Optimal Progressive Capital Income Taxes in the Infinite Horizon Model

Emmanuel Saez

Harvard University and NBER*

October 15, 2001

First Draft (incomplete)

Abstract

This paper analyzes optimal progressive capital income taxation in the infinite horizon dynastic model. It shows that progressive taxation is a much more powerful and useful tool to redistribute wealth than linear taxation on which previous literature has focused. 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. 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. A number of numerical simulations illustrate the theoretical result. (*JEL* H21, H62)

*Harvard University, Department of Economics, Littauer, Cambridge MA02138. email:saez@fas.harvard.edu.
This paper owes much to stimulating discussions with Thomas Piketty.

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 U.S. 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 U.K. has imposed schedular income taxes since 1842 and introduced a progressive super-tax on comprehensive individual income in 1909. The common characteristic of these early income tax systems is that they had large exemption levels and thus hit only the top of the income distribution. While tax rates were initially set at low levels (in general below 10%), 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 often above 60-70%. In most cases, the very top rates applied only to an extremely small fraction of taxpayers.[1] Therefore, the income tax was devised to have its strongest impact on the very top income earners. As documented by Piketty (2001) for France, and Piketty and Saez (2001) for the U.S., these top income earners derived the vast majority of their income in the form of capital income (mostly dividends and to a lesser extent capital gains).[2] 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. Most countries have also introduced graduated forms of estate or inheritance taxation that further increase the degree of progressivity of taxation. Such a progressive tax structure should have a strong wealth equalizing effect.[3]

A central question in tax policy analysis is whether using capital income taxation to achieve redistribution of wealth is desirable. As in most tax policy problems, there is a classical equity and efficiency trade-off: capital income taxes should be used to redistribute wealth only if the efficiency cost of doing so is not too large. A number of studies on optimal dynamic taxation have suggested that capital taxation might have very large efficiency costs (see e.g., Lucas (1990),

[1]For example, in the U.S., in the 1930s, the top bracket was for incomes above \$5,000,000 (in current dollars). Unsurprisingly, but a handful of taxpayers had incomes large enough to be in the top bracket in any given year.

[2]This is still true in France today but no longer in the U.S. where highly compensated executives have replaced rentiers at the top of the income distribution.

[3]Indeed Piketty (2001) and Piketty and Saez (2001) 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 U.S.

and Chari et al. (1997)). In the infinite horizon dynastic model, linear capital income taxes generate distortions increasing without bound with time. The influential studies by Chamley (1986) and Judd (1985) show that, in the long-run, optimal linear capital income tax should be zero. Therefore, the predictions coming out of these optimal dynamic taxation models is much at odds with the historical and even current record of actual tax practices in most developed countries.[4]

This paper argues that capital income taxes can be a very powerful and desirable tool to redistribute wealth. The critical departure from the literature that grew out of the seminal work of Chamley (1986) and Judd (1985) considered here is that, in accordance with actual income and estate tax policy practice, I consider non-linear capital income taxation. Progressive capital income taxation is much more effective than linear taxation to redistribute wealth. Under realistic assumptions for the intertemporal elasticity of substitution, with optimal non-linear taxation, even if the initial wealth distribution is unbounded, the optimal 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 to a given threshold. If the initial wealth distribution is unbounded, at any time, there are still some individuals who continue to be taxed and therefore, strictly speaking, the tax is never zero. Therefore the policy prescriptions that are obtained from the model developed here are well in line with the historical practice. Introducing a steeply progressive capital income tax does not introduce large efficiency costs and is very effective in reducing the concentration of capital income, as in the historical experience of France and the U.S.[5] Whether the progressive income taxes and very high top tax rates enacted in most OECD

[4]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), and 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) and 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 part of wealth, especially for the rich, is due to bequests (see Kotlikoff and Summers (1981)). The OLG model therefore is not well suited to the analysis of the taxation of large fortunes. I come back to this issue in conclusion.

[5]As mentioned above, the revival of income inequality in the last three decades in the U.S. is a labor income (and not a capital income) phenomenon.

countries during the 1930s and in place at least until the 1970s have had negative effects on economic activity is a controversial issue (see e.g., Slemrod (2000)). This tax policy apparently did not prevent developed countries from growing very quickly in the post World War II period.

The mechanism explaining why progressive taxation is desirable can be understood as follows. In the dynastic model, linear taxation of capital income is undesirable because it introduces a price distortion exponentially increasing with time. That is why, at the optimum, linear capital income taxation must be zero in the long-run. However, if one considers a simple progressive tax structure with a single marginal tax rate above a given exemption threshold, then large wealth holders will be in the tax bracket and therefore face a lower net-of-tax rate of return than lower wealth holders (that are in the exempted bracket). As a result, the infinite horizon dynastic 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 efficient in general for the following reason. For large wealth holders, the price distortion induced by capital income taxation generates a relatively smaller negative human wealth effect on wealth accumulation than for poorer taxpayers because the consumption stream of the wealthy is large relative to their labor income stream. As a result, the rich 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 at time passes producing an unbounded long-run wealth distribution.

[6] Piketty (2001) 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 capital stock in the economy because the reduction of large fortunes is compensated by an increase of smaller fortunes. This, of course, is not true with linear capital income taxation.

The paper is organized as follows. Section 2 presents the model and the government objective. Section 3 considers wealth specific linear taxation and provides useful preliminary results on the desirability of taxing richer individuals longer. Section 4 introduces progressive capital income taxation and derives the key theoretical results. Section 5 proposes some numerical simulation to illustrate the results and discusses policy implications. Section 6 analyzes how relaxing the simplifying assumptions of the basic model affects the results. Finally, Section 7 offers some concluding comments.

2 The General Model

2.1 Individual program

We consider a standard dynastic model with no uncertainty and perfectly competitive markets. All individuals have the same instantaneous utility function with constant intertemporal elasticity of substitution σ

$$u(c) = \frac{c^{1-1/\sigma}}{1-1/\sigma}. \tag{1}$$

When $\sigma = 1$, we have of course $u(c) = \log c$. All individuals discount the future at rate $\rho > 0$ and maximize the intertemporal utility

$$U = \int_0^\infty u(c_t) e^{-\rho t} dt \tag{2}$$

subject to the budget constraint

$$\dot{a}_t = r_t a_t - I_t(r_t a_t) + y_t - c_t \tag{3}$$

where a_t denotes wealth, r_t is the interest rate, $I_t(.)$ is the capital income tax (possibly non-linear, and time varying), y_t is instantaneous income equal to wage income w_t plus government lumpsum benefits b_t. The individual starts with exogenous initial wealth a_0. Utility maximization leads to the usual Euler equation

$$\frac{\dot{c}_t}{c_t} = \sigma[r_t(1 - I_t'(r_t a_t)) - \rho]. \tag{4}$$

Equations (3) and (4) combined with the initial condition $a(0) = a_0$, and the transversality condition define a unique optimal path of consumption and wealth. It is important to note that the intertemporal elasticity of substitution measures the sensitivity of the consumption pattern with respect to the net-of-tax interest rate. A higher marginal tax rate shifts consumption away from later periods to earlier periods.

We assume that all individuals earn the same wage and differ only through their initial wealth endowment a_0.[7] The population is normalized to one and the cumulated distribution of wealth is denoted by $H(a_0)$, and the density by $h(a_0)$. The support of the wealth distribution is denoted by A_0. 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 path of government benefits (b_t).

The derivation of optimal capital taxes relies critically on the behavioral responses to taxation and the induced effect on wealth accumulation. Before providing a more general analysis, it is useful to focus on the particular case where the interest r is exogenous and equal to the discount rate ρ, the annual income stream y_t is constant (equal to y).

In this situation, with no taxation $I_t(.) = 0$, the Euler equation (4) implies that the path of consumption is constant ($c_t = c_0$ for all t). The budget constraint (3) becomes $\dot{a}_t = \rho a_t + y - c$. As both y and c are constant overtime, the transversality condition can be satisfied only if wealth a_t is constant overtime and thus equal to a_0 (otherwise, wealth would grow exponentially). Consumption is equal to wage income plus interest income on wealth ($c = y + \rho a_0$). Therefore, the wealth distribution remains constant over time and equal to the initial wealth distribution $H(a_0)$.

From the Euler equation (4), we see that introducing positive marginal tax rates produces a decreasing pattern of consumption over time. 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.

In the general case, I denote by $\bar{r}_t = r_t(1 - I_t'(r_r a_t))$ the net-of-tax instantaneous interest

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

rate, by $R_t = \int_0^t r_s ds$ the cumulated pre-tax interest rate up to time t, and by $\bar{R}_t = \int_0^t \bar{r}_s ds$ the cumulated after-tax interest rate. Finally, as in the labor supply literature with non-linear taxes, it is useful to consider the linearized budget constraint defined as the tangent of the actual non-linear tax schedule ($r_t a_t \to r_t a_t - I(r_t a_t)$) at the point $r_t a_t$. The slope of this linearized budget is obviously $1 - I_t'(r_t a_t)$ and the intercept with the y-axis is defined as the virtual income. The virtual income is denoted by m_t and is equal to $m_t = r_t a_t I_t'(r_t a_t) - I_t(r_t a_t)$. As lumpsum payments b_t are included in annual income y_t, we can adopt the normalization $I_t(0) = 0$; that is, taxes are zero for individuals with no capital income. The wealth accumulation equation (3) can be simply rewritten as

$$\dot{a}_t = \bar{r}_t a_t + m_t + y_t - c_t. \tag{5}$$

Integrating this equation and using the transversality equation, we obtain

$$\int_0^\infty c_t e^{-\bar{R}_t} dt = a_0 + \int_0^\infty [y_t + m_t] e^{-\bar{R}_t} dt. \tag{6}$$

Equation (6) will be of much use. It simply states that the discounted stream (using the net-of-tax rate of return) of consumption must be equal to initial wealth a_0 plus the discounted stream of annual income y_t plus virtual income m_t. Thus, the price of consumption (or income) at time t faced by the individual is $q_t = e^{-\bar{R}_t}$. The pre-tax price is obviously $p_t = e^{-R_t}$. As is well known, a constant tax rate over-time introduces an exponentially growing price distortion. The Euler equation (4) can also be integrated to obtain $c_t = c_0 e^{\sigma(\bar{R}_t - \rho t)}$. Plugging this expression in (6), we obtain

$$c_0 = \frac{a_0 + \int_0^\infty [y_t + m_t] e^{-\bar{R}_t} dt}{\int_0^\infty e^{\sigma(\bar{R}_t - \rho t) - \bar{R}_t} dt}. \tag{7}$$

Let us study the effects of capital taxes on the pattern of consumption c_0. Suppose that the marginal tax rate I_t' is increased for a small period of time at time t (assuming no change for the moment in virtual income), then $\bar{R}_s$ is reduced for all $s > t$, and thus the price of consumption $q_s = e^{-\bar{R}_s}$ is increased for all $s > t$. As is well known, there are three effects on the pattern of consumption c_0. These three effects correspond to the three occurences of $\bar{R}_t$ in equation (7): one in the numerator and two in the denominator. First, there is a substitution effect (this is the first $\bar{R}_t$ term in the denominator). The price of consumption after time t increases relative to the price of consumption before time t and thus the individual shifts consumption earlier in time by increasing c_0. This substitution effect is increasing with σ. Second, there is a negative income effect (second $\bar{R}_t$ term in the denominator). The price of consumption after time t is increased and thus the individual has to reduce its consumption level in general. This effect decreases c_0. As usual, and as can be seen in (7), when $\sigma = 1$, income and substitution effects exactly cancel out. Third, there is a positive human wealth effect ($\bar{R}_t$ term in the numerator), the price of the income stream $y_t + m_t$ is increased after time t and this allows the individual to increase c_0. In general, an increase in the marginal tax rate also increases the virtual income m_t, which produces an additional positive human wealth effect and increases c_0.[8] From now on, we will call this latter effect, the virtual income effect.

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

$$Tax(a_0) = \int_0^{\infty} I_t(r_t a_t) e^{-R_t} dt. \tag{8}$$

Integrating equation (3), and using the transversality condition, one obtains that taxes collected are also equal to the difference between initial wealth a_0 plus the discounted value of the income stream y_t and the discounted value of the consumption stream c_t. As a result, we have

$$Tax(a_0) = a_0 + \int_0^{\infty} [y_t - c_t] e^{-R_t} dt = a_0 + \int_0^{\infty} y_t e^{-R_t} dt - c_0 \int_0^{\infty} e^{\sigma(\bar{R}_t - \rho t) - R_t} dt. \tag{9}$$

This equation shows clearly how a behavioral response in c_0 due to a tax change triggers a change in tax revenue collected. A large c_0 (consequence of high marginal tax rates and a distorted consumption pattern), implies a lower level of taxes collected. In principle, the change in tax revenue triggered by a small tax reform can be decomposed into a mechanical effect (change in tax revenue if there were no behavioral response), and a behavioral effect (change in tax revenue due to the behavioral response). In the present model, however, this important conceptual distinction does not provide the simplest way to derive our results. It turns out that using

[8]In the special case where the tax is linear with constant marginal rate τ, changing τ has no effect on virtual income.

formula (9) where the behavioral and mechanical response do not appear separately explicitly is simpler. However, to provide the economic intuition behind the proofs, we will see that it is useful to come back to the distinction between mechanical and behavioral responses.

Finally, using the expression for c_t, the total discounted utility U of the individual can be rewritten as

$$U = u(c_0) \int_0^\infty e^{(\sigma-1)\bar{R}_t - \sigma \rho t} dt. \tag{10}$$

To simplify significantly the presentation, we make the following assumption:

Assumption 1 *The real interest rate is exogenous and constantly equal to the discount rate ρ, the wage is exogenous and constantly equal to a given value w.*

We show in Section 6 how assumption 1 can be relaxed without affecting the results.

2.2 Government Program

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

$$W = \int_{A_0} U(a_0) dH(a_0) \tag{11}$$

subject to the budget constraint

$$\int_{A_0} Tax(a_0) dH(a_0) \geq B + G \tag{12}$$

where B and G denote the present discounted value (at pre-tax interest rates) of government benefits b_t and exogenous spending g_t. The budget constraint states that the present discounted value of 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 (12). It is possible to extend the analysis to more general social welfare functions than the utilitarian welfare function described above. However, as most results are independent of the social welfare criterion, to keep the presentation simple, it is preferable to focus on the utilitarian case.

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. This wealth levy is first-best Pareto efficient. As redistribution with a wealth levy entails no efficiency costs, a government would redistribute wealth so as to equalize perfectly marginal utilities. This implies that consumption and wealth levels after the levy are equal across individuals.[9]

In the analysis that follows, we assume, as in the literature, that the government cannot implement this 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 implement, 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 2 *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 almost never and only in very extraordinary situations such as wars, or after-war periods.[10] The political debates preceding the introduction of progressive income taxes in the U.K. in 1909, France in 1914, or the U.S. in 1913 provide interesting evidence on these issues. Parties from the left were the promoters of progressive income taxation for redistributive reasons and to curb the largest wealth holdings. Fierce opposition for the right prevented the implementation of more drastic redistributive policies such as wealth levies, and that is why, in most cases, the initial income tax systems started with relatively low top marginal tax rates.

We make the following additional simplification assumption:

Assumption 3 *The path of government lumpsum grants b_t is restricted to be constant overtime.*

[9]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.

[10]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.

Assumption 3 requires some explanations. Implicit in equation (12) 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, individuals will typically face different after-tax interest rates and debt is no longer neutral and can be used to improve welfare. We will discuss in detail in Section 6 how debt can be used in conjunction with non-linear taxes to improve redistribution. Assumption 3 is a way to freeze the debt instrument by forcing the government to redistribute tax proceeds uniformly over time.

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. We will in the following section how the insights that we obtain in that (unrealistic) situation are useful to analyze the desirability of progressive capital income taxation.

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, we consider the following policy: the government imposes the maximum tax rate τ on capital income up to a time T and zero taxation afterwards. We show later on that this "bang-bang" pattern of taxation is optimal in the models we consider.[11] For notational simplicity and without loss of generality, we assume that $\tau = 1$, that is, the maximum rate is 100%.

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, we have $m_t = 0$ because the tax is linear, $\bar{R}_t = 0$ if $t \leq T$ and $\bar{R}_t = \rho(t - T)$ if $t \geq T$. After

[11] Chamley (1986) was the first to prove that this type of policy is optimal for a wide class of dynamic models.

time T, the Euler equation (4) implies that $\dot{c}_t = 0$, and thus constant consumption $c_t = c_T$. With assumptions 1 and 3, $y_t = w_t + b_t$ is also constant overtime (equal to $y = w + b$), the wealth equation becomes $\dot{a}_t = \rho a_t + y - c_T$. This equation has a unique solution $a_t = (c_T - y)/\rho$ (constant path of wealth after T) that is compatible with the transversality condition.

Before time T, the Euler equation implies $\dot{c}/c = -\sigma\rho$, and therefore $c_t = c_0 e^{-\sigma\rho t}$. The wealth equation implies $\dot{a}_t = y - c_t$, and therefore using the initial condition for wealth, we have

$$a_t = a_0 + y \cdot t - \frac{c_0}{\sigma\rho}\left(1 - e^{-\sigma\rho t}\right). \tag{13}$$

There is a unique value c_0 such that the path for wealth (13) for $t = T$ matches the constant path of wealth $a_T = (c_0 e^{-\sigma\rho T} - y)/\rho$ after T. Using equation (7), this unique value c_0 is such that

$$c_0 = \frac{\sigma[y + \rho(y \cdot T + a_0)]}{1 - (1-\sigma)e^{-\sigma\rho T}}. \tag{14}$$

We denote by $a_\infty(a_0)$ and $c_\infty(a_0)$ the (constant) values of wealth and consumption after time T. Using (13) and (14), we obtain:

$$a_\infty(a_0) = a_0 + y \cdot T - \frac{y \cdot T + a_0 + y/\rho}{1 - (1-\sigma)e^{-\sigma\rho T}}\left(1 - e^{-\sigma\rho T}\right). \tag{15}$$

Using equation (9), the present discounted value of total capital income taxes collected is

$$Tax(a_0, T) = \int_0^T \rho a_t e^{-\rho t} dt = \frac{y}{\rho} + a_0 - \frac{c_0}{\rho} \cdot \frac{1 + \sigma e^{-(\sigma+1)\rho T}}{1 + \sigma} \tag{16}$$

and using (10), the total utility of the individual is

$$U(a_0, T) = u(c_0) \cdot \frac{1 - (1-\sigma)e^{-\sigma\rho T}}{\sigma\rho}. \tag{17}$$

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 the 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

$$L = \int_{A_0} U(a_0) dH(a_0) + p \left[\int_{A_0} Tax(a_0) dH(a_0) - (b+g)/\rho \right],$$

and taking the first order conditions with respect to b and T.

Presumably, the optimal time span of taxation T depends positively on exogenous revenue requirements g. 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 equation (15) for large values of a_0, we see 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 divided by a proportional factor, but the shape of the top tail on 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 α. However, 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, does not seem to have been investigated by the literature.

In that model, it is straightforward to check that the pattern of government benefits b_t has no effect on the final allocation.[12] To see this, suppose that the government modifies the pattern of benefits b_t to b'_t so as to keep the budget constraint of the individual unchanged: $\int b_t e^{-\bar{R}_t} dt = \int b'_t e^{-\bar{R}_t} dt$. Then the consumption decision c_t of the individual in unaffected. As the income stream w_t is also unchanged, equation (9) shows total taxes collected net of benefits b'_t are also unchanged, and thus the government budget constraint is also satisfied. Therefore, changing the stream b_t has no real effect on the economy and thus debt policy cannot affect the real outcomes.

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 dynastic model. As a direct extension of the

[12] Chamley (1986) made this point.

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.[13]

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, and again with no effect on the key results, we assume that $\tau = 1$. In this context, the government chooses the optimal set of time periods $T(a_0)$, and benefits levels b that maximize social welfare (11) subject to the budget constraint (12). The first order condition with respect to $T(a_0)$ is

$$\frac{\partial U(a_0)}{\partial T(a_0)} + p\frac{\partial Tax(a_0)}{\partial T(a_0)} = 0. \tag{18}$$

The first order condition (18) has a straightforward interpretation: 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. From equation (9), we see that it is critical to analyze the effect of T on c_0 to assess the effect of increasing T on tax revenue $Tax(a_0)$. Using equation (14), the effect of an extra time of taxation dT on c_0 is given by

$$\frac{\partial c_0}{\partial T} = \sigma\rho \cdot \frac{y - c_0 e^{-\sigma\rho T} + \sigma c_0 e^{-\sigma\rho T}}{1-(1-\sigma)e^{-\sigma\rho T}}. \tag{19}$$

Therefore, as displayed in the numerator of (19) and paralleling the analysis of equation (7), the marginal effect of T on c_0 can be decomposed into three effects. The first term in the numerator of equation (19) is the human wealth effect and is always positive because $y = w + b > 0$. When the time of taxation increases, the present discounted value of the income stream y increases and thus consumption goes up. Note that the human wealth effect 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

[13] The proof is given in appendix.

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. We can now state our first proposition.

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

$$T(a_0) \sim \frac{1}{\sigma\rho} \log a_0, \tag{20}$$

$$a_\infty(a_0) \to \frac{\sigma}{1-\sigma} \cdot \frac{y}{\rho}. \tag{21}$$

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^∞,*

$$a_\infty(a_0) \sim a_0 \cdot \frac{\sigma e^{-\sigma\rho T^\infty}}{1 + (\sigma - 1)e^{-\sigma\rho T^\infty}}. \tag{22}$$

- *If $\sigma = 1$, then asymptotically (i.e., for large a_0)*

$$T(a_0) \sim \frac{1}{2\rho} \log a_0, \tag{23}$$

$$a_\infty(a_0) \sim \sqrt{\frac{a_0 y}{2\rho}}. \tag{24}$$

The proof of Proposition 1 can be obtained by analyzing the first order condition (18) for large a_0. The technical proof is provided in appendix. If the maximum tax rate were any $\tau > 0$ (instead of 1), the time of taxation in (20) would be multiplied by a factor $1/\tau$, but equation (21) on the final wealth level would be identical. Similarly, the qualitative results for the cases $\sigma > 1$, and $\sigma = 1$ would be unchanged.

It is worth describing in detail the intuition for these results. When $\sigma > 1$, increasing the time of taxation T by dT produces a negative substitution effect on tax revenue that dominates the income income effect. As the wealth effect is also negative, increasing T unambiguously produces a reduction in tax revenue through the behavioral response in c_0. As can be seen from equation (19), the effect on c_0 is on the order of dT, and thus, as can be seen from equation (16), the effect on taxes collected is also on the order of dT.

As 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 dwarves the mechanical increase in tax revenue unless T is small. 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. That is the only way the mechanical increase in tax revenue can compensate the large behavioral response to taxation.

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, capital income taxes are not implemented (even for the largest fortunes) beyond a finite time T^∞. In that sense, capital income taxation is really zero after time T^∞ in spite of the fact that some individuals may still own very large fortunes. As in the Chamley (1986) uniform linear tax situation described in Section 3.2, the resulting wealth distribution is not drastically affected by optimal capital taxation.

When $\sigma < 1$, increasing T may increase tax revenue through the behavioral response because the substitution effect dominates the income effect. For large a_0, initial consumption c_0 is large relative to y (because the capital income stream dwarves the annual income stream y and allows the individual to sustain a much higher consumption level). As can be seen from equation (19), unless T is large, the substitution effect (net of the income effect) is going to dominate the human wealth effect, and therefore the response in c_0 is going to be negative, generating more tax revenue (equation (16)). 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.[14] Therefore, using the denominator of (19), T must be such that $(1-\sigma)c_0 e^{-\sigma\rho T} \approx y$,

[14] One can check that, for large a_0, the welfare effect is small relative the increase in tax revenue. Thus the optimal time of taxation in that case is such that the behavioral response of the consumption plan c_0 to an extra-time of taxation is zero. Therefore, the time of taxation for large wealth owners is set such as to extract the maximum amount of tax revenue, and thus corresponds to the top of the Laffer curve. This shows that the rule that richer individuals should be taxed longer does not depend on redistributive considerations but only on

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 (21).

Therefore when the elasticity of substitution σ is below unity, the government would like to tax larger fortunes longer until they are reduced to a finite threshold given in (21). If the initial wealth distribution is bounded above, then it is true that taxation is zero in the long run (after time $T(\max(a_0))$). But if the 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 $\sigma = 1/2$ (not an unrealistic value, see below), the largest fortunes would only allow the owners to double their labor plus government benefits annual stream of income.

It is central 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 small relative to the income effect. This result needs to be qualified when y is correlated with 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 emphasizes that we should tax rich rentiers (those who get predominantly capital income) and that we should spare the working rich (those whose labor income stream is significant relative to capital income). On this respect, it is interesting to note that the composition of income within the very top income groups can change overtime. Piketty and Saez (2001), exploiting tax returns statistics in the U.S. from 1913 to 1998, document that top income earners were mostly rentiers at the beginning of the period but have been slowly replaced by highly compensated salary earners over the course of the century. Today in the U.S., labor income forms a very significant share of total income even within the very top income earners. This secular change did not happen in all countries. Piketty (2001) shows that top income earners in France are still mostly rentiers as in the beginning of the century. Therefore, the desirability of capital income taxation

efficiency concerns.

is weaker in the U.S. today than in France or in the U.S. one century ago. We come back to this important issue in Section 6.

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 sophisticated than uniform linear taxes as in the Chamley (1986) 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 impose estate or inheritance taxation with substantial exemption levels and a progressive structure of marginal tax rates.[15] Most individual income tax systems have increasing marginal tax rates and capital income is often in large part included in the tax base,[16] producing a progressive capital income tax structure.

Non-linear capital income taxes in the dynastic 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. Obviously, progressive taxation cannot be as efficient than the wealth specific linear taxation of the previous system because progressive taxation generates a link between taxes paid by the poor and rich: low marginal tax rates on the poor means lower infra-marginal tax receipts from the rich.

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

[15] The U.S. for example exempts estates below \$675,000, and imposes progressive estate tax rates from 37% to 55% on larger estates.

[16] Most countries, such as the U.S., include dividends, rents, and interest income in the individual tax base. Capital gains receive in general a special treatment. Capital gains are in general taxed upon realization and not on an accrual basis. Some countries exempt capital gains fully from taxation; others, such as the U.S., tax capital gains according to special schedules, in general less progressive than ordinary individual income taxation.

taxation all individuals with wealth a_t below a given threshold a_t^* (possibly time varying), and imposes a 100% marginal tax rate on all capital income derived from wealth in excess of a_t^*. It can be shown (see below) that none of our results are changed if we assume that the government can set a marginal tax rate $\tau > 0$, however small, in the top bracket. More precisely $I_t(\rho a_t) = 0$ if $a_t \leq a_t^*$, 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 $a_t^* \geq 0$ so that individuals with zero wealth have no tax liability. 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^*. Note that virtual income m_t is zero for those in the exemption bracket at time t ($a_t \leq a_t^*$) and is $m_t = \rho a_t^*$ for those in the tax bracket ($a_t > a_t^*$).

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

$$a_t = a_0 + \rho A_t^* + y \cdot t - \frac{c_0}{\sigma\rho}\left(1 - e^{-\sigma\rho t}\right). \tag{25}$$

The only difference with equation (13) is the presence of the extra-term ρA_t^* due to the presence of the exemption threshold. As a_t^* is non-decreasing and c_t is decreasing, $\dot{a}_t$ is increasing. It is easy to show that wealth a_t declines up to point where it reaches a_t^*. This happens at time T (which depends of course on a_0) such that:

$$a_T^* = a_0 + \rho A_T^* + y \cdot T - \frac{c_0}{\sigma\rho}\left(1 - e^{-\sigma\rho T}\right). \tag{26}$$

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 (by the transversality condition): $a_t = a_T^* = (c_T - y)/\rho$. As $\dot{a}_t = \rho a_t^* + y - c_t$, this implies that $\dot{a}_T = 0$ (both from the left and the right). Therefore, as depicted on Figure 1, the pattern of consumption is exponentially decreasing up to time T and flat afterwards. The wealth pattern is also declining up to time T. At $t = T$, wealth is flat ($\dot{a}_T = 0$) and hits the exemption threshold a_T^* and remains flat afterwards. We denote as above the (constant) levels of consumption and wealth after time T by $c_\infty(a_0)$ and $a_\infty(a_0)$. Obviously, 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. We can note here that 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. In the present particular case, (7) can be written as

$$c_0 = \frac{\sigma[y + \rho(y \cdot T + \rho A_T^* + a_0)]}{1 - (1-\sigma)e^{-\sigma\rho T}}. \tag{27}$$

Using (9), the present discounted value of taxes paid by an individual with initial wealth a_0 is:

$$Tax(a_0, T) = \int_0^T \rho[a_t - a_t^*]e^{-\rho t}dt = \frac{y}{\rho} + a_0 - \frac{c_0}{\rho} \cdot \frac{1 + \sigma e^{-(\sigma+1)\rho T}}{1+\sigma}. \tag{28}$$

Note that expression (28) is identical to expression (16). 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 equation (27): the initial level of consumption c_0 contains a extra-term ρA_T^* reflecting the extra "virtual" income due to the exemption of taxation below the threshold a_t^*. As in Section 2, 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.[17] 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 wealth effect through the virtual income, 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

[17]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.

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 Figure 2. Obviously, all individuals with wealth $a_0 < \bar{a}_0$ are unaffected by the reform. 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*

$$\delta c_0 \approx \rho \left[\sigma\rho(T - \bar{t}) - \sigma\right] \delta a^*. \tag{29}$$

The formal proof is simple and provided in appendix. Let us provide the economic intuition. The reform increases the virtual income m_t by δa^* between times $\bar{t}$ and T. As can be seen from (27) assuming T is large, this produces a direct positive virtual income effect $\sigma\rho(T - \bar{t})\delta a^*$ on c_0. This is the first term in (29).

As can be seen on Figure 2, after the reform, the time needed to reach the exemption threshold is reduced by $\delta T < 0$ because the exemption threshold is higher. As we know from Section 3, a change in T produces again three effects: a substitution effect, an income effect, and a human wealth (and virtual income) effect. However, in this case, the human wealth (and virtual income) and income effects exactly cancel out because at time T, the consumption level (to which the income effect is proportional) and the income stream (including virtual income) level (to which the human wealth (and virtual income) effect is proportional) are identical: $c_T = \rho a^* + y$. As a result, we are left only with the substitution effect. This substitution effect is due to the fact that the regime with positive tax rate is now shorter and therefore the individual reduces c_0. For large $\bar{t}$ and hence T, equation (27) shows that the substitution effect on c_0 is approximately $\sigma\rho\sigma e^{-\sigma\rho T} c_0 \delta T = -\sigma\rho\delta a^*$.[18] This is the second term in (29).

Equation (29) shows that increasing the exemption threshold induces a positive effect on consumption for individuals with large T (i.e. large a_0) 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 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 equation (27), we have, for T large, $c_0 = \sigma\rho a_0(1 + o(1))$. Thus, as $c_0 e^{-\sigma\rho T} = y + \rho a^*$, we have

$$T = \frac{1}{\sigma\rho}[\log a_0 + \log(\sigma\rho) - \log(y + \rho a^*) + o(1)]. \tag{30}$$

Applying this equation at T and $T = \bar{t}$ (remembering that $T(\bar{a}_0) = \bar{t}$), we can rewrite (29) as

$$\delta c_0 \approx \rho \left[\log \frac{a_0}{\bar{a}_0} - \sigma\right] \delta a^*. \tag{31}$$

Using equation (28), and the expressions for δc_0 that we obtained in (31), for large $\bar{t}$ and T, we have, up a first order of approximation[19]

$$\delta Tax(a_0) \approx -\frac{\delta c_0}{\rho(1+\sigma)} \approx \frac{\delta a^*}{\sigma + 1}\left[\sigma - \log \frac{a_0}{\bar{a}_0}\right]. \tag{32}$$

Equation (32) 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

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

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

of rich. Integrating equation (32) over the distribution of wealth above $\bar{a}_0$, we obtain the effect of the reform on aggregate tax revenue:

$$\delta Tax \approx \frac{\delta a^*}{\sigma+1}\int_{\bar{a}_0}^{\infty}\left[\sigma - \log\frac{a_0}{\bar{a}_0}\right]h(a_0)da_0 = \frac{\delta a^*}{\sigma+1}\left[\sigma - A(\bar{a}_0)\right]\cdot\left[1-H(\bar{a}_0)\right] \tag{33}$$

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 equation (32), 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.

- **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, equation (33) 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 feature is closely linked 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 at no tax revenue cost for the government because there is nobody above that income level to benefit from the local reduction in the marginal tax rate. Similarly, in the capital income tax problem analyzed here, if the wealth distribution is bounded, then the government would improve welfare by increasing the exemption level a^* in place when the richest individual reaches it. This would improve the incentives of the richest individual to accumulate wealth and thus would increase its tax liability while producing no effect on all the other taxpayers.

- **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 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, 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 $1/\alpha$ and equation (33) becomes

$$\delta Tax \approx \frac{\delta a^*}{\sigma+1}[\sigma - \frac{1}{\alpha}]. \tag{34}$$

It is well known (since the work of Pareto in the late 19th century) that Pareto distributions approximate extremely well the top tails of income and wealth distributions. Using the large microfiles on individual tax returns publicly released by the Internal Revenue Service in the U.S., 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[20] as the sum of dividends, interest income, rents, fiduciary income (trust and estate income), and I plot on Figure 3 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 $\alpha = 1.5$.[21] Figure 3 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.[22] As is well known, the Pareto parameters of wealth (or capital income) distributions are lower (in general around 1.5 and 2) than those of labor income distributions (in general around 2 and 3).

Formula (34) 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. We show below that the welfare effect of this reform is negligible relative to the tax revenue effect. Therefore,

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

[21] Statistics compiled by the IRS by size of dividends since 1927, and exploited in Piketty and Saez (2001) show that the Pareto parameter for dividend income from 1927 to 1995 has always been around 1.5-1.7.

[22] 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.

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.

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_∞^*.*
- *If $\sigma \cdot \alpha > 1$ then the threshold a_t^* diverges to infinity and thus the asymptotic wealth distribution is unbounded.*

Proof: The technical proof of the result is in appendix.

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 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 the 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.

The reader might wonder to what extent the results of Proposition 2 strikingly favorable to redistribution via capital income taxation rely on the assumption that the exogenous upper bound for the tax rate had been taken equal to the very high value of one. In fact, it is possible to redo the formal analysis with any upper bound $\tau > 0$, and show that the results of Proposition 1 remain true for any $\tau > 0$, however small. This parallel analysis is outlined in appendix. We analyze in the numerical simulations of Section 5, how changing the upper-bound τ affects the *levels* of the asymptotic threshold level a^*.

4.3 Is the simple two-bracket tax system optimal?

We have considered in this paper a very particular case of non-linear capital income taxation. The natural question that arises is whether the simple two bracket structure considered here is optimal among the set of all non-linear capital income taxes? If the answer is negative, is it possible to characterize the optimal fully non-linear capital income tax, and what would the asymptotic wealth distribution be if the optimal non-linear capital income tax were implemented?

Solving for the optimal fully non-linear tax appears to be a daunting problem. The main difficulty is that, when the capital tax is not progressive in the sense that the budget constraint $\rho a \to \rho a - I(\rho a)$ is not concave, the individual maximization problem is not well behaved. For example, imposing high marginal tax rates at the bottom and a zero marginal tax rate above might look attractive because it allows to extract revenue from the rich in a quasi lumpsum way without producing efficiency costs. However, because the budget set is not concave, a number of individuals above the tax bracket would find optimal to let their wealth drift down and enter eventually the tax bracket. Computing fully the effects on tax revenue appears to be extremely complicated.

However, even if the full computation of the optimal non-linear tax is too complicated, the wealth asymptotic distribution analysis might be manageable, and the general result of Proposition 2 might be still true in the general case. We do not have yet a formal proof of this assertion but an intuitive argument can be made as follows.

In the long-run, an exemption threshold is certainly optimal because otherwise individuals would be taxed forever and that is clearly suboptimal because positive taxation generates a distortion growing exponentially with time. So, if there is an exemption threshold in the long-run, the asymptotic analysis made above can very possibly be adapted and the same general result proven.

If we restrict ourselves to non-linear taxation with increasing marginal tax rates, that is, schedules such that for each t, $I_t'(\rho a)$ is increasing in ρa, then the analysis of individual behavior is simplified. In that case, it should not be too hard to show that it is never optimal to have a marginal tax rate $I'(\rho a)$ strictly in between zero and the exogenous upper bound τ. As a result, the optimal tax in that situation takes the simple two-bracket form analysis previously.

The fact that the threshold a_t^* separating the exemption bracket from the tax bracket is non-decreasing is perhaps not always true at the optimum. However, it should be true for t large enough and thus our asymptotic results are very likely to be robust to this assumption.

5 Numerical Simulations

The goal of the numerical simulations is to analyze how large is the asymptotic threshold level a^* in the case where this threshold is finite (that is, from Proposition 2, when $\alpha \cdot \sigma < 1$). This analysis is important because it is important to assess how large (relative to the annual income stream $y = w + b$) will the largest wealth holdings be. In particular, we want to know how the threshold a^* varies with the key parameters σ (intertemporal elasticity of substitution), α (Pareto parameter of the initial wealth distribution), τ (the exogenous upper-bound for the tax rate), and ρ (the discount factor and interest rate).

For the numerical simulations, we normalize the wage level w to one. We calibrate the initial wealth distribution $H(a_0)$ as follows. We assume that the density distribution is Pareto above some threshold $\bar{a}_0$, and constant below $\bar{a}_0$.[23] The threshold $\bar{a}_0$ is chosen so that average wealth holdings produces an income stream equal to 0.25. This calibration replicates the approximate (80%, 20%) division of personal income into labor income and capital income.

As explained in appendix, we do not solve for the full path of a_t^* because this is a very challenging numerical exercise. We adopt instead a much simpler method to obtain upper-bounds for the asymptotic values of a^*. Preliminary numerical simulations in the general case suggest that these upper-bounds are in fact close the real limit value of the optimal path a_t^* for large t. All the details are provided in appendix.

Table 1 reports upper bounds for the asymptotic values of the capital income stream ρa^* for the richest individuals in the long-run.[24] Unsurprizingly, the optimal value of ρa^* is increasing with the intertemporal elasticity of substitution σ, and the thinness of the wealth distribution

[23] A constant density does not replicate exactly the empirical wealth distribution but this is not a concern as we focus on asymptotic results involving only the top of the wealth distribution.

[24] We present values of ρa^* instead of a^* in order to compare directly the capital income stream to the labor income stream (normalized to one).

measured by the Pareto parameter α. As we expect from Proposition 2, when the product $\sigma \cdot \alpha$ gets close to one, the value a^* becomes large. Therefore, the numerical simulations provide a useful complement to the knife-edge result of Proposition 2. While the threshold of one for the product $\sigma \cdot \alpha$ is *qualitatively* critical, the value of the threshold is very important *quantitatively* to assess how far the redistribution can go. For example, for very low values of the product, the capital income stream of the rich in the long-run is only a very small fraction of the labor income stream, impying a very low level of income inequality in the long-run. For values of the product close to one, that capital income stream is much larger than the labor income stream, implying that, even though humongous fortunes disappear, substantial income inequality is left in the long-run.

6 Extending the Basic Model

6.1 Endogenous interest rate and wages

Previous sections have considered the case with an exogenous interest rate $r_t = \rho$ and wage rate w. In general, dynamic models introduce 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).

It is possible to show that introducing such a neo-classical function would not change our results. 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.[25]

With a neoclassical production function and no 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. If the tax on capital income is positive and equal to τ in the long run, then the stock of capital is lower and given by $(1-\tau)f'(k_\infty) = \rho$.

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

It is interesting to note that the optimal set of taxes considered here always lead to the efficient level of capital $f'(k_\infty) = \rho$. This is because, even if the tax is never exactly zero, the number of people in the tax regime shrinks to zero.[26] If the capital stock is smaller than the modified Golden rule level, people in the exempt bracket would start accumulating capital.[27]

6.2 Role of debt

[TO BE DONE]

7 Conclusion

This paper has shown that introducing progressive taxation in the optimal dynamic capital income tax model can have a dramatic impact of the policy prescriptions. 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, under realistic assumptions on the intertemporal elasticity of substitution and the thickness of the top tail of the distribution, progressive taxation allows the government 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 dynastic 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 otherwise equilibria are degenerated. However, the dynastic model should not be judged on the realism of its assumptions but rather on the accuracy of its predictions for wealth accumulation. Relative to other models, the infinite horizon model has the realistic feature that wealth inequality is persistent in the absence of government

[26] This result is a direct application of the point made by Piketty (2001) that, contrary to linear capital taxation, progressive capital income taxation might not lower the long-run stock of capital in the economy.

[27] If the long-run exemption threshold a^* is very low, it might happen that the capital stock is still below the modified Golden rule level when all the poor reach a^*, producing a suboptimal long-run capital stock. However, for realistic values of the parameters, we expect the long-run a^* to be large enough so that this issue that does not come up.

interventions. It is perhaps the case that the infinite horizon model predicts too large responses to capital income taxes. However, this feature would bias the results against finding redistributive policies desirable.[28] It is therefore remarkable that the dynastic model produces tax policy recommendations so favorable to the breaking of large fortunes and redistribution of wealth.

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. There are perhaps political constraints preventing the government from applying such a drastic policy. 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 other words, even if the government is limited to using a maximum capital income tax rate as small as 10%,[29] large fortunes will eventually disappear. In the historical record of tax policy development of western countries, wealth inequality inherited from the past and the tremendous levels of large fortunes was certainly one of key arguments used 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.

Obviously, it is an interesting and important research question to understand how the results of this paper would be affected if the wealth distribution were endogenous. Numerous papers have extended the basic infinite horizon model to endogeneize the wealth distribution.[30] The wealth distribution might be unequal because of unequal past labor income streams. It would be interesting to know whether it is more efficient to have a progressive labor income tax to prevent wealth accumulation or to let skilled individuals keep a large share of their labor income and then apply a progressive capital income tax to curb down accumulated wealth holdings.[31]

[28] The Chamley-Judd results stating that optimal capital income taxes should be zero in the long-run have often been criticized on these grounds.

[29] This type of limit was certainly politically binding when western economies started introducing progressive income taxation a century ago.

[30] See e.g. Quadrini and Rios-Rull (1997) for a survey.

[31] The famous Atkinson and Stiglitz (1976) result on commodity taxation with non-linear labor income taxation suggests that, under some conditions, progressive labor income taxation should be enough.

Appendix

- **Optimality of the "bang-bang" tax system**

As in the Chamley (1986) model, we can show that the bang-bang tax system: maximum tax rate up to time $T(a_0)$ and no tax afterward is optimal (among the set of all possibly time varying and wealth specific linear taxes $\tau_t(a_0)$). The proof can be obtained as follows.

We consider the optimal bang-bang policy taxing individual with initial wealth a_0 up to time $T(a_0)$ at the maximum rate, and then introduce some small tax rate $d\tau$ on individual a_0 between time t and $t + dt$ for some $t > T(a_0)$. Tedious computations show that this policy has no first order effect on total welfare and thus cannot improve upon the bang-bang policy.

To complete the proof, we can also show that *reducing* the tax rate before time $T(a_0)$ by $d\tau$ between times t and $t + dt$ reduces total welfare.

[DETAILS TO BE COMPLETED]

- **Proof of Proposition 1**

It is useful to note that the denominator in equation (14), $1 - (1 - \sigma)e^{-\sigma\rho T}$, is between 1 and σ for any value of T. Using the expression (14), it is then clear that $c_0 \to +\infty$ as a_0 tends to infinity.

The envelope theorem implies that an extra second of taxation leads to a welfare loss equal to the marginal utility of consumption at time T times the value of the mechanical extra tax liability.[32] Therefore we have

$$\frac{\partial U(a_0, T)}{\partial T} = -u'(c_T)e^{-\rho T}\rho a_T = c_0^{-1/\sigma}[y - c_0 e^{-\sigma\rho T}]. \tag{35}$$

Let us examine now the tax revenue effect. Using (16), we have

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

Using the expressions (35) and (19), we can rewrite the first order condition (18) as

[32] Alternatively, this result can be obtained by deriving expression (17) with respect to T.

$$\frac{c_0^{-1/\sigma}}{p}\left[y - c_0 e^{-\sigma\rho T}\right] + \sigma\rho \cdot \frac{-y + c_0 e^{-\sigma\rho T} - \sigma c_0 e^{-\sigma\rho T}}{1-(1-\sigma)e^{-\sigma\rho T}} \cdot \frac{1+\sigma e^{-(\sigma+1)\rho T}}{1+\sigma} + \sigma\rho c_0 e^{-(\sigma+1)\rho T} = 0. \quad (36)$$

The first term is the welfare effect and the last two terms are the tax revenue effect.

It is important to note that, because, $c_0 \to \infty$, the welfare effect is negligible relative to $[y - c_0 e^{-\sigma\rho T}]$. This expression appears in the numerator of the second term of (36) 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 (36).

- Case $\sigma < 1$:

In that situation, $c_0 e^{-\sigma\rho T}$ must be bounded otherwise the numerator of the second term in (36) takes arbitrarily large positive values (as y is constant) and the third term of (36) is also positive, implying that (36) cannot hold. Therefore $c_0 e^{-\sigma\rho T}$ is bounded and thus the first term (welfare effect) in (36) tends to zero. As c_0 goes to infinity with a_0, it must be the case that T also grows without bound. As a result the third term in (36) converges to zero and thus (36) holds only if the second term also converges to zero. Therefore:

$$(1-\sigma)c_0 e^{-\sigma\rho T} \to y, \quad (37)$$

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 (21).

- Case $\sigma > 1$:

In that situation, the behavioral response unambiguously reduces tax revenue. Therefore the behavioral response must be compensated by the positive mechanical effect. In that case T must be bounded because otherwise the third term in (36) would be negligible relative to $c_0 e^{-\sigma\rho T}$ and (36) could not hold. As T is bounded and as $c_0 \to \infty$, the dominant terms proportional to c_0 in (36) must cancel each other, implying that:

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

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. One can note that T^{∞} decreases with σ and tends to infinity when σ decreases to one. Using equation (15), it is then easy to obtain (22).

- Case $\sigma = 1$:

In that case, the tax revenue effect simplifies to:

$$\frac{\partial Tax(a_0, T)}{\partial T} = -\rho y \cdot \frac{1 + e^{-2\rho T}}{2} + \rho c_0 e^{-2\rho T}$$

Thus the first order condition (36) can be satisfied for large a_0 only if $c_0 e^{-2\rho T} \to y/2$. As $c_0 \sim \rho a_0$, we obtain both (23) and (24).

When the maximum rate is $\tau > 0$, similar computations go through and the same results are obtained. QED.

- **Proof of Lemma 1**

Differentiating (25), we have

$$\delta a_t = \rho \delta a^* (t - \bar{t}) 1(t > \bar{t}) - \frac{\delta c_0}{\sigma \rho} \left(1 - e^{-\sigma \rho t}\right). \tag{38}$$

At $t = T$, it must the case that $\delta a_T = \delta a^*$ because the individual reaches the tax regime at time T (up to a first order effect). Using (38), this implies that

$$\delta c_0 = \frac{\sigma \rho}{1 - e^{-\sigma \rho T}} \left[\rho (T - \bar{t}) - 1\right] \delta a^*. \tag{39}$$

When $\bar{t}$ (and hence T) is large, equation (39) can be approximated as (29) and the Lemma is proved.

- **Formal Proof of Proposition 2**

TO BE DONE

- **Asymptotic analysis with a general upper bound τ**

Suppose the government adopts a two-bracket tax system as in Section 4 but with a tax rate τ. In that case the virtual income m_t is equal to $\tau \rho a_t^*$ for those in the tax bracket. I denote again by T the time at which the individual reaches the exemption threshold. The cumulated net interest rate $\bar{R}_t$ is equal to $\rho(1-\tau)t$ for $t \leq T$, and to $\rho t - \rho\tau T$ for $t \geq T$. Routine computations show that equation (7) implies

$$c_0 = (1-\tau+\tau\sigma) \cdot \frac{\rho a_0 + y(1-\tau e^{-\rho(1-\tau)T})/(1-\tau) + \tau\rho\int_0^T \rho a_t^* e^{-\rho(1-\tau)t}dt}{1-\tau(1-\sigma)e^{-\rho T(1-\tau+\sigma\tau)}}, \tag{40}$$

and that equation (9) implies

$$Tax(a_0) = a_0 + \frac{y}{\rho} - c_0 \cdot \frac{1+\sigma\tau e^{-\rho(1+\sigma\tau)T}}{\rho(1+\sigma\tau)} \tag{41}$$

After time T, consumption and wealth are constant, and we have $c_T = c_0 e^{-\rho\sigma\tau T} = \rho a_T^* + y$. We consider, as in the text, a small increase δa^* for a^* for $t \geq \bar{t}$. Differentiating equation (40) and $c_0 e^{-\rho\sigma\tau T} = \rho a^* + y$, and assuming that T is large, we obtain

$$\delta c_0 \approx \rho\delta a^*(1-\tau+\sigma\tau)e^{-\rho(1-\tau)T}\left[\frac{\tau}{1-\tau}\left(e^{\rho(1-\tau)(T-\bar{t})}-1\right)-1\right]. \tag{42}$$

Using equation (40) and $c_0 e^{-\rho\sigma\tau T} = \rho a^* + y$, we can change variables from T to a_0,

$$T = \frac{1}{\sigma\rho\tau}[\log a_0 + \log[\rho(1-\tau+\tau\sigma)] - \log(y+\rho a^*) + o(1)]. \tag{43}$$

Using this equation for T and $T = \bar{t}$ (corresponding to wealth level $\bar{a}_0$, we can rewrite (42) as

$$\delta c_0 \approx \delta a^* \rho(1-\tau+\sigma\tau)e^{-\rho(1-\tau)\bar{t}}\left(\frac{a_0}{\bar{a}_0}\right)^{-\frac{1-\tau}{\sigma\tau}}\left[\frac{\tau}{1-\tau}\left(\left(\frac{a_0}{\bar{a}_0}\right)^{\frac{1-\tau}{\sigma\tau}}-1\right)-1\right]. \tag{44}$$

The change in tax paid is obtained by differentiating (41). For large T, we have the following approximation

$$\delta Tax(a_0) \approx -\frac{\delta c_0}{\rho(1+\sigma\tau)} \approx \frac{\delta a^*(1-\tau+\sigma\tau)e^{-\rho(1-\tau)\bar{t}}}{1+\sigma\tau}\left(\frac{a_0}{\bar{a}_0}\right)^{-\frac{1-\tau}{\sigma\tau}}\left[1-\frac{\tau}{1-\tau}\left(\left(\frac{a_0}{\bar{a}_0}\right)^{\frac{1-\tau}{\sigma\tau}}-1\right)\right]. \tag{45}$$

Assuming that a_0 is Pareto distributed in the tail with parameter α, a simple computation shows that,

$$\delta Tax \approx \frac{\delta a^*(1-\tau+\sigma\tau)e^{-\rho(1-\tau)\bar{t}}}{1+\sigma\tau} \cdot \frac{\alpha\sigma\tau}{\alpha\sigma\tau+1-\tau}\left[1-\frac{1}{\sigma\alpha}\right]. \tag{46}$$

Equation (46) shows that the result of Proposition 2 goes through for any τ. The intuition is exactly the same.

- **Numerical Simulations**

[DETAILS TO BE EXPANDED]

The upper bounds for a^* reported in Table 1 are computed as follows. We assume that the threshold is constant and equal to a^* after some time t^* (not necessarily large). We then consider a small increase in a^* above time $\bar{t} > t^*$ and compute the effect on total tax collected (using the initial wealth distribution described in the text), and compute the effect on total tax collected using exact formulas, denoted by $dTax(a^*, \bar{t})/da^*$. When $\sigma \cdot \alpha <$, we know from Proposition 2 that $dTax(a^*, \bar{t})/da^* < 0$ for $\bar{t}$ large enough. However, for smaller $\bar{t}$ the change in taxes might be positive. It turns out that the function $\bar{t} \to dTax(a^*, \bar{t})/da^*$ has an inverted U-shape, first increasing and then decreasing. It is negative for small and large values of $\bar{t}$. The maximum of the function reached at a time denoted by $\bar{t}^*(a^*)$ might be positive or negative. If is negative for large values of a^*, and positive for low values of a^*.

When a^* is such that the function is always negative, a decrease in a^* always increases tax revenue. Therefore, such values of a^* cannot be optimal. The particular value of a^* denoted by a^{**} such that the function is always negative except at its maximum where it is zero is therefore an upper bound for the optimal a^*. These are the values reported on Table 1.

References

Atkeson, Andrew, V.V. Chari, and Patrick Kehoe (1999), "Taxing Capital Income: A Bad Idea." *Federal Reserve Bank of Minnepolis Quarterly Review*, Summer 3-17.

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

Atkinson, Anthony B. and Joseph E. Stiglitz. (1976), "The Design of the Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics*, 6, 55-75.

Chamley, Christophe P. (1986), "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives." *Econometrica* 54, 607-622.

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. and James Mirrlees (1971), "Optimal Taxation and Public Production." *American Economic Review*, 61, 8-27, 261-278.

Feldstein, Martin (1978), "The Welfare Cost of Capital Income Taxation", *Journal of Political Economy*, 86, S29-S52.

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 G. Hughes and G. Heal, eds., *Public Policy and the Tax System*, (London: George Allen Unwin), 1-36.

Kotlikoff, Lawrence and Lawrence Summers (1981), "The Role of Intergenerational Transfers in Aggregate Capital Accumulation", *Journal of Political Economy*, 89, 706-732.

Lucas, Robert E., Jr. (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. and Edmund Phelps (1979), "The Concept of Optimal Taxation in the Overlapping Generations Model of Capital and Wealth," *Journal of Public Economics*, 12, 1-26.

Piketty, Thomas (2001), "Income Inequality in France, 1901-1998", CEPR Discussion Paper No. 2876.

Piketty, Thomas, and Emmanuel Saez (2001), "Income Inequality in the U.S., 1913-1998", NBER Working Paper No. 8467.

Quadrini, Vincenzo, and José-Victor Rios-Rull (1997), "Understanding the U.S. Distribution of Wealth.", *Federal Reserve Bank of Minneapolis Quarterly Review*, 21(2), 22-36.

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, 1986, 137-143.

Slemrod, Joel (2000), *Does Atlas Shrug? The Economics Consequences of Taxing the Rich*, Cambridge University Press.

FIGURE 1: Wealth and Consumption Dynamics

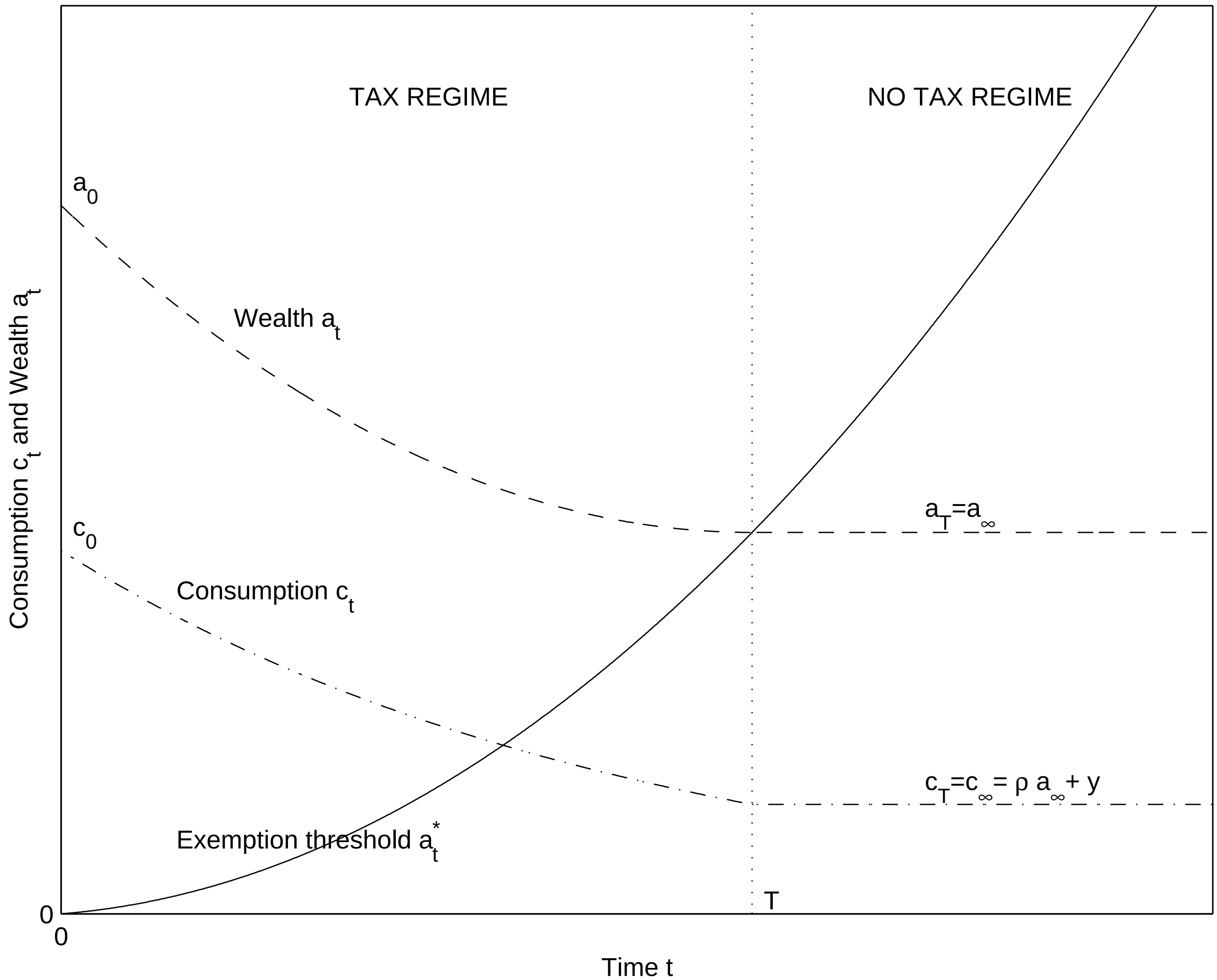

FIGURE 2: Increasing the exemption threshold

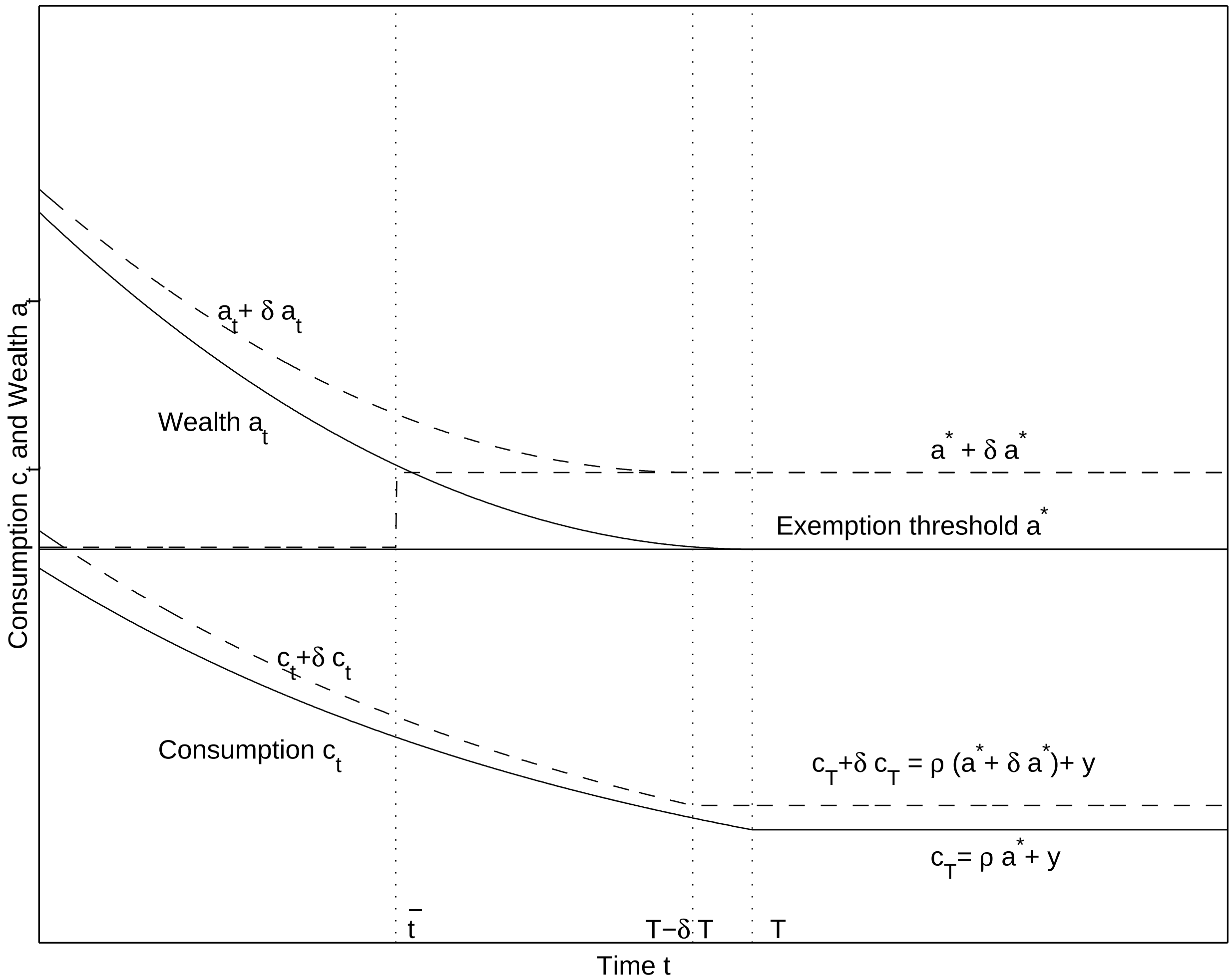

FIGURE 3: Empirical Statistic A(z0)=E(log(z/z0)|z>z0) for capital income

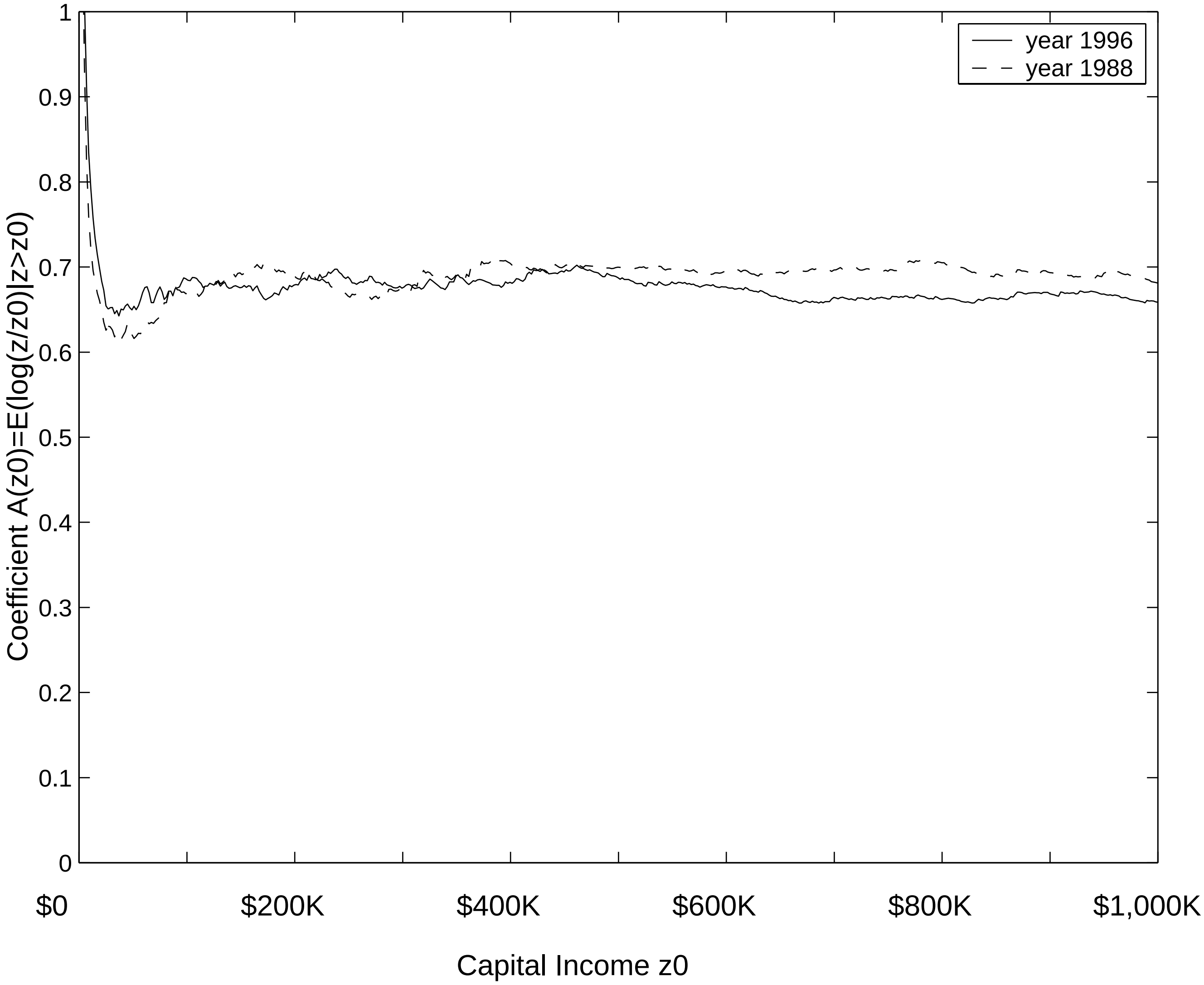

TABLE 1 - Upper bounds for asymptotic top capital income stream ρa^*

	Intertemporal Elasticity σ=0.125			Intertemporal Elasticity σ=0.25			Intertemporal Elasticity σ=0.45			Intertemporal Elasticity σ=0.6		
	Marginal Tax Rate τ			Marginal Tax Rate τ			Marginal Tax Rate τ			Marginal Tax Rate τ		
	0.25	0.5	1	0.25	0.5	1	0.25	0.5	1	0.25	0.5	1
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pareto Parameter α												
1.5	0.020	0.029	0.060	0.052	0.077	0.163	0.166	0.273	0.734	0.527	1.156	3.385
2	0.065	0.092	0.199	0.189	0.284	0.745	1.410	4.163	6.783	infty	infty	infty
2.5	0.132	0.188	0.449	0.463	0.763	5.124	infty	infty	infty	infty	infty	infty
3	0.223	0.324	0.922	1.109	2.384	12.010	infty	infty	infty	infty	infty	infty

The table reports upper bounds for the asymptotic top capital income stream ρa^* expressed in terms of the average annual income stream y=w+b for various parameter values for σ, τ, and α. In the long-run with optimal progressive taxation, the highest wealth level is below a*. Computation details are provided in appendix.
The discount factor ρ is equal to 0.05.