenous, uniform across individuals and over time, and denoted by w.

We show in Section 5.3 how relaxing Assumption 1 might affect the results. The assumption on the interest rate can be interpreted as the small open economy assumption where individuals can lend and borrow from abroad at a constant world market interest rate ρ.[9] The exogenous rate is taken as equal to the discount rate so that the economy converges to a steady-state (see below). We denote by a_t, the individual wealth level at time t. We assume that individuals differ only through their initial wealth endowment a_0.[10] The population is normalized to one and the cumulative distribution of initial wealth is denoted by $H(a_0)$, and the density by $h(a_0)$, with support A_0.

The government implements a capital income tax schedule possibly non-linear, and time varying denoted by $I_t(.)$, and distributes uniform (across individuals) lump-sum benefits b_t. We adopt without loss of generality the normalization $I_t(0)=0$, i.e., taxes are zero for individuals with no capital income. We denote by $y_t=w+b_t$ the annual stream of non capital income. The individual wealth accumulation equation can be simply written as

$$\dot{a}_t=\rho a_t-I_t(\rho a_t)+y_t-c_t. \tag{1}$$

Maximizing utility subject to the budget constraint (1) leads to the usual Euler equation

$$\frac{\dot{c}_t}{c_t}=\sigma\left[\rho\left(1-I'_t(\rho a_t)\right)-\rho\right]. \tag{2}$$

The transversality condition states that the present discounted value of consumption equals the present discounted value of the income stream so that debt (or assets) does not explode. Eqs. (1) and (2) combined with the initial condition $a(0)=a_0$, and the transversality condition defines a unique optimal path of consumption and wealth. We denote by $U(a_0)$ the utility of individual with initial wealth a_0, and by $Tax(a_0)$ the present discounted value (using the pre-tax interest rate) of tax payments of the individual with initial wealth a_0. Of course, utility and taxes depend on the path of tax schedules ($I_t(.)$) and the size of government benefits b_t.

2.2. Government tax instruments

2.2.1. Government objective

The government uses capital income taxation to raise an exogenous revenue requirement g_t and to redistribute a uniform lumpsum grant b_t to all individuals. We assume that the government maximizes a utilitarian social welfare function $\int_{A_0}U(a_0)dH(a_0)$ subject to the budget constraint

$$\int_{A_0}Tax(a_0)dH(a_0)\geq B+G \tag{3}$$

where B and G denote the present discounted value (at pre-tax interest rates) of government benefits b_t and exogenous spending g_t. Total taxes collected must finance the path of lumpsum grants b_t and government spending g_t. We denote by p the multiplier of the budget constraint (3). The analysis can be extended to more general social welfare functions. However, to keep the presentation simple, we focus first on the utilitarian case, and present the results for the general case in Section 5.2. We make the following additional simplifying assumption:

Assumption 2. The path of government lumpsum grants b_t is restricted to be constant overtime.

Assumption 2 requires some explanations. Implicit in Eq. (3) is the assumption that the government can use debt paying the same pre-tax rate as capital. We will see below that when all individuals face the same after-tax interest rate as in Chamley (1986), debt is neutral and does not allow the government to improve welfare. However, with non-linear capital income taxation, different individuals typically face different after-tax interest rates and debt is no longer neutral and can be used to improve welfare. We discuss in detail in Section 5.1 how debt can be used in conjunction with non-linear taxes to improve redistribution. Assumption 2 is a way to freeze the debt instrument by forcing the government to redistribute tax proceeds uniformly over time.

2.2.2. First best wealth taxation

Ideally, the government would like to make a wealth levy at time zero in order to finance all future government spending and equalize wealth if it cares about redistribution. As initial wealth is exogenous, this wealth levy is first-best Pareto efficient.[11] If political constraints limit the ability of the government to tax initial wealth at 100%, then there will remain wealth inequality after this wealth tax and our optimal tax problem carries over unchanged by simply replacing the initial wealth distribution by the post-wealth tax wealth distribution.

[8] As an important caveat, we note that our results might not necessarily easily carry over to general utility functions $u(c)$.

[9] We assume implicitly that capital income taxation is on a residency basis, i.e., individuals are taxed in their country of residence based on their worldwide capital income, regardless of where their assets are invested. In that case, pre-tax rates of returns will be equalized across countries (even if countries impose different tax rates). We also assume that the government can observe the total level of capital income individual by individual to impose a progressive tax on capital income. This effectively assumes that individuals cannot arbitrage the progressive tax by sharing their assets. This is how most individual tax systems operate. In practice however, tax evasion through off-shore tax heavens limits the ability of governments to implement this ideal residence base tax.

[10] We discuss later on how introducing wage income heterogeneity may affect the results.

[11] This perfect equalization is similar to the perfect equalization of after-tax income that takes place in a static optimal income tax model with no behavioral response and decreasing (social) marginal utility of consumption.

2.2.3. Capital income taxation

Therefore, in the analysis that follows, we assume that the government cannot implement a wealth levy and has to rely on distortionary capital income taxation. If there is no constraint on the maximum capital tax rate that the government can use, then, as shown in Chamley (1986), the government can replicate the first-best wealth levy using an infinitely large capital income tax rate during an infinitely small period of time. It is therefore necessary to set an exogenous upper-bound on the feasible capital income tax rate.

Assumption 3. The capital income tax schedules are restricted to having marginal tax rates always below an exogenous level $\tau > 0$.

We believe that this assumption captures a real constraint faced by tax policy makers. In practice, wealth levies happened only in very extraordinary situations such as wars, or after-war periods.[12] The political debates preceding the introduction of progressive income taxes in the United Kingdom in 1909, France in 1914, or the United States in 1913 provide interesting evidence on these issues. Populist and left-wing parties were the promoters of progressive income taxation for redistributive reasons and to curb the largest wealth holdings. Fierce opposition from the right prevented the implementation of more drastic redistributive policies such as wealth levies. Therefore, the situation where the government can only use income taxation to redistribute wealth is perhaps relevant in practice because of political constraints.

2.2.4. Consumption taxation

As explained by Chari and Kehoe (1999), the first-best wealth levy can be replicated with large consumption taxes (uniform over time) combined with a large lumpsum subsidy. Such a combination of taxes would make initial wealth less valuable, but would not distort relative prices. In the limit where these taxes and subsidies go to infinity, initial wealth becomes irrelevant and complete equalization is obtained as in the first-best wealth levy. Such an extreme policy is certainly unrealistic. However, the point remains that consumption taxes, even without going to the extreme case described above, would be more efficient than capital income taxation alone because they would allow one to replicate more closely a wealth levy than capital income taxation.[13] It is an interesting question why the political debates surrounding the introduction of progressive income taxation to curb large wealth holdings did not consider consumption taxation as a feasible means to redistribute wealth.

If there is a finite limit τ_c on the level of the consumption tax rate that the government can use, due perhaps to political constraints, then the government should use the maximum consumption tax to effectively confiscate a fraction $\tau_c/(1+\tau_c)$ of initial wealth. After including this maximum consumption tax, the optimal capital income tax problem takes exactly the same form as the one described above. Therefore, in this paper, we ignore the possibility of consumption taxation without loss of generality. It is important to keep in mind the caveat that an infinite consumption tax (or a 100% initial wealth tax as described above) would achieve complete first-best redistribution, with no necessity to use distortionary capital income taxes.

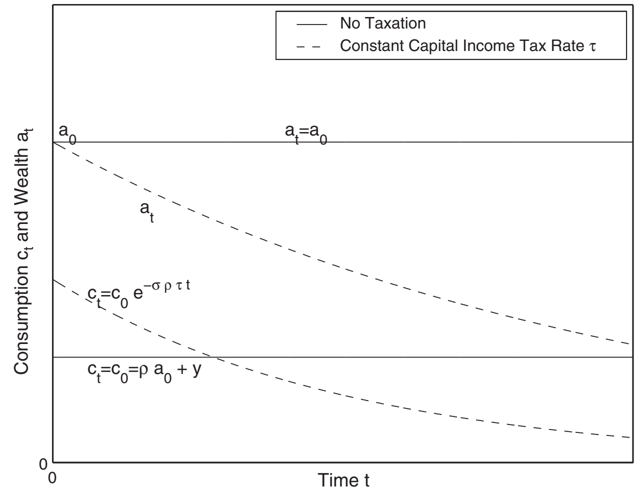

Fig. 1. Wealth and consumption dynamics.

2.3. Responses to taxation

2.3.1. The central trade-off

The derivation of optimal capital taxes relies critically on the behavioral responses to taxation and the induced effect on wealth accumulation. With no taxation ($I_t(.)=0$), the Euler Eq. (2) implies that the path of consumption is constant ($c_t = c_0$ for all t), and thus wealth a_t is also constant (otherwise the transversality condition would be violated). Consumption is equal to labor income and benefits plus interest income on wealth ($c = y + \rho a_0$). This case is depicted on Fig. 1 in straight lines. Therefore, in that situation, the wealth distribution remains constant over time and equal to the initial wealth distribution $H(a_0)$.

In the presence of taxation, let us denote by $\bar{r}_t = \rho\left(1 - I'_t(\rho a_t)\right)$ the instantaneous after-tax interest rate, and by $\bar{R}_t = \int_0^t \bar{r}_s ds$ the cumulative after-tax interest rate. The Euler Eq. (2) can be integrated to obtain $c_t = c_0 e^{\sigma(\bar{R}_t - \rho t)}$. Thus a positive and constant over time marginal tax rate τ produces a decreasing pattern of consumption over time $c_t = c_0 e^{-\rho\sigma\tau t}$, as depicted on Fig. 1 (in dashed line). In that case, the high initial level of consumption in early periods has to be financed from the initial wealth stock. Therefore, positive marginal tax rates produce a declining pattern of wealth holding as shown on Fig. 1.

Fig. 1 illustrates well the equity–efficiency trade-off that the government is facing. On the one hand, the government would like to use capital income taxation to redistribute wealth from the rich to the poor because this is the only instrument available. On the other hand, using capital income taxation leads the rich rentiers to run down their wealth, which reduces the capital income tax base in later periods.

Importantly, we make the assumption $r = \rho$ to ensure a stable steady-state in the case with no taxation. As we shall see, capital income taxation does vanish in the long-run in our models. The assumption $r = \rho$ then ensures a stable steady-state under the optimal tax regime.[14]

[12] For example, just after World War II, the French government confiscated property of the rich individuals accused of having collaborated with the Nazi regime during the occupation. These confiscations were de facto a wealth levy. Similarly, Japan, in the aftermath of World War II applied, confiscatory tax rates on the value of property in order to redistribute wealth from those who did not suffer losses from war damage to those who did.

[13] It is well known that switching from income taxation to consumption taxation would amount to taxing existing wealth. See Auerbach and Kotlikoff (1987) for such an analysis in an OLG model.

[14] Assuming $r \neq \rho$ would lead to a degenerate steady-state. Note that in a closed economy context, the steady-state is always stable as the capital stock responds to meet the condition $f'(k) = \rho$ (in the case with no capital taxes in the long-run). We discuss the closed-economy case in Section 5.3 but leave the full optimal tax analysis in that context to future research.

2.3.2. Tax revenue

In order to derive optimal tax results, it is useful to assess how a change in taxes affects tax revenue. The present discounted value (at pre-tax interest rates) of taxes collected on a given individual is equal to $Tax(a_0)=\int_0^\infty I_t(\rho a_t)e^{-\rho t}dt$. Integrating Eq. (1), and using the transversality condition, one obtains that taxes collected are also equal to initial wealth a_0 plus the discounted value of the income stream y less the discounted value of the consumption stream c_t:

$$Tax(a_0)=a_0+\int_0^\infty\left[y-c_t\right]e^{-\rho t}dt=a_0+\frac{y}{\rho}-c_0\int_0^\infty e^{\sigma(\bar{R}_t-\rho t)-\rho t}dt. \quad (4)$$

This equation shows clearly how a behavioral response in c_0 due to a tax change triggers a change in tax revenue collected. A very large c_0 (consequence of high marginal tax rates and a distorted consumption pattern as in Fig. 1) may imply a lower level of taxes collected.

2.3.3. Effect of taxes on initial consumption

Initial consumption c_0 is defined so that the transversality condition is satisfied. The response of c_0 to capital income taxation is critical to assess the effect of changes in taxation on the tax base (as illustrated on Fig. 1), and hence, on taxes collected (as shown in Eq. (4)).

An increase in the capital income tax rate at time t^* produces an increase in the consumption prices $e^{-\bar{R}_t}$ *after* time t^*. As is well known, this increase in prices after time t^* leads to three effects on c_0. First, there is a *substitution* of consumption after t^* toward consumption before t^* leading an *increase* in c_0. Second, the increase in prices leads to a negative *income* effect on consumption and thus on c_0. As usual, when $\sigma=1$ (log utility case), income and substitution effects exactly cancel out. Third, the increase in prices also increases the value of the income stream y_t and thus produces a positive *human wealth* effect on consumption and hence on c_0. These three effects will show up in the optimal tax analysis below.

3. Linear taxation and preliminary results

In this section, we examine individual consumption and wealth accumulation decisions under linear taxation. We then investigate whether it would be efficient for the government to tax (using individual-specific linear taxation) richer individuals for a longer period of time. As progressive taxation allows one to precisely discriminate taxpayers based on the size of their capital income (or equivalently wealth), the results obtained in this section will be of much use to tackle the optimal progressive income tax problem.

3.1. Linear income taxes and individual behavior

We consider first the case where the government implements linear capital income taxes (possibly time varying). As the policy which comes closest to the first-best wealth levy is to tax capital as much as possible early on, the optimal policy consists in imposing the maximum tax rate τ on capital income up to a time T and zero taxation afterwards. This "bang-bang" pattern of taxation was shown to be optimal in a wide class of dynamics models by Chamley (1986). For notational simplicity, we assume that $\tau=1$, that is, the maximum rate is 100%.[15]

Let us assume therefore that the government imposes a linear capital income tax with rate 100% up to time T, and with rate zero after time T. In the notation introduced in Section 2, $\bar{R}_t=0$ if $t\le T$ and $\bar{R}_t=\rho(t-T)$ if $t\ge T$. After time T, the Euler Eq. (2) implies that $\dot{c}_t=0$, and thus constant consumption $c_t=c_T$. As $y=w+b$ is also constant, wealth a_t must also be constant after time T and such that $c_T=\rho a_T+y$.

Before time T, the Euler equation implies $\dot{c}/c=-\sigma\rho$, and therefore $c_t=c_0e^{-\sigma\rho t}$. The wealth equation implies $\dot{a}_t=y-c_t$, and therefore using the initial condition for wealth, we have

$$a_t=a_0+y\cdot t-\frac{c_0}{\sigma\rho}\left(1-e^{-\sigma\rho t}\right). \quad (5)$$

There is a unique value c_0 such that the path for wealth (Eq. (5)) for $t=T$ matches the constant path of wealth $a_T=(c_0e^{-\sigma\rho T}-y)/\rho$ after T

$$c_0=\frac{\sigma\left[y+\rho(y\cdot T+a_0)\right]}{1-(1-\sigma)e^{-\sigma\rho T}}. \quad (6)$$

We denote by $a_\infty(a_0)$ and $c_\infty(a_0)$ the (constant) values of wealth and consumption after time T. The individual patterns of consumption and wealth are depicted in straight lines on Fig. 2. Using Eq. (4), the present discounted value of total capital income taxes collected is

$$Tax(a_0,T)=\int_0^T\rho a_te^{-\rho t}dt=\frac{y}{\rho}+a_0-\frac{c_0}{\rho}\cdot\frac{1+\sigma e^{-(\sigma+1)\rho T}}{1+\sigma}. \quad (7)$$

3.2. Uniform linear taxes

In this subsection, we consider the case where the government has to set the same linear taxes on all individuals. This is the standard case studied in Chamley–Judd and the subsequent literature. In that case, the time of taxation T has to be the same for all individuals. The optimal time T and benefit level b are obtained by forming the Lagrangian $L=\int_{A_0}U(a_0)dH(a_0)+p\left[\int_{A_0}Tax(a_0,T)dH(a_0-b/\rho-G)\right]$, and taking the first order conditions with respect to b and T.

The interesting point to note is that this type of taxation does not *qualitatively* change the nature of the wealth distribution in the long-run. Using Eqs. (5) and (6) for large values of a_0, it is easy to show that $a_\infty(a_0)\sim\mu\cdot a_0$ where $0<\mu=\sigma e^{-\sigma\rho T}/(1-(1-\sigma)e^{-\sigma\rho T})<1$. Therefore, large fortunes are reduced by a proportional factor $\mu<1$, but the shape of the top tail of the wealth distribution is not qualitatively altered. For example, if the initial wealth distribution is Pareto distributed at the top with parameter α, then the distribution of final wealth will also be Pareto distributed with the same parameter

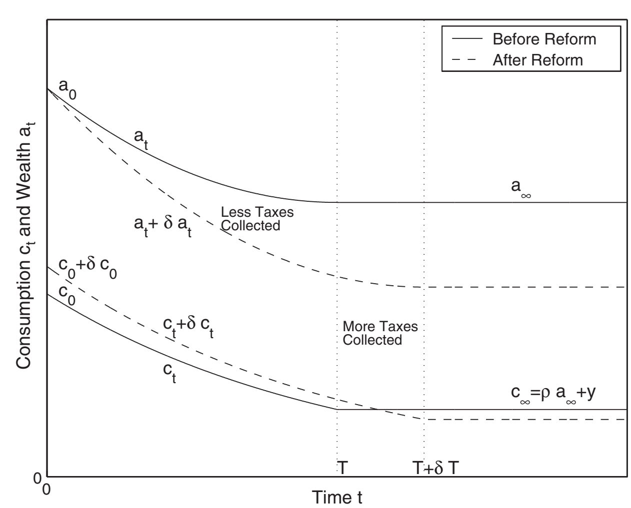

Fig. 2. Increasing the time of taxation *T*.

[15] The key results are independent of the maximum tax rate τ (see below).

α. To our knowledge, the interesting question of how much redistribution of wealth is achieved by the optimal set of linear taxes, as a function of the parameters of the model and the redistributive tastes of the government, has not been investigated with numerical simulations in the literature.

3.3. Wealth specific linear income tax

In this subsection, we assume that the government can implement linear capital income taxes (possibly time varying) that *depend* on the initial wealth level a_0. This set-up does not correspond to a realistic situation but it is a helpful first step to understand the mechanisms of wealth redistribution using capital income taxes in the infinite horizon model.[16] As a direct extension of the Chamley (1986) bang-bang result, it is easy to show that the optimal policy for the government in that context is to impose the maximum allowed tax rate τ on capital income up to a time period $T(a_0)$ (which now depends on the initial wealth level) and no tax afterward. There are two interesting questions in that model. First, how does T vary with a_0? That is, does the government want to tax richer individuals longer? And for which reasons (redistribution, efficiency, or both)? Second, what is the asymptotic wealth distribution when the set of optimal wealth specific income taxes is implemented?

To simplify the notation, we assume again that $\tau = 1$ (this does not affect the nature of the results). In this context, the government chooses the optimal set of time periods $T(a_0)$, and benefits levels b that maximize social welfare subject to the budget constraint (3). The first order condition with respect to $T(a_0)$ is $\partial U(a_0)/\partial T(a_0) + p\partial Tax(a_0)/\partial T(a_0) = 0$. This condition states that an individual with initial wealth a_0 should be taxed up to the time $T(a_0)$ such that the social welfare loss created by an extra time of taxation is equal to the extra revenue obtained. We show formally in Appendix A the following proposition.

Proposition 1. • *If $\sigma < 1$, then asymptotically (i.e., for large a_0)*

$$T(a_0) \sim \frac{1}{\sigma\rho} \log a_0, \quad a_\infty(a_0) \to \frac{\sigma}{1-\sigma} \cdot \frac{y}{\rho}. \tag{8}$$

Therefore, the asymptotic wealth distribution is bounded.

- *If $\sigma > 1$ then asymptotically (i.e., for large a_0), $T(a_0)$ converges to a finite limit T^∞, and $a_\infty(a_0) \sim a_0 \cdot \sigma e^{-\sigma\rho T^\infty}/[1+(\sigma-1)e^{-\sigma\rho T^\infty}]$.*

It is important to understand the economic intuitions behind the proof Proposition 1. Note first that the social marginal value of an extra dollar given at time zero to an individual with wealth a_0 is given by $\partial U/\partial a_0 = u'(c_0) = c_0^{-1/\sigma}$. As c_0 grows to infinity when initial wealth a_0 grows without bound, the social marginal utility of the rich goes to zero as wealth goes to infinity. Therefore, under a utilitarian criterion the government hardly values marginal wealth of the very rich and hence the optimal tax system extracts the maximum amount of tax revenue from the richest individuals.[17] Naturally, this result would also be true under a Rawlsian criterion where the social marginal value of consumption is zero for everybody except the poorest.[18] Therefore, the social welfare effect term $\partial U(a_0)/\partial T(a_0)$ can be ignored in the derivation and only revenue effects $\partial Tax(a_0)/\partial T(a_0)$ matter. As we shall see, revenue effects depend on both income and substitution effects, explaining the role of income effects in the derivation.[19]

As shown on Fig. 2, when the time of taxation T is increased by dT, there are two effects on taxes collected. First, as the time of taxation increases, taxes are collected for a longer time, increasing *mechanically* tax revenue. Second, the tax change produces a behavioral response which might increase (or decrease) c_0 and hence decrease (or increase) the path of wealth a_t, inducing a decrease (or increase) in taxes collected before time T (Fig. 2 depicts the case where c_0 increases). Let us analyze the effect of T on c_0. Using Eq. (6), the effect of an extra time of taxation dT on c_0 is given by

$$\frac{\partial c_0}{\partial T} = \sigma\rho \cdot \frac{y - c_0 e^{-\sigma\rho T} + \sigma c_0 e^{-\sigma\rho T}}{1-(1-\sigma)e^{-\sigma\rho T}}. \tag{9}$$

Therefore, as displayed in the numerator of Eq. (9) and as discussed informally in Section 2.3, the marginal effect of T on c_0 can be decomposed into three effects. The first term in the numerator of Eq. (9) is the human wealth effect: when the time of taxation increases, the present discounted value of the income stream y increases and thus consumption goes up. The human wealth effect is positive and goes away when the individual does not receive any income stream ($y = 0$). The second term is the income effect and is negative: a longer time of taxation increases the relative price of consumption after time T and thus reduces c_0 through an income effect. The third and last term is the substitution effect and is positive: increasing the price of consumption after time T relative to before time T shifts consumption away from the future toward the present and produces an increase in c_0. As always, when $\sigma = 1$, the income and substitution effects exactly cancel out.

When $\sigma > 1$, the substitution effect dominates the income effect. Thus, increasing T unambiguously increases c_0, producing a reduction in tax revenue (case depicted on Fig. 2). The mechanical increase in tax revenue is due to extra tax collected between times T and $T + dT$. Because of discounting at rate ρ, this amount is small relative to dT when T is large. As a result, the behavioral response tax revenue effect dwarfs the mechanical increase in tax revenue if T is large. As the welfare effect of increasing T is also negative, T can clearly not grow without bounds when a_0 grows. Therefore, T has to converge to a finite limit T^∞ no matter how strong the redistributive tastes of the government.

Therefore, in the case where $\sigma > 1$, wealth specific capital income taxes are not a very useful tool for redistributing wealth because the behavioral response to capital income taxes is very large. As a result, taxes are zero after a finite time T^∞ and the resulting wealth distribution is not drastically affected by optimal capital taxation (as in the uniform linear tax case of Section 3.2).

When $\sigma < 1$, the income effect dominates the substitution effect. For large a_0, initial consumption c_0 is large relative to y (because the capital income stream dwarfs the annual income stream y). Thus, and as can be seen from Eq. (9), unless T is large, the income effect (net of the substitution effect) dwarfs the human wealth effect, and therefore the response in c_0 is negative, generating more tax revenue. Therefore, if T does not grow without bounds, increasing T would generate more revenue and hence would increase social welfare implying that this cannot be an optimum. Thus, at the optimum, T must grow without bounds when a_0 grows so that the income effect (net of the substitution effect) is compensated by the human wealth effect.[20] Therefore, using the numerator of Eq. (9), T must be such that $(1-\sigma)c_0 e^{-\sigma\rho T} \approx y$, implying that long-run consumption must be such that $c_T \approx y/(1-\sigma)$, and therefore the long-run wealth level needed to finance this consumption stream is $a_T \approx (y/\rho) \cdot \sigma/(1-\sigma)$ as stated in Eq. (8).

Therefore, when the elasticity of substitution σ is below unity, the government would like to tax larger fortunes longer until they are

[16] As pointed out earlier, it is informationally inconsistent to assume that the government can observe a_0 to tailor T but yet cannot confiscate wealth entirely.

[17] We discuss extensions where the government cares about the social marginal consumption of the rich in Section 5.2.

[18] See Saez, 2001 for a more detailed discussion of the equivalence of the Rawlsian criterion and the utilitarian criterion for top marginal income tax in a standard Mirrlees (1971) optimal tax model when marginal utility goes to zero with income.

[19] This is similar to the standard optimal labor income tax model of Mirrlees (1971) where the optimal asymptotic tax rate depends on both income and substitution effects (see Saez, 2001, Section 3).

[20] One can check that, for large a_0, the welfare effect is small relative to the increase in tax revenue.

reduced to a finite threshold given in Eq. (8). If the initial wealth distribution is unbounded, at any time t no matter how large, there will remain (at least a few) large fortunes that continue to be taxed. This result is a significant departure from the zero tax result of Chamley (1986) and Judd (1985). In the long run, the largest fortunes produce a stream of interest income equal to $\sigma y/(1-\sigma)$. For example, with $\sigma = 1/2$ (which falls in the range of empirical estimates, see below), the largest fortunes would only allow their owners to double their labor plus government benefits annual stream of income.

It is important to note that this result relies on the fact that, for the very wealthy, annual labor plus benefits income y is small relative to the stream of capital income, and therefore the human wealth effect is small relative to the income effect. This result needs to be qualified when y is positively related to a_0. If the wealthy have a labor income stream proportional to their initial wealth, then the human wealth effect will be of the same order as the income effect for finite T. In that case, asymptotic wealth will be proportional to y, and hence to a_0 producing an unbounded asymptotic wealth distribution. Therefore, the theory developed here shows that taxing wealthy rentiers is much more desirable than taxing capital income from the working rich. Conversely, if $y = 0$ for the rich, then with $\sigma < 1$, taxing the rich longer always raises more revenue as there is no human wealth effect and the income effect dominates the substitution effect. This situation does not arise in our model as taxes are rebated lumpsum so that $y = b + w > 0$ even with $w = 0$.

4. Optimal progressive taxation

Obviously, the wealth specific linear income tax analyzed in the previous section is not a realistic policy option for the government. However, in practice, the government can use a tool more powerful than the uniform linear taxes of the Chamley (1986) and Judd (1985) model, namely progressive or non-linear capital income taxation. As discussed in the Introduction, actual tax systems often impose a progressive tax burden on capital income. Many countries, including the United States, impose estate or inheritance taxation with substantial exemption levels and a progressive structure of marginal tax rates. Most individual income tax systems have increasing marginal tax rates and capital income is often in large part included in the tax base, producing a progressive capital income tax structure.[21]

Non-linear capital income taxes in the infinite horizon model are appealing, in light of our results on wealth specific linear taxation, because a non-linear schedule allows to discriminate among taxpayers on the basis of wealth. A progressive tax structure can impose high tax burdens on the largest fortunes while completely exempting from taxation modest fortunes.[22]

4.1. A simple two-bracket progressive capital tax

The progressive tax structure that comes closest to the wealth specific linear taxation is the following simple two-bracket system. At each time period t, the government exempts from taxation all individuals with wealth a_t below a given threshold a_t^* (possibly time varying), and imposes the maximum marginal tax rate τ on all capital income in excess of ρa_t^*, as depicted on Fig. 3. Note that the progressive schedule creates a *virtual income* $m_t = \tau\rho a_t^*$ for those in the tax bracket.

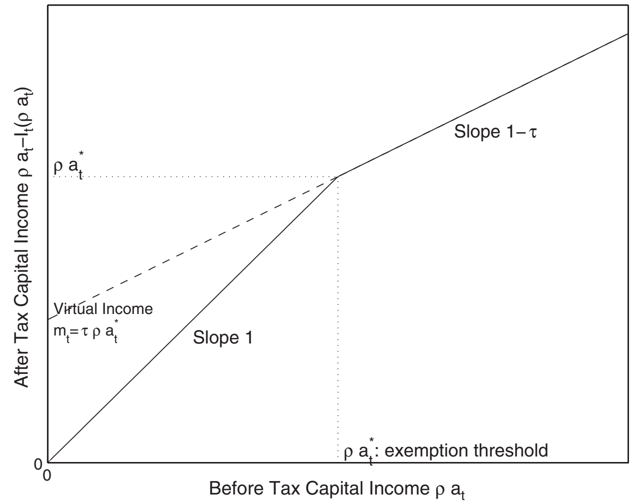

Fig. 3. Two-bracket progressive capital income tax.

As long as $\tau > 0$, none of our results stated in Propositions 2 and 3 are sensitive to the level of τ. Therefore, to simplify the presentation, we consider in the text the case $\tau = 1$. In that case, $I_t(\rho a_t) = 0$ if $a_t \le a_t^*$, and $I_t(\rho a_t) = \rho(a_t - a_t^*)$ if $a_t > a_t^*$. Because we have adopted the normalization $I_t(0) = 0$, we assume that $a_t^* \ge 0$ so that individuals with zero wealth have no tax liability.[23] We also impose the condition that the exemption threshold a_t^* is non-decreasing in t (see below for a justification), and we denote by $A_t^* = \int_0^t a_s^* ds$ the integral of the function a_t^*.

The dynamics of consumption and wealth accumulation of this progressive tax model are very similar to those with the wealth specific linear tax and are depicted on Fig. 4. Individuals (with initial wealth $a_0 > a_0^*$) first face a 100% marginal tax rate regime. From the Euler Eq. (2), individual consumption is such that $c_t = c_0 e^{-\sigma\rho t}$, and individual wealth evolves according to $\dot{a}_t = \rho a_t^* + y - c_t$, implying

$$a_t = a_0 + \rho A_t^* + y \cdot t - \frac{c_0}{\sigma\rho}\left(1 - e^{-\sigma\rho t}\right). \tag{10}$$

The only difference with Eq. (5) is the presence of the extra-term ρA_t^* due to the presence of the exemption threshold.

It is easy to show that wealth a_t declines up to the point where it reaches a_t^*. This happens at time T, which naturally depends on a_0, such that

$$a_T^* = a_0 + \rho A_T^* + y \cdot T - c_0\left(1 - e^{-\sigma\rho T}\right) / (\sigma\rho).$$

After time T, the individual is exempted from taxation and therefore has a flat consumption pattern $c_t = c_0 e^{-\sigma\rho T}$ and a flat wealth pattern $a_t = a_T^* = (c_T - y)/\rho$. Therefore, as depicted on Fig. 4, the pattern of consumption is exponentially decreasing up to time T and flat afterwards. The wealth pattern is also declining up to time T, and flat afterwards.[24] Routine computations paralleling the analysis of Section 3.2 show that

$$c_0 = \frac{\sigma\left[y + \rho(y \cdot T + \rho A_T^* + a_0)\right]}{1 - (1-\sigma)e^{-\sigma\rho T}}, \quad c_0 e^{-\rho\sigma T} = \rho a_T^* + y \cdot \tag{11}$$

[21] In the United States (and in many other countries as well), the development of tax-exempted instruments to promote retirement savings such as Individual Retirement Accounts and 401(k) plans that are subject to maximum annual contribution levels also create a progressive structure.

[22] Obviously, progressive taxation cannot be as efficient as the wealth specific linear capital income taxation of Section 3.3 because reduced marginal tax rates for low incomes lowers the tax burden on higher incomes.

[23] It would be optimal for the government to set a_t^* large and negative for low t in order to replicate a lumpsum tax at time zero which would be equivalent to a wealth levy. Imposing the constraint $a_t^* \ge 0$ effectively rules out this possibility.

[24] Note that, as depicted on Fig. 4, at $t = T$, the wealth pattern is flat because $\dot{a}_t = \rho a_t^* + y - c_0 e^{-\sigma\rho t} = 0$ when $t = T$. As a_t^* is non-decreasing and c_t is decreasing, $\dot{a}_t = \rho a_t^* + y - c_t$ is increasing. As $\dot{a}_T = 0$, this implies that $\dot{a}_t < 0$ for $t < T$ confirming that a_t decreases with t for $t < T$.

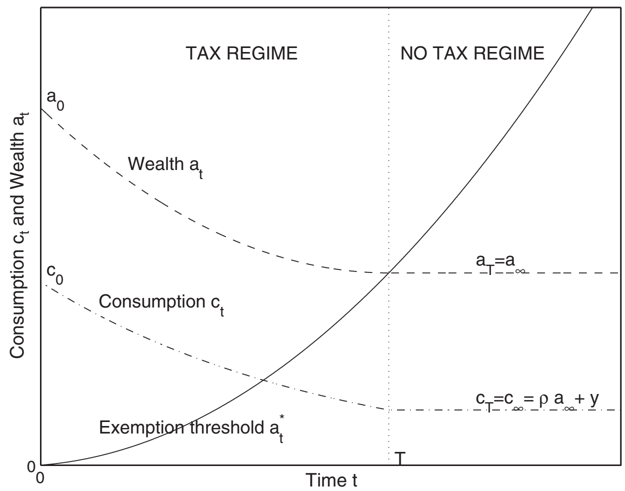

Fig. 4. Wealth and consumption dynamics with progressive taxation.

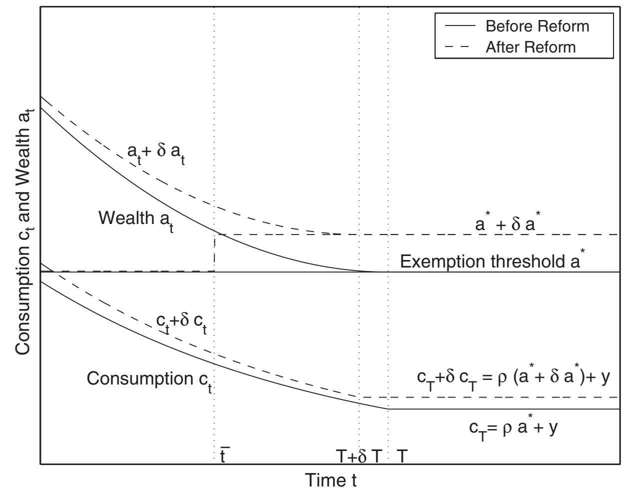

Fig. 5. Increasing the exemption threshold.

Eq. (11) implicitly defines T as an increasing function of a_0. Intuitively, individuals with higher wealth remain in the tax regime longer than individuals with lower wealth. For any given path a_t^*, the time of taxation $T(a_0)$ is increasing in a_0.[25] We denote as above the (constant) levels of consumption and wealth after time T by $c_\infty(a_0)$ and $a_\infty(a_0)$. Note that $a_\infty(a_0) = a_{T(a_0)}{}^*$ is a non-decreasing function of a_0 as a_t^* is non-decreasing in t and $T(a_0)$ increases with a_0.

Using Eq. (4), the present discounted value of taxes paid by an individual with initial wealth a_0 is:

$$Tax(a_0, T) = \int_0^T \rho\left[a_t - a_t^*\right]e^{-\rho t}dt = \frac{y}{\rho} + a_0 - \frac{c_0}{\rho}\cdot\frac{1+\sigma e^{-(\sigma+1)\rho T}}{1+\sigma}. \qquad (12)$$

Note that expression (12) is identical to expression (7). For a given initial consumption level c_0 and a given time of taxation T, the non-linear tax system raises exactly the same amount of taxes than the linear tax system. The key difference appears in Eq. (11). The initial level of consumption c_0 contains an extra-term ρA_T^* reflecting the extra virtual income due to the exemption of taxation below the threshold a_t^*. From now on, we call this effect the virtual income effect.

This non-linear tax system may improve substantially over the uniform linear tax system à la Chamley (1986) because large wealth holders can be taxed longer than poorer individuals.[26] For low values of σ, our previous results suggest that this is a desirable feature of the tax system. The non-linear tax system, however, is inferior to the wealth specific capital income tax of Section 3.3 because it exempts wealth holdings below a_t^* from taxation and creates a positive virtual income effect on c_0, and thus is not as efficient to raise revenue.

The central question we want to address is about the optimal asymptotic pattern for a_t^*. Does a_t^* tend to a finite limit a_∞^*, implying that, in the long-run, the wealth distribution is truncated at a_∞^*? Or does it diverge to infinity, implying that the wealth distribution remains unbounded in the long-run?

4.2. Optimal asymptotic tax

To tackle this question, let us assume that a_t^* is constant (say equal to a^*) after some large time level $\bar{t}$. I denote by $\bar{a}_0$ the wealth level of the person who reaches the exemption threshold a^* at time $\bar{t}$, that is, such that $T(\bar{a}_0) = \bar{t}$. Let us consider the effects of the following small tax reform. The exemption threshold a^* is increased by δa^* for all t above $\bar{t}$ as depicted on Fig. 5. Only individuals with initial wealth high enough (such that $a_0 > \bar{a}_0$) are affected by the reform. We denote by δc_0, δT, and δa_t the changes in c_0, $T(a_0)$, and a_t induced by the reform. We first prove the following lemma.

Lemma 1. *For large $\bar{t}$ (and hence T), we have*

$$\delta c_0 \approx \rho\left[\sigma\rho(T-\bar{t}) - \sigma\right]\delta a^*. \qquad (13)$$

The formal proof follows from the differentiation of Eq. (11). These differentiated equations express the endogenous δc_0 and δT in terms of the exogenous δa^*. Eliminating δT, we can obtain $\delta c_0 = \sigma\rho\cdot\left[\rho(T-\bar{t})-1\right]\delta a^*/(1-e^{-\sigma\rho T})$. When $\bar{t}$ (and hence T) is large, this equation can be approximated as Eq. (13). QED.

Let us provide the economic intuition. The small reform increases the virtual income m_t by δa^* between times $\bar{t}$ and T. As can be seen from Eq. (11) assuming T is large, this produces a direct positive virtual income effect $\rho\sigma\rho(T-\bar{t})\delta a^*$ on c_0. This is the first term in Eq. (13).

As can be seen on Fig. 5, after the reform, the time needed to reach the exemption threshold is reduced by $\delta T < 0$ because the exemption threshold is higher. This change in T produces a pure negative substitution effect on c_0.[27] For large $\bar{t}$ and hence T, Eq. (11) shows that the substitution effect on c_0 is approximately $\sigma\rho\sigma e^{-\sigma\rho T}c_0\delta T = -\sigma\rho\delta a^*$.[28] This is the second term in Eq. (13).

Eq. (13) shows that increasing the exemption threshold induces a positive effect on consumption for individuals with T far above $\bar{t}$ (i.e., the richest individuals) and a negative effect for those whose T is close to $\bar{t}$ (i.e., the poorest individuals affected by the reform). The explanation is the following. Individuals with large T benefit from the increased exemption for a long time and thus the direct virtual income wealth effect is large, and therefore they can afford to consume more. Individuals with T close to $\bar{t}$ do not benefit from this wealth effect and face only the indirect substitution effect. They

---

[25] The assumption that a_t^* be non-decreasing in time is important and simplifies considerably the analysis. If a_t^* were decreasing in some range, then individuals who were out of the tax bracket may enter the tax regime again, producing complicated dynamics. As we discuss below, we are interested on whether a_t^* diverges to infinity when t grows, therefore the constraint a_t^* increasing is not an issue for our analysis.

[26] The uniform tax system of Section 3.2 can be seen as a particular case of non-linear taxation with $a_t^* = 0$ up to time T and $a_t^* = \infty$ after T.

[27] As $c_T = \rho a_T^* + y$, the income effect and the human wealth effect (which must also include the virtual income ρa_T^*) exactly cancel out.

[28] δT is obtained by differentiating $c_0 e^{-\sigma\rho T} = y + \rho a^*$.

reach the higher exemption threshold sooner and thus the reform reduces the price of consumption after T relative to consumption before T and thus they reduce their initial consumption level.

It is useful to change variables from T to a_0. Using Eq. (11), we have, for T large, $c_0 \approx \sigma\rho a_0$. Thus, as $c_0 e^{-\sigma\rho T} = y + \rho a^*$, we have $\sigma\rho T \approx \log a_0 + \log(\sigma\rho) - \log(y + \rho a^*)$. Applying this equation at T and $T = \bar{t}$ (remembering that $T(\bar{a}_0) = \bar{t}$), we can rewrite Eq. (13) as $\delta c_0 \approx \rho[\log(a_0/\bar{a}_0) - \sigma]\delta a^*$. Using Eq. (12), and the expression for δc_0 just obtained, for large $\bar{t}$ and T, we have, up a first order approximation[29]

$$\delta Tax(a_0) \approx -\frac{\delta c_0}{\rho(1+\sigma)} \approx \frac{\delta a^*}{\sigma+1}\left[\sigma - \log\frac{a_0}{\bar{a}_0}\right]. \quad (14)$$

Eq. (14) shows that increasing the exemption threshold above $\bar{a}_0$ increases the tax liability of the rich for whom a_0 is slightly above $\bar{a}_0$ (the substitution effect reducing c_0 dominates) and decreases the tax liability of the super-rich for whom a_0 is far above $\bar{a}_0$. The net effect over the population is therefore going to depend on the number of super-rich relative to the number of rich. Integrating Eq. (14) over the distribution of wealth above $\bar{a}_0$, we obtain the effect of the reform on aggregate tax revenue:

$$\delta Tax \approx \frac{\delta a^*}{\sigma+1}\int_{\bar{a}_0}^{\infty}\left[\sigma - \log\frac{a_0}{\bar{a}_0}\right]h(a_0)da_0 = \frac{\delta a^*}{\sigma+1}\left[\sigma - A(\bar{a}_0)\right]\cdot\left[1 - H(\bar{a}_0)\right] \quad (15)$$

where $A(\bar{a}_0) = E(\log(a_0/\bar{a}_0)|a_0 \geq \bar{a}_0)$ is the normalized average log of wealth holding above $\bar{a}_0$. From Eq. (14), it is easy to see that the direct virtual income effect of the reform is captured by the term $A(\bar{a}_0)$ in the square brackets while the indirect substitution effect is simply the term σ in the square brackets.

4.2.1. Bounded initial wealth distribution

If the initial wealth distribution is bounded with a top wealth a_0^{top}, then when $\bar{t}$ is close to the maximum time of taxation, $\bar{a}_0$ is close to a_0^{top}, and $A(\bar{a}_0)$ is close to zero. As a result, Eq. (15) shows that the effect of the reform on tax revenue is unambiguously positive because, as discussed above, the virtual income effect is dominated by the substitution effect.

As the welfare effect is also obviously positive, it is always beneficial for the government to increase the exemption level at the top starting from a situation with constant a^* close to the top. This reform improves the incentives of the richest individual to accumulate wealth and thus would increase his tax liability while producing no effect on all the other taxpayers. This feature is similar to the zero top rate result in the Mirrlees (1971) model of optimal income taxation. In the Mirrlees model, a positive top marginal tax rate is suboptimal because reducing it would improve the incentives to work of the highest income individual (and hence his tax liability) without affecting anybody else.

4.2.2. Unbounded initial wealth distribution

If the initial wealth distribution is unbounded, then, in the present model, by increasing the exemption level above $\bar{t}$, the government collects more taxes from the individuals whose T is close to $\bar{t}$ but looses tax revenue for the very rich whose T is well above $\bar{t}$. Obviously, whether the net effect is positive depends on the relative number of taxpayers in these two groups: that is the number of super-rich individuals relative to the number of rich individuals. Exactly the same logic applies in the Mirrlees (1971) model with unbounded income distributions (Diamond, 1998; Saez, 2001).

It turns out that, as in the Mirrlees (1971) model, the Pareto distributions are of central importance. When the top tail is Pareto distributed with parameter α, then $H(a_0) = 1 - C/a_0^{\alpha}$ and the statistic $E(\log(a_0/\bar{a}_0)|a_0 \geq \bar{a}_0)$ is constant over all values of $\bar{a}_0$ and equal to $1/\alpha$. Eq. (15) then becomes

$$\delta Tax \approx \frac{\delta a^*}{\sigma+1}\left[\sigma - \frac{1}{\alpha}\right]\cdot\left[1 - H(\bar{a}_0)\right]. \quad (16)$$

It is well known (since the work of Pareto (1896)) that Pareto distributions approximate extremely well the top tails of income and wealth distributions.[30] Using the large microfiles of individual tax returns publicly released by the Internal Revenue Service in the United States, it is possible to estimate empirically the key statistic $A(\bar{a}_0)$ as a function of $\bar{a}_0$. More precisely, I consider capital income defined[31] as the sum of dividends, interest income, rents, fiduciary income (trust and estate income), and I plot on Fig. 6 the average normalized log income above income $\bar{z}$ for a large range of values of $\bar{z}$. This statistic is remarkably stable for large values $\bar{z}$, around 0.65, showing that the top tail is Paretian with a parameter $\alpha = 1.5$.[32] Fig. 6 shows that the empirical function $A(\bar{a}_0)$ whose value must be zero for the top wealth level, remains stable around 0.6 and does not get to zero even for very large values.[33] Therefore, the Pareto distribution assumption is clearly the best one to understand optimal taxation of the very wealthy in the current model.

Formula (16) shows that when $\sigma\alpha < 1$, then starting from a constant exemption level a^* (after a large time level $\bar{t}$), increasing the exemption level reduces tax revenue. It can be shown that the welfare effect of this reform is negligible relative to the tax revenue effect. Therefore, it is optimal for the government to reduce a^*. As the exemption a_t^* must be increasing, this implies that a_t^* must converge to a finite value. On the other hand, if $\sigma\alpha < 1$, then increasing a^* does increase tax revenue and is therefore desirable, this implies that the function a_t^* diverges to infinity as t grows. We can now state our main result on optimal progressive taxation whose rigorous proof is presented in Appendix A.

Proposition 2. *Assume that the top tail of the initial wealth distribution is Pareto with parameter α.*

- *If $\sigma \cdot \alpha < 1$ then the threshold a_t^* converges to a finite limit a_∞^* and thus the asymptotic wealth distribution is truncated at a_∞^*. More precisely, a_t^* is constant and equal to a_∞^* for t large enough.*
- *If $\sigma \cdot \alpha > 1$ then the threshold a_∞^* grows to infinity and thus the asymptotic wealth distribution is unbounded. The Pareto parameter of the asymptotic wealth distribution is also equal to α.*

Two important lessons should be drawn from Proposition 2. First, the central long-run vanishing capital tax result of Chamley–Judd carries over to our model. In our model, the fraction of individuals subject to capital income taxation vanishes to zero in the long-run so that capital income tax revenue also converges to zero in the long-run.

Second, however, in contrast to linear capital taxation, optimal nonlinear capital income taxes can have a dramatic impact on the long-run wealth distribution. Proposition 2 shows that two parameters affect critically the desirability of capital income taxation to curb large wealth holdings. First, and as expected from Section 3, the intertemporal elasticity of substitution matters. The higher this elasticity, the larger the

[29] The exact formula, valid for any $\bar{t}$ and T is given in Appendix A.

[30] A number of studies have shown how Pareto distributions arise naturally when year to year individual income or wealth growth is stochastic and independent of size (see e.g., Champernowne, 1953; Gabaix, 2009).

[31] I exclude realized capital gains because realizations are lumpy and are not an annual stream of income.

[32] Statistics compiled by the Internal Revenue Service by size of dividends since 1927, and exploited in Piketty and Saez (2003) show that the Pareto parameter for dividend income from 1927 to 1995 has always been around 1.5–1.7.

[33] In fact, if the second wealth holder has half as much wealth than the top wealth holder, then $A(\bar{a}_0) = \log(2) \approx 0.7$ at the level of the second top wealth holder. This shows again that, as in the Mirrlees (1971) model, the top result applies only to the top income and thus is not relevant in practice.

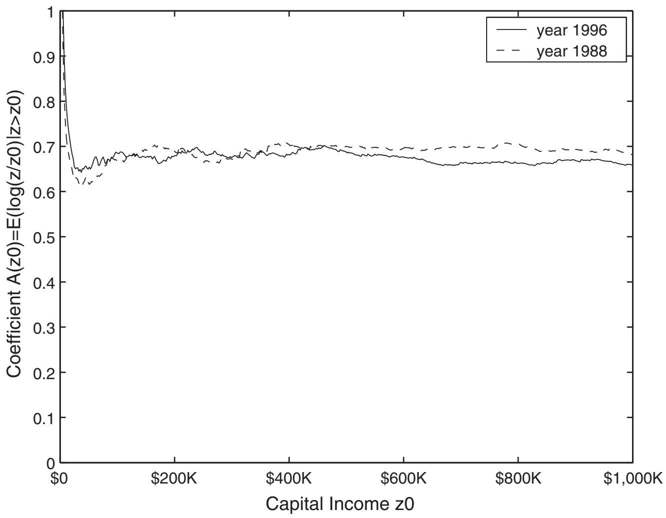

Fig. 6. Empirical statistic $A(z_0)=E(\log(z/z_0)|z>z_0)$ for capital income.

behavioral response to capital income taxation, and the less efficient are capital income taxes. Second and interestingly, the thickness of the top tail of the wealth distribution matters. The thinner the top tail of the distribution (as measured by the Pareto parameter α), the less desirable are capital income taxes. The intuition for this result is clear and is similar to the one obtained in the Mirrlees (1971) model of static labor income taxation. If the wealth distribution is thin, providing a tax break in the form of a higher exemption level for the rich is good for the wealth accumulation of the rich and bad for tax revenue collected from the super-rich. Therefore, granting the tax break is good when the number of super-rich is small relative to the number of rich individuals.

Finally, although we have presented all the results assuming that the capital income marginal tax rate τ above the exemption level is 100%, it is important to note that the asymptotic wealth distribution results apply for any $\tau>0$, no matter how small.

There is a large literature that tries to estimate the inter-temporal elasticity of substitution σ (see Deaton (1992) for a survey). Most studies find that consumption patterns are not very sensitive to the interest rate, and hence find a small inter-temporal elasticity of substitution σ, in general below 0.5.[34] It is important to note however that much uncertainty remains about the estimation of the inter-temporal elasticity of substitution with some studies obtaining large estimates. In particular, most macro-economic models are calibrated with log-utility with $\sigma=1$ because this parametrization is consistent with balanced growth paths. Note also that reduced form estimates tend to find close to zero uncompensated elasticities of savings with respect to the interest rate (see e.g., Bernheim, 2002 for a survey). In the CRRA case, this would imply that σ is close to one.

Pareto parameters of wealth distributions are almost always between 1.5 and 2. Therefore, the key condition $\sigma\cdot\alpha<1$ is empirically satisfied for some but not all estimates of the σ.

As discussed in Section 3, the case for using capital income taxation would be weaker if labor income y were positively related to initial wealth a_0.[35] In other words, capital income taxation should be used to tax rich rentiers but would be less desirable to tax the working rich.

5. Extending the basic model

5.1. Role of debt

As discussed in Section 2, with progressive (or wealth specific) capital income taxation, different individuals face different after-tax interest rates and debt is no longer neutral and can be used to improve welfare. An individual exempted from taxation is indifferent between one extra dollar at time 0 and $e^{\rho t}$ extra dollars at time t, while an individual facing a marginal capital income tax rate τ is indifferent between one extra dollar at time 0 and $e^{\rho(1-\tau)t}$ extra dollars at time t. Therefore, by distributing the lumpsum benefits b_t earlier on and creating debt, the government favors the low income untaxed relative to the high incomes who are taxed.[36] If no limit is set for the debt instrument, the government would distribute infinitely large lumpsum benefits earlier on, and implement an infinitely large lumpsum tax later on. Therefore, to avoid this degenerate and unrealistic outcome, a limit on the debt instrument must be introduced. That is why we introduced Assumption 2 in Section 2. Introducing other forms of debt limits such as period by period budget balance (where taxes equal transfers plus government spending at any point in time), or a finite limit on the size of debt, or an absolute limit on the size of lumpsum benefits or transfers, would not affect the asymptotic results obtained in Sections 3 and 4.[37] Hence, exactly as in the case of initial wealth taxation or consumption taxes, our results are robust to the introduction of debt as long as there is a limit on the debt instrument so that debt cannot achieve complete redistribution.

5.2. General welfare functions

In the derivations carried out so far, we have assumed that the government maximizes a utilitarian criterion. In that case, the social marginal value of an extra dollar given at time zero to an individual with wealth a_0 is given by $\partial U/\partial a_0=u'(c_0)=c_0^{-1/\sigma}$. As c_0 grows to infinity when initial wealth α_0 grows without bound, we see that the social marginal utility of the rich goes to zero as wealth goes to infinity. Therefore, the government hardly values marginal wealth of the very rich and thus the optimal tax systems that we have considered are designed to extract the maximum amount of tax revenue from the highest fortunes (soak the rich).

The important question we want to address here is how our results are modified if we assume that the social marginal value of wealth of the rich converges to some positive limit instead of zero. Therefore, let us extend our initial model and consider that the government maximizes some general social welfare function of the form $\int_{A_0} G(U(a_0))dH(a_0)$, where $G(.)$ is a (weakly) increasing function. The *direct* social marginal value of wealth of individual a_0 (expressed in terms of the value of public funds) is now given by $\beta(a_0)=G'(U(a_0))\cdot u'(c_0)/p$ where p is the multiplier of the government budget constraint. In the presence of income effects, giving one dollar at time zero to an individual with initial wealth a_0 produces, in addition to the direct welfare effect, a change in behavior and hence a change in tax revenue $\Delta T=dTax(a_0)/da_0$.

A simple way to incorporate the value of the change in tax revenue in the social marginal valuation of individual a_0 is to consider the following thought experiment. Suppose that the extra-tax revenue ΔT is rebated to the same individual, producing an extra welfare effect, and an extra income effect. Assuming that the extra tax is always

[34] The earliest studies based on macro data such as Hall (1988) found very small elasticities around 0.1. Later studies based on micro data tend to find bigger elasticities but most of the time below 0.5 (Attanasio and Weber, 1995).

[35] More precisely, it can be shown that if $y\sim a_0^\gamma$, then a^* converges to a finite limit only if $\sigma\cdot\alpha<1-\gamma$.

[36] Clearly, and as shown by Chamley (1986), this issue does not arise with uniform linear capital income taxation where debt is neutral.

[37] The presentation would have been more tedious as the income stream y_t would no longer have been constant. The value of y (which appears in Proposition 1) would also have been different.

rebated to the taxpayer, the total social value of giving one dollar is $g(a_0)=(\beta(a_0)/p)(1+\Delta T+\Delta T^2+\ldots)=(\beta(a_0)/p)/(1-dTax(a_0)/da_0)$. Hence, we can define $g(a_0)$ as the *total* social marginal welfare effect of wealth of individual a_0 (expressed in terms of the value of public funds).

The curve of total marginal social weights $g(a_0)$ describes how the government values giving a marginal dollar at any level of the wealth distribution and thus summarizes in a transparent way the redistributive tastes of the government. If the government has redistributive tastes, then $g(a_0)$ is decreasing. We denote by $\bar{g}$ the limit value of $g(a_0)$ when a_0 grows to infinity.[38] When $\bar{g}>0$, our two propositions are modified as follows.

Proposition 3. *In the wealth specific linear tax situation of Proposition 1, if $\sigma<1-\bar{g}$, then the asymptotic wealth distribution is bounded, and the asymptotic top wealth level is such that $\rho a_\infty(a_0)=\sigma\cdot y/(1-\bar{g}-\sigma)$. If $\sigma>1-\bar{g}$, then the optimal time of taxation converges to a finite limit and the asymptotic wealth distribution is unbounded.*

In the situation of Proposition 2, if $\sigma\cdot\alpha<1-\bar{g}$ then the exemption threshold a_t^* converges to a finite level and the asymptotic wealth distribution is truncated. If $\sigma\cdot\alpha>1-\bar{g}$ then a_t^* grows to infinity and the asymptotic wealth distribution is unbounded and Paretian with parameter α.

The proof is presented in Appendix A. Therefore, caring for the rich at the margin does have an impact on our results, and the condition needed to obtain a bounded asymptotic wealth distribution is stronger.

When the government does not care about redistribution, it sets equal marginal weights $g(a_0)$ for all individuals. Suppose that the government is then restricted to using distortionary capital income taxation to finance an exogenous amount of public spending G. In that situation, whether the asymptotic wealth distribution is truncated depends on the level of exogenous spending G. If G is low, the marginal efficiency cost of taxation is low and the asymptotic wealth distribution is unbounded. However, there is a threshold for public spending $\bar{G}$ above which the efficiency cost of taxation becomes high enough that it becomes efficient for the government to tax the rich sharply so that the asymptotic wealth distribution is truncated.

5.3. Endogenous interest rate and wages

Previous sections have considered the case with an exogenous interest rate $r_t=\rho$ and wage rate w, corresponding to the small open economy assumption with capital taxation based on residence of owners. It is an interesting question to know how our results are affected in the closed economy case with a neo-classical production function $f(k)$ where k denotes capital per capita. In that situation, $r=f'(k)$ and $w=f(k)-rk$. The initial capital stock per capita k_0 is given (and equal to the average a_0 if the economy starts with no debt).

We conjecture that our results carry over to the closed economy case. This is due to a general principle in optimal taxation theory stating that optimal tax formulas depend essentially on consumer elasticities and not on the elasticities of substitution in the production sector.[39] Importantly, at this stage this is solely a conjecture as the calculations in our proofs are tractable only when w and r are constant overtime. We therefore leave the important extension to a closed economy for future work.

It is nevertheless useful to point out striking steady-state results on the capital stock and wealth distribution with progressive taxation in a closed economy that have been established by Piketty (2001a).

With no capital taxation, the long-run stock of capital k_∞ is given by the modified Golden rule $f'(k_\infty)=\rho$. The intuition is the following. If the rate of return is below the discount rate, individuals accumulate wealth and the capital stock increases up to the point where the rate of return is reduced down to the discount rate. With a *linear* tax on capital income at rate τ in the long run, the stock of capital is lower and given by $(1-\tau)f'(k_\infty)=\rho$. It is interesting to note that the optimal set of taxes considered here always leads to the efficient level of capital $f'(k_\infty)=\rho$ in the long-run.

To see this, note first that a progressive capital income tax always leads to a stable steady-state. This can be seen from the Euler Eq. (2). Those with higher wealth – and hence higher capital income – face a lower net-of-tax rate of return $r\left(1-I'_t(ra_t)\right)$ and hence adopt a consumption path that decreases faster. Using the individual wealth accumulation Eq. (1), this implies that the wealthy reduce their wealth holdings faster. Therefore, all individual wealth levels have to converge so that in steady-state, all individuals are in the same tax bracket and face the same net-of-tax rate.[40]

With the simple two-bracket tax structure we have used, as demonstrated by Piketty (2001a), two cases can arise. First, everybody in the exemption bracket and the capital stock is given by the Golden rule $f'(k)=\rho$. This case arises if the exemption threshold a^* is larger than k. Second, everybody is in the tax bracket and the capital stock is given by $f'(k')=\rho(1-\tau)$. This case arises if the exemption threshold a^* is below k'.[41] If, as we conjecture, the optimal capital tax system in the closed economy case resembles the open economy case, then the fraction of taxpayers will converge to zero, so that everybody ends up in the exempt tax bracket and the long-term capital stock is given by the modified Golden rule $f'(k)=\rho$.

Therefore, in the infinite horizon closed economy model, even if the rich hold a substantial fraction of the capital stock, taxing them with progressive taxation does not have a negative impact on the long-run capital stock because lower wealth people will accumulate more and replace the capital stock lost by the rich (as long as the exemption threshold a^* is not too low). The model generates this striking result because everybody has the same discount rate ρ. This is a strong and unrealistic assumption of the standard infinite horizon model.[42]

6. Conclusion

This paper has shown that introducing progressive taxation in a basic optimal dynamic capital income tax model can affect policy prescriptions and especially long-term distributional outcomes. In the standard model with linear taxes, capital income taxes are zero after a finite time, and therefore the wealth distribution cannot be radically changed by capital income taxation. In contrast, with nonlinear taxation, if the intertemporal elasticity of substitution is low enough and of the top tail of the distribution thick enough, progressive taxation should be used to reduce all large fortunes down to a finite level. As a result, the long-run wealth distribution is truncated above and wealth inequality is drastically reduced.

There are a number of limitations in the model that should be emphasized. First, the infinite horizon model might not be a good representation of savings and wealth accumulation behavior. It is certainly not fully realistic to think that consumers can be so far-sighted. Moreover, the model requires everybody to have the same discount rate otherwise equilibria are degenerated. It is perhaps the case that the infinite horizon model predicts too large responses to capital income taxes. However, this feature should bias the results

[38] In the utilitarian case, we have $\bar{g}=0$ as described above.

[39] This result was noticed by Samuelson (1951), and rigorously established by Diamond and Mirrlees (1971).

[40] Importantly, a *regressive* capital income tax schedule will lead to divergence of wealth accumulation with no stable steady-state as the richest person ends up holding all wealth (and all others end up with no wealth).

[41] If $a^*\in(k',k)$, then all individuals end up with identical wealth a^* at the kink of the tax schedule.

[42] Piketty and Saez (2012) propose an optimal capital income tax theory in a model with heterogeneous discount rates across individuals and across dynasties that generates more realistic wealth accumulation patterns than the standard model considered in this paper.

against finding redistributive policies desirable.[43] It is therefore remarkable that the infinite horizon model can produce tax policy recommendations generating drastic wealth redistribution under some parameter values.

Second, in the model presented here, the initial unequal wealth distribution is given exogenously. As mentioned in Section 2, the obvious first-best policy would be to confiscate and redistribute wealth from the start once and for all either through wealth confiscation, infinitely large consumption taxes couples with infinitely large labor subsidies. There are perhaps political constraints preventing the government from applying such drastic policies. In that case, it is of interest to note that the effects of the optimal capital income taxes proposed here do not depend on the maximum tax rate that the government can set. In the historical record of tax policy development of western countries, wealth inequality inherited from the past and the large levels of the largest fortunes accumulated during the industrial revolutions was certainly one of the key arguments put forward by the proponents of progressive income taxation. Therefore, the analysis of limited wealth redistribution tools such as progressive capital income taxation (as opposed to direct wealth confiscation) is certainly relevant in practice.

Appendix A

Proof of Proposition 1

The denominator in Eq. (6), $1-(1-\sigma)e^{-\sigma\rho T}$, is between 1 and σ for any value of T, therefore $c_0\to+\infty$ when a_0 tends to infinity. The envelope theorem implies that the welfare effect is

$$\frac{\partial U(a_0,T)}{\partial T}=-u'(c_T)e^{-\rho T}\rho a_T=c_0^{-1/\sigma}\left[y-c_oe^{-\sigma\rho T}\right].$$

Using Eq. (7), the tax revenue effect is

$$\frac{\partial Tax(a_0,T)}{\partial T}=-\frac{\partial c_0}{\partial T}\cdot\frac{1+\sigma e^{-(\sigma+1)\rho T}}{\rho(1+\sigma)}+\sigma c_0e^{-(\sigma+1)\rho T}$$

Using these expressions and Eq. (9), we can rewrite the first order condition for the optimal $T(a_0)$ as

$$\begin{aligned}&\frac{c_0^{-1/\sigma}}{p}\left[y-c_0e^{-\sigma\rho T}\right]\\&+\frac{\sigma}{\sigma+1}\cdot\frac{1+\sigma e^{-(\sigma+1)\rho T}}{1-(1-\sigma)e^{-\sigma\rho T}}\left[-y+c_0e^{-\sigma\rho T}-\sigma c_0e^{-\sigma\rho T}\right]\\&+\sigma c_0e^{-(\sigma+1)\rho T}=0.\end{aligned}\tag{17}$$

The first term is the welfare effect and the last two terms are the tax revenue effects. As $c_0\to\infty$, the welfare effect is negligible relative to $[y-c_0e^{-\sigma\rho T}]$. This expression appears in the numerator of the second term of Eq. (17) multiplied by a factor bounded away from zero and infinity for all values of T. Therefore, the welfare effect is negligible in the asymptotic analysis of Eq. (17).

Case $\sigma<1$

In that situation, $c_0e^{-\sigma\rho T}$ must be bounded otherwise the bracketed expression of the second term in Eq. (17) takes arbitrarily large positive values (as y is constant) and the third term of Eq. (17) is also positive, implying that Eq. (17) cannot hold. Therefore $c_0e^{-\sigma\rho T}$ is bounded implying that $T\to\infty$ because $c_0\to\infty$. Thus the first term (welfare effect) and the third term in Eq. (17) both tend to zero. Therefore Eq. (17) holds only if the second term also converges to zero, that is, $(1-\sigma)c_0e^{-\sigma\rho T}\to y$, implying that $c_\infty=c_T\to y/(1-\sigma)$. As consumption and wealth are constant after T, we have $c_\infty=\rho a_\infty+y$, and thus $a_\infty(a_0)\to\sigma y/((1-\sigma)\rho)$ which proves Eq. (8).

A.1.2 . Case $\sigma>1$

In that situation, the behavioral response in c_0 unambiguously reduces tax revenue and thus the second term in Eq. (17) is negative and must be compensated by the positive third term in Eq. (17). In that case T must be bounded because otherwise the third term in Eq. (17) would be negligible relative to $c_0e^{-\sigma\rho T}$ and Eq. (17) could not hold. As T is bounded and as $c_0\to\infty$, the dominant terms proportional to c_0 in Eq. (17) must cancel each other, implying that:

$$\frac{(1-\sigma)e^{-\sigma\rho T}}{1+(\sigma-1)e^{-\sigma\rho T}}\cdot\frac{1+\sigma e^{-(\sigma+1)\rho T}}{\sigma+1}+e^{-(\sigma+1)\rho T}=0.$$

A simple analysis shows that this equation defines a unique T^∞ which must be the limit of $T(a_0)$ when a_0 grows to infinity. T^∞ decreases with σ and tends to infinity when σ decreases to one. Using Eqs. (5) and (6), it is then easy to obtain the asymptotic formula for $a_\infty(a_0)$. QED.

Proof of Proposition 2

The objective of the government is to choose the path (a_t^*) and b so as to maximize the sum of utilities subject to the budget constraint as described in Section 2.2. Let us assume that a_t^* is the optimal path for the exemption level. We assume that the tax rate above a_t^* is equal to the exogenous value $\tau=1$. The proof and results would be identical for any $\tau>0$ but the expressions would be greatly complicated.

As shown in the text when discussing Eq. (11), for a given path of exemption thresholds $(a_t^*)t\geq 0$, two equations define implicitly c_0 and T as a function of a_0:

$$c_0=\frac{\sigma\left[y+\rho(y\cdot T+\rho A_T^*+a_0)\right]}{1-(1-\sigma)e^{-\sigma\rho T}},\quad c_0e^{-\rho\sigma T}=\rho a_T^*+y.\tag{18}$$

We consider, as in the text, a small increase (or decrease) δa^* of a_t^* for $t\geq\bar{t}$. More precisely, as the post-reform exemption path must be non-decreasing, we assume that the derivative of the exemption path $a_t^{*'}$ is increased locally (between $\bar{t}-\delta\bar{t}$ and $\bar{t}$) by an amount $\delta a^{*'}$ such that $\delta a^{*'}\cdot\delta\bar{t}=\delta a^*$, effectively producing an increase δa^* in a_t^* for $t\geq\bar{t}$.[44] To save on notation, it is useful to define

$$v_t=\frac{a_t^{*'}}{\sigma(\rho a_t^*+y)}.\tag{19}$$

Differentiating the expressions in Eq. (18), and eliminating δT, we obtain:

$$\delta c_0=\frac{\sigma\rho\left(1+v_T\right)}{1+v_T-\left(1+v_T(1-\sigma)\right)e^{-\sigma\rho T}}\left[\rho(T-\bar{t})-\frac{1}{1+v_T}\right]\delta a^*.\tag{20}$$

Differentiating Eq. (12), we obtain

$$\delta Tax(a_0)=-\frac{\delta c_0}{\rho}\cdot\frac{1+v_T-(1-\sigma v_T)e^{-\rho(1+\sigma)T}}{(1+v_T)(1+\sigma)}-\frac{e^{-\rho T}}{1+v_T}\delta a^*.\tag{21}$$

[43] The Chamley–Judd results stating that optimal capital income taxes should be zero in the long-run have often been criticized on these grounds (see Piketty and Saez, 2012 for a discussion).

[44] In the case where $a_{\bar{t}}^{*'}=0$, it is impossible to decrease a_t^* uniformly above $\bar{t}$ and the constraint $a_{\bar{t}}^{*'}\geq 0$ binds.

Using the envelope theorem, the effect of the reform on utility $U(a_0)$ of individual with initial wealth a_0 is given by

$$\delta U(a_0) = \int_{\bar{t}}^{T} u'(c_t)e^{-\rho t}\rho\delta a^* dt = \delta a^* u'(c_0)\rho(T-\bar{t}). \quad (22)$$

Consider now the asymptotic analysis $\bar{t}\to\infty$. In that case, $T\to\infty$ and we assume first that v_T converges to $\bar{v}$. We denote by $o(1)$ a quantity converging to zero when $\bar{t}\to\infty$. Eq. (20) can be rewritten as

$$\delta c_0 = \delta a^*\rho\sigma\left[\rho(T-\bar{t}) - \frac{1}{1+\bar{v}} + o(1)\right]. \quad (23)$$

We now change variables from T to a_0. Using Eq. (18), for T large, $c_0 = \sigma\rho a_0(1+o(1))$. This can be seen as follows. For T large, Eq. (18) implies $c_0 = \sigma\rho[a_0 + T(\rho A_T{}^*/T + y)](1+o(1))$ and $c_0 = e^{\rho\sigma T}(\rho a_T{}^* + y)$. Hence c_0 grows exponentially with T. This implies therefore that $T(\rho A_T{}^*/T + y) \ll a_0$, and hence $c_0 = \sigma\rho a_0(1+o(1))$.

Therefore, using $c_0 e^{-\rho\sigma T} = \rho a_T{}^* + y$, we have

$$\sigma\rho T = \log a_0 + \log(\rho\sigma) - \log(y + \rho a_T^*) + o(1). \quad (24)$$

Integrating Eq. (19) from $\bar{t}$ to T, we have $\log(y+\rho a_T^*) - \log(y+\rho a_{\bar{t}}^*) = \rho\sigma(T-\bar{t})[\bar{v}+o(1)]$. Hence, taking the difference of Eq. (24) for T and $\bar{t}$ (corresponding to wealth levels a_0 and $\bar{a}_0$ respectively), we have

$$\sigma\rho(T-\bar{t}) = \frac{1}{1+\bar{v}+o(1)}\log\left(\frac{a_0}{\bar{a}_0}\right) + o(1). \quad (25)$$

Therefore, we can rewrite Eq. (23) as

$$\delta c_0 = \delta a^*\rho\left[\frac{1}{1+\bar{v}+o(1)}\log\left(\frac{a_0}{\bar{a}_0}\right) - \frac{\sigma}{1+\bar{v}} + o(1)\right]. \quad (26)$$

For large T and $\bar{t}$, using Eqs. (21) and (22), we have the following approximation formulas for the change in tax revenue and welfare

$$\delta Tax(a_0) = \frac{\delta a^*}{1+\sigma}\left[\frac{\sigma}{1+\bar{v}} - \frac{1}{1+\bar{v}+o(1)}\log\left(\frac{a_0}{\bar{a}_0}\right) + o(1)\right], \quad (27)$$

$$\delta U(a_0) = \delta a^*\frac{c_0^{-\frac{1}{\sigma}}}{\sigma}\left[\frac{1}{1+\bar{v}+o(1)}\log\left(\frac{a_0}{\bar{a}_0}\right) + o(1)\right]. \quad (28)$$

As $c_0\to\infty$ when $a_0\to\infty$, asymptotically, Eqs. (27) and (28) show that the welfare effect $\delta U(a_0)$ is negligible relative to the tax effect $\delta Tax(a_0)$ and can be ignored in the asymptotic analysis.

Assuming that a_0 is Pareto distributed in the tail with parameter α, a simple integration of Eq. (27) from $\bar{a}_0$ to infinity implies that the total effect on tax revenue is given by

$$\delta Tax = \delta a^* \cdot \frac{1}{(1+\sigma)(1+\bar{v})}\cdot\left[\sigma - \frac{1}{\alpha} + o(1)\right]\cdot[1-H(\bar{a}_0)]. \quad (29)$$

- If $\sigma\alpha<1$, then Eq. (29) implies that decreasing $a_t{}^*$ increases tax revenue. Therefore, it must be the case that the constraint $a_t^{*'}\geq 0$ is binding asymptotically, meaning that $a_t{}^*$ is constant for t large enough which proves the first part of Proposition 2.
- If $\sigma\alpha>1$, then Eq. (29) implies that increasing $a_t{}^*$ increases tax revenue. As it is always possible to increase $a_t{}^*$, it must be the case that v_t is not converging to a finite value but diverging to infinity. In that case, integrating Eq. (19) implies that $\sigma\rho T/\log(y+\rho a_T{}^*) = o(1)$. Therefore Eq. (24) implies $\log(\rho a_T{}^* + y) = (1+o(1))\log(a_0)$, and hence $\log(a_\infty(a_0)) = (1+o(1))\log(a_0)$. Therefore, the asymptotic wealth distribution is also Pareto distributed with parameter α. QED.

General welfare function

Wealth specific tax

With the general welfare function, the first term (corresponding to the welfare effect) in the first order condition (17) must be replaced by $\beta(a_0)(y - c_0e^{-\sigma\rho T}) = g(a_0)(1 - dTax(a_0)/da_0)(y - c_0e^{-\sigma\rho T})$. Using Eq. (4), we have $1 - dTax(a_0)/da_0 = (\sigma/(1+\sigma))\cdot(1+\sigma e^{-(\sigma+1)\rho T})/(1-(1-\sigma)e^{-\sigma\rho T})$. Therefore the first order condition (17) becomes:

$$\frac{\sigma}{\sigma+1}\cdot\frac{1+\sigma e^{-(\sigma+1)\rho T}}{1-(1-\sigma)e^{-\sigma\rho T}}\left[(1-g(a_0))\left(-y+c_0e^{-\sigma\rho T}\right) - \sigma c_0 e^{-\sigma\rho T}\right] + \sigma c_0 e^{-(\sigma+1)\rho T} = 0. \quad (30)$$

The remaining of the proof parallels the proof of Proposition 1. The two cases to be distinguished are $\sigma<1-\bar{g}$ and $\sigma>1-\bar{g}$. In the former, we have $(1-\bar{g}-\sigma)c_0e^{-\sigma\rho T}\to(1-\bar{g})y$, and hence $\rho a_\infty(a_0)\to y\cdot\sigma/(1-\bar{g}-\sigma)$, as stated in Proposition 3.

Progressive income tax

In that case, routine but tedious computations show that

$$1 - dTax(a_0)/da_0 = \frac{\sigma}{1+\sigma}\cdot\frac{1+v-(1-\sigma v)e^{-\rho(1+\sigma)T}}{1+v-(1-(\sigma-1)v)e^{-\rho\sigma T}} \to \frac{\sigma}{1+\sigma}.$$

Therefore, adding the welfare effect $\delta W(a_0) = G'(U(a_0))\delta U(a_0)/p$ to the tax effect $\delta Tax(a_0)$, and using Eqs. (27) and (28), we obtain

$$\delta W(a_0) + \delta Tax(a_0) = \frac{\delta a^*}{1+\sigma}\left[\frac{\sigma}{1+\bar{v}} - \frac{1}{1+\bar{v}+o(1)}\log\left(\frac{a_0}{\bar{a}_0}\right)(1-\bar{g}+o(1)) + o(1)\right]. \quad (31)$$

Therefore, integrating over the population with $a_0\geq\bar{a}_0$ as in Proposition 2, the total welfare and tax revenue effect is

$$\delta W + \delta Tax = \frac{\delta a^*}{(1+\sigma)(1+\bar{v})}\left[\sigma - \frac{1-\bar{g}}{\alpha} + o(1)\right]\cdot[1-H(\bar{a}_0)]. \quad (32)$$

Therefore the same analysis as in Proposition 2 applies and the two cases to be distinguished are $\sigma\alpha<1-\bar{g}$ and $\sigma\alpha>1-\bar{g}$. QED.

References

Aiyagari, S.R., 1995. Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting. Journal of Political Economy 103, 1158–1175.

Atkeson, Andrew, Chari, V.V., Kehoe, Patrick, 1999. Taxing capital income: a bad idea. Federal Reserve Bank of Minneapolis Quarterly Review 3–17 (Summer).

Atkinson, Anthony B., Sandmo, Agnar, 1980. Welfare implications of the taxation of savings. The Economic Journal 90, 529–549.

Atkinson, Anthony B., Stiglitz, Joseph E., 1976. The design of the tax structure: direct versus indirect taxation. Journal of Public Economics 6, 55–75.

Atkinson, Anthony B., Piketty, Thomas, Saez, Emmanuel, 2011. Top incomes in the long run of history. Journal of Economic Literature 49 (1), 3–71.

Attanasio, Orazio, Weber, Guglielmo, 1995. Is consumption growth consistent with intertemporal optimization? Journal of Political Economy 103, 1121–1157.

Auerbach, Alan, Kotlikoff, Lawrence, 1987. Dynamic Fiscal Policy. Cambridge University Press, Cambridge.

Bernheim, Douglas, 2002. Taxation and saving. In: Auerbach, A., Feldstein, M. (Eds.), Handbook of Public Economics, vol. 3. North Holland, Amsterdam.

Brownlee, Eliott, 2000. Historical perspective on U.S. tax policy toward the rich. In: Slemrod, J. (Ed.), Does Atlas Shrug? The Economics Consequences of Taxing the Rich. Harvard University Press, Cambridge.

Cagetti, Marco, De Nardi, Cristina, 2006. Entrepreneurship, frictions and wealth. Journal of Political Economy 106 (5), 835–870.

Chamley, Christophe P., 1986. Optimal taxation of capital income in general equilibrium with infinite lives. Econometrica 54, 607–622.

Chamley, Christophe P., 2001. Capital income taxation, wealth distribution and borrowing constraints. Journal of Public Economics 79, 55–69.

Champernowne, D., 1953. A model of income distribution. The Economic Journal 63, 318–351.

Chari, V.V., Kehoe, Patrick, 1999. Optimal fiscal and monetary policy. In: Taylor, J., Woodford, M. (Eds.), Handbook of Macroeconomics. Elsevier, New York.

Cremer, Helmuth, Pestieau, Pierre, 2004. Wealth transfer taxation: a survey of the theoretical literature. In: Kolm, S.C., Mercier Ythier, J. (Eds.), Handbook on Altruism, Giving and Reciprocity. Amsterdam, North Holland.

Deaton, Angus, 1992. Understanding Consumption. Clarendon Press, Oxford.

Diamond, Peter A., 1998. Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates. American Economic Review 88, 83–95.

Diamond, Peter A., Mirrlees, James, 1971. Optimal taxation and public production. American Economic Review 61 (8–27), 261–278.

Diamond, Peter, Saez, Emmanuel, 2011. The case for a progressive tax: from basic research to policy recommendations. Journal of Economic Perspectives 25 (4), 165–190.

Feldstein, Martin, 1978. The welfare cost of capital income taxation. Journal of Political Economy 86, S29–S52.

Gabaix, Xavier, 2009. Power laws in economics and finance. Annual Review of Economics 1 (1), 255–294.

Golosov, Michael, Tsyvinski, Aleh, Werning, Ivan, 2006. New dynamic public finance: a user's guide. NBER Macroeconomics Annual.

Hall, Robert, 1988. Intertemporal substitution in consumption. Journal of Political Economy 96, 339–357.

Judd, Kenneth, 1985. Redistributive taxation in a simple perfect foresight model. Journal of Public Economics 28, 59–83.

King, Mervyn, 1980. Savings and taxation. In: Hughes, G., Heal, G. (Eds.), Public Policy and the Tax System. George Allen Unwin, London, pp. 1–36.

Kocherlakota, Narayana, 2010. The New Dynamic Public Finance. Princeton University Press, Princeton.

Kotlikoff, Lawrence, Summers, Lawrence, 1981. The role of intergenerational transfers in aggregate capital accumulation. Journal of Political Economy 89, 706–732.

Lucas Jr., Robert E., 1990. Supply-side economics: an analytical review. Oxford Economic Papers 42, 293–316.

Mirrlees, James A., 1971. An exploration in the theory of optimal income taxation. Review of Economic Studies 38, 175–208.

Ordover, J., Phelps, Edmund, 1979. The concept of optimal taxation in the overlapping generations model of capital and wealth. Journal of Public Economics 12, 1–26.

Pareto, Vilfredo, 1896. Cours d'Economie Politique. Droz, Geneva, Switzerland.

Piketty, Thomas, 2001a. Income inequality in France, 1901–1998. CEPR Discussion Paper No. 2876.

Piketty, Thomas, 2001b. Les hauts revenus en France au 20eme siècle — Inégalités et redistributions, 1901–1998. Editions Grasset, Paris.

Piketty, Thomas, Saez, Emmanuel, 2003. Income inequality in the United States, 1913–1998. Quarterly Journal of Economics 118 (1), 1–39.

Piketty, Thomas, Saez, Emmanuel, 2012. A theory of optimal capital taxation. NBER Working Paper, No. 17989.

Saez, Emmanuel, 2001. Using elasticities to derive optimal income tax rates. Review of Economic Studies 68, 205–229.

Samuelson, Paul A., 1951. Theory of optimal taxation. Memorandum to the U.S. Treasury, published in Journal of Public Economics 30, 137–143 (1986).