NBER WORKING PAPER SERIES

A SIMPLER THEORY OF OPTIMAL CAPITAL TAXATION

Emmanuel Saez
Stefanie Stantcheva

Working Paper 22664
http://www.nber.org/papers/w22664

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2016, Revised September 2017

We thank Alan Auerbach, Stephen Coate, Emmanuel Farhi, Mike Golosov, Roger Gordon, Louis Kaplow, Ben Lockwood, Andreas Peichl, Thomas Piketty, Chris Sleet, Joel Slemrod, Matthew Weinzierl, Nicolas Werquin, Daniel Waldenström, and numerous seminar participants for useful discussions and comments. We acknowledge financial support from the MacArthur Foundation, and the Center for Equitable Growth at UC Berkeley. We thank Nina Roussille for research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Simpler Theory of Optimal Capital Taxation
Emmanuel Saez and Stefanie Stantcheva
NBER Working Paper No. 22664
September 2016, Revised September 2017
JEL No. H20,H21

ABSTRACT

This paper develops a theory of optimal capital taxation that expresses optimal tax formulas in sufficient statistics. We first consider a simple model with utility functions linear in consumption and featuring heterogeneous utility for wealth. In this case, there are no transitional dynamics, the steady-state is reached immediately and has finite elasticities of capital with respect to the net-of-tax rate. This allows for a tractable optimal tax analysis with formulas expressed in terms of empirical elasticities and social preferences that can address many important policy questions. These formulas can easily be taken to the data to simulate optimal taxes, which we do using U.S. tax return data on labor and capital incomes. Second, we show how these results can be extended to the case with concave utility for consumption. The same types of formulas carry over by appropriately defining elasticities. We show that one can recover all the results from the simpler model using a new and non standard steady state approach that respects individual preferences even with a fully general utility function or uncertainty.

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

Stefanie Stantcheva
Department of Economics
Littauer Center 232
Harvard University
Cambridge, MA 02138
and NBER
sstantcheva@fas.harvard.edu

1 Introduction

The public debate has long featured an important controversy about the proper design of capital taxation. Arguments typically center around an equity-efficiency trade-off: who owns the capital and how strongly would capital react to higher taxes? The economics literature has developed dynamic, complex models, which have emphasized different results depending on the structure of individual preferences and shocks, the government's objective, and the policy tools available. Many of the highly salient questions in the policy debate on capital taxation have been very difficult to address in these complex models. A few examples are how to take into account income shifting between the capital and labor income bases, different types of capital assets, heterogeneity in individuals' preferences or returns to capital, nonlinear taxation, and broader social fairness and equity considerations. Bridging the gap between economic theory and the policy debate seems especially important in the current context with growing income and wealth inequality, and where a large fraction of top incomes comes from capital income (Saez and Zucman, 2016; Piketty et al., 2016).

The goal of this paper is to connect the theory of optimal capital taxation to the public debate by providing a framework in which many policy questions related to capital taxation can be addressed. This framework permits the derivation of robust optimal capital tax formulas expressed in terms of elasticities of capital supply with respect to the net-of-tax rate of return that can be estimated in the data, and distributional considerations which society may have. The aim is to build a model which generates an empirically realistic response of capital to taxes,[1] is sufficiently tractable to yield results for a variety of policy topics related to capital taxation, but general enough for these results to be robust to a broader set of models.

We start in Section 2 with a simple model in continuous time with the following two ingredients. First, individuals derive utility from wealth. We provide several microfoundations for this wealth in the utility specification: bequest motives, entrepreneurship, or services from wealth. It implies that the steady-state features finite supply elasticities of capital with respect to tax rates.[2] It also implies that there is bi-dimensional heterogeneity in capital and labor incomes

[1]Our model generates finite elasticities of capital responses to capital taxes. In contrast, the famous result of Chamley (1986) and Judd (1985) that in the long-run the optimal capital tax should be zero–arises because the elasticity to a long-run tax increase is infinite due to anticipation effects (see Piketty and Saez (2013b)).

[2]The magnitude of capital income elasticities is an empirical question. Our model nests the case of infinite steady state elasticities from the standard Chamley-Judd model as a special case. Other possible modeling

across individuals. As a result, the famous zero tax result of Atkinson and Stiglitz (1976) does not apply. Second, utility of consumption is linear so that there are no consumption smoothing issues and individual responses to tax changes are immediate.[3] While necessary to analyze insurance issues as in the New Dynamic Public Finance literature,[4] consumption smoothing due to concave utility seems, at a first pass, less important for thinking about taxation of top incomes, where most of the capital is ultimately concentrated, and long-run taxation.[5]

While we generalize this model later on to allow for concave utilities (and the anticipatory and sluggish responses of capital to taxation they generate), the simpler version with linear consumption is extremely tractable and amenable to studying a wide range of issues about optimal capital taxation, such as nonlinear capital taxation, income shifting, cross-elasticities between capital and labor income, consumption taxation and others in Section 3. It highlights the main forces shaping capital taxation which are obscured in more complex models. Another key advantage is that it resolves the highly thorny issue of how to tax the existing capital stock and does not require making a judgment about what type of reform to consider (e.g.: anticipated, unanticipated, steady state focused).[6] Confronting these dilemmas in earlier papers required making normative judgments on the social welfare objective, none of which are entirely satisfactory (see Section 5).

We can describe four sets of findings that we obtain by putting this newly gained simplicity to use.

---

devices to obtain finite elasticities would be introducing uncertainty as in the Aiyagari (1995) model as shown by Piketty and Saez (2013b) or discount rates that depend on consumption (as in Judd (1985)). We argue that utility of wealth is much simpler and fits the data better in Section 2.3, but do relate our results to these alternative models in Section 5.

[3]Anticipated tax reforms do not create any effect until they actually take place, which greatly simplifies the analysis by eliminating the need to model anticipation effects and expectations about policy. This is unlike in the Chamley (1986) and Judd (1985) theory where unanticipated capital taxes are desirable while anticipated long-distance capital taxes are not.

[4]Golosov et al. (2003) founded this literature. Golosov et al. (2006) provide a comprehensive overview.

[5]To draw the analogy to the labor income tax literature, responses of labor to taxes are also part of a dynamic decision process if we acknowledge longer-term and slowly adjusting margins such as occupational choice and human capital acquisition. Two strands of the literature have thought of labor taxation in a dynamic way: the heterogeneous agents macro literature as in Jones et al. (1993) and the modern New Dynamic Public Finance literature. While providing useful insights, this literature has not been central in the public policy debate on taxation. The missing piece in optimal capital tax theory that we propose here is an approach that is dynamic, but can yield a static-equivalent model, which abstracts from transitional dynamics, and as was adopted for labor income following the seminal contribution of Mirrlees (1971).

[6]In Section 5 we show that we can recover this highly valuable neutrality property in the generalized model, but at the cost of having to adopt a non-standard normative approach.

First, we derive formulas for optimal linear and nonlinear capital income taxation that can be expressed in terms of the elasticity of the supply of capital income with respect to the net-of-tax rate of return, the shape of the capital income distribution, and the social welfare weights at each capital income level. We also derive formulas which take into account policy issues that have traditionally been hard to deal with in dynamic optimal capital tax models. These include, among others, income shifting between capital and labor, economic growth, heterogeneous returns to capital across individuals, and different types of capital assets and heterogeneous tastes for each of them.

Second, we derive a formula for the optimal tax on comprehensive income (labor plus capital income) that takes exactly the same form as the traditional optimal labor income formula. This formally justifies the use of the optimal labor income formulas to discuss optimal income taxation as has been done without rigorous justification in a number of studies (e.g., Diamond and Saez (2011)). The comprehensive income tax is the fully optimal tax if there is perfectly elastic income shifting between the labor and capital income bases when labor and capital are taxed differentially.

Third, we can analyze consumption taxation in this model as well by making the assumption that real wealth (i.e., the purchasing power of wealth) enters individual utilities. In this case, a consumption tax makes people accumulate more nominal wealth so that their steady-state real wealth is unchanged. Hence, consumption taxation ends up being equivalent to labor taxation plus an initial wealth levy.[7] It is thus not a sufficient tool to address capital inequality when capital inequality re-emerges even after an initial wealth equalization as in our model and as seems to be the case from empirical experiences (e.g., Eastern Europe's transition to a market economy and in particular Russia as recently analyzed in Novokmet et al. (2017)).

Fourth, our approach is very amenable to considering a broader range of justice and fairness principles related to capital taxation, through the use of generalized social welfare weights as in Saez and Stantcheva (2016). Given the prevalence of discussions about fairness and equity with regard to capital taxation, having a tractable way to incorporate broader and more diverse equity considerations is key.[8] We consider several salient ethical standpoints from the policy

[7]The same equivalence holds in standard OLG models traditionally used to discuss transitions from income to consumption taxation (Auerbach and Kotlikoff, 1987).

[8]Put simply, to obtain the optimal tax for different justice and fairness principles, the reader can use all the formulas derived and "plug in" the corresponding generalized social welfare weights.

debate. To give just one example, if differences in capital are considered fully fair (i.e., the generalized social welfare weights are uncorrelated with capital and capital income is not a tag) the optimal capital tax is zero.

In Section 4, we put our formula in sufficient statistics to use by calibrating optimal taxes based on U.S. tax data on labor and capital income. Because capital income is much more concentrated than labor income, we find that, if the supply elasticities of labor and capital with respect to tax rates were the same, the top tax rate on capital income would be higher than the top tax rate on labor income. The model highlights which elasticities should fruitfully be estimated in the data, including the cross-elasticities between capital and labor (Section 3.2.2) and the elasticities and cross-elasticities for different types of capital assets (Section 3.3.3).

In Section 5, we show that the tax formulas obtained in the specific model of Section 2 carry over to the general case with concave utility for consumption as long as the elasticity of the capital income tax base is appropriately defined. However in the general case, there are transitional dynamics and as a result, there is no single way to pose the optimal tax problem. We consider three cases: unanticipated tax changes, anticipated tax changes, and a focus on the long-run steady-state with a new and non standard approach which we call the "utility based steady state approach."[9] In each case, the same qualitative formulas expressed in terms of elasticities apply, but the quantitative elasticities are different. Choosing one of these cases is itself a policy choice that entails a normative judgment. We discuss the pros and cons of each case.[10]

All proofs and many extensions are gathered in the Online Appendix.

[9] The simpler model in Section 2, in addition to its tractability, does not require such a normative judgment in the choice of the type of reform, as they are all equivalent.

[10] The results of this section relate to earlier work by Piketty and Saez (2013b) who consider a model of bequest taxation with stochastic shocks in earnings abilities over generations. In contrast, we consider a model with certainty. This makes the derivations simpler. We can do this because our wealth in the utility model generates finite elasticities. As we shall see, for the anticipated and unanticipated reform approaches, many of the formulas from Piketty and Saez (2013b) are similar to ours, illustrating the virtue of the sufficient statistics approach. Our utility based steady state approach is a new and non standard way of dealing with capital taxation in a fully general setting.

2 A Simpler Model of Capital Taxation

In this section, we present a simpler model of capital taxation. The key simplification comes from having utility linear in consumption, which implies immediate convergence to the steady state. The key additional component is to introduce wealth in the utility, which allows for smooth responses of capital to taxation. This model usefully highlights the key efficiency-equity trade-off for capital taxation, often obscured in more complex models.

2.1 Model

Time is continuous. Individual i has instantaneous utility with functional form $u_i(c, k, z) = c + a_i(k) - h_i(z)$, linear in consumption c, increasing in wealth k with $a_i(k)$ increasing and concave, and with a disutility cost $h_i(z)$ of earning income z increasing and convex in z. The individual index i can capture any arbitrary heterogeneity in the preferences for work and wealth, as well as in the discount rate δ_i. We justify the assumption of wealth in the utility in great detail below. The discounted utility of i from an allocation $\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}$ is:

$$V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}) = \delta_i \cdot \int_0^{\infty} [c_i(t) + a_i(k_i(t)) - h_i(z_i(t))]e^{-\delta_i t} dt. \tag{1}$$

We normalize utility by the discount rate δ_i so that an extra unit of consumption in perpetuity increases utility by one unit uniformly across all individuals. The net return on capital is r. At time 0, initial wealth of individual i is k_i^{init}. For any given time-invariant tax schedule $T(z, rk)$ based on labor and capital incomes, the budget constraint of individual i is:

$$\frac{dk_i(t)}{dt} = rk_i(t) + z_i(t) - T(z_i(t), rk_i(t)) - c_i(t). \tag{2}$$

$T'_L(z, rk) \equiv \partial T(z, rk)/\partial z$ denotes the marginal tax with respect to labor income and $T'_K(z, rk) \equiv \partial T(z, rk)/\partial(rk)$ denotes the marginal tax with respect to capital income.

The Hamiltonian of individual i at time t, with co-state $\lambda_i(t)$ on the budget constraint, is:

$$H_i(c_i(t), z_i(t), k_i(t), \lambda_i(t)) = c_i(t) + a_i(k_i(t)) - h_i(z_i(t)) + \lambda_i(t) \cdot [rk_i(t) + z_i(t) - T(z_i(t), rk_i(t)) - c_i(t)].$$

Taking the first order conditions, the choice $(c_i(t), k_i(t), z_i(t))$ is such that:

$$\lambda_i(t) = 1, \quad h_i'(z_i(t)) = 1 - T_L'(z_i(t), rk_i(t)), \quad a_i'(k_i(t)) = \delta_i - r[1 - T_K'(z_i(t), rk_i(t))], \quad \text{and}$$

$$c_i(t) = rk_i(t) + z_i(t) - T(z_i(t), rk_i(t)).$$

In this model, $(c_i(t), k_i(t), z_i(t))$ jumps immediately to its steady-state value (c_i, k_i, z_i) characterized by $h_i'(z_i) = 1 - T_L', a_i'(k_i) = \delta_i - r(1 - T_K'), c_i = rk_i + z_i - T(z_i, rk_i)$. This is achieved by a Dirac quantum jump in consumption at instant $t = 0$, so as to bring the wealth level from the initial k_i^{init} to the steady state value k_i. Because of this immediate adjustment and the lack of transition dynamics, we have that:

$$V_i(\{c_i(t), k_i(t), z_i(t)\}_{t \geq 0}) = [c_i + a_i(k_i) - h_i(z_i)] + \delta_i \cdot (k_i^{init} - k_i),$$

where the last term $(k_i^{init} - k_i)$ represents the utility cost of going from wealth k_i^{init} to wealth k_i at instant 0, achieved by the quantum Dirac jump in consumption.

Heterogeneous wealth preferences and a smooth steady state. Wealth accumulation in this model depends on the heterogeneous individual preferences, as embodied in the taste for wealth $a_i(\cdot)$ and in the impatience δ_i. It also depends on the net-of-tax return $\bar{r} = r(1 - T_K'(z, rk))$: capital taxes discourage wealth accumulation through a substitution effect (there are no income effects). Because of a possibly arbitrary heterogeneity in preferences for capital, steady state wealth holdings are heterogeneous across individuals and capital exhibits a smooth behavior in the steady state, with a finite elasticity of capital supply with respect to the net-of-tax return. This also implies that there is heterogeneity in wealth even conditional on labor earnings. Therefore, the famous zero tax result of Atkinson and Stiglitz (1976) does not apply: Even without any heterogeneity in labor earnings, there is heterogeneity in wealth and capital income and hence scope for a capital income tax.

The wealth-in-the-utility feature puts a limit on individuals' impatience to consume. Intuitively, if $\delta_i > \bar{r}$, with linear consumption and no utility for wealth, the individual would like to consume all his wealth at once at time 0. With utility of wealth, there is value in keeping some wealth. At the margin, the value lost in delaying consumption $\delta_i - \bar{r}$ is equal to the marginal value of holding wealth $a_i'(k)$ and the optimum for capital holding is interior. Note that we need to impose the condition that $\delta_i > \bar{r}$ for all individuals to avoid wealth going to infinity.[11]

Instant adjustments to the steady state and equivalence to the static model. With utility linear in consumption, there are no consumption smoothing considerations. As a result, all dynamic adjustments occur instantaneously and there are no transitional dynamics.

The dynamic model of equation (1) is mathematically equivalent to a static representation. I.e., the optimal choice (c_i, k_i, z_i) from the dynamic problem also maximizes the static utility equivalent:

$$U_i(c_i, k_i, z_i) = c_i + a_i(k_i) - h_i(z_i) + \delta_i \cdot (k_i^{init} - k_i), \tag{3}$$

subject to the static budget constraint $c_i = rk_i + z_i - T(z_i, rk_i)$.

Therefore a social welfare objective based on the original discounted utility V_i from equation (1) is equivalent to a social welfare objective based on the static equivalent U_i from equation (3). It also seems natural to impose a constraint $k \geq 0$ for those who do not like wealth (i.e., who have $a_i(k) \equiv 0$). Such individuals optimally choose $k = 0$ and behave entirely like in the static labor supply model.

Anticipated vs. unanticipated tax reforms and taxing the existing capital stock. A thorny issue in the capital taxation literature has been how to tax the existing stock of capital. Should the government announce reforms in advance or not? This is a policy choice that is itself normative and has important implications for the optimal capital tax. In Section 5, we discuss why none of the approaches adopted so far has been satisfactory. A key advantage of our simpler model is that, with linear utility of consumption and the resulting lack of transitional dynamics, announced and unannounced tax reforms have exactly the same effect. If at time $t = 0$ a capital tax reform is announced to take place at time T, there is no behavioral response until the actual time of the reform. At time T, the capital stock jumps to its new steady level thanks to a Dirac quantum jump in consumption, exactly as in the unannounced tax reform

[11] In practice, wealth does not go to infinity because of shocks to the rate of return or to preferences (Piketty (2011, 2014)). Uncertainty makes the model less tractable. We relate our results to models with stochastic shocks in Section 5.

case. The same optimal taxes apply in the short-run and long-run. As a result, as long as the tax on the return to capital is bounded (e.g. limited to 100%), issues of policy commitment and policy discretion are irrelevant in our model.[12]

2.2 Optimal Tax Formulas

The government sets the time invariant tax $T(z, rk)$, subject to budget-balance, to maximize its social objective:

$$SWF = \int_i \omega_i \cdot U_i(c_i, k_i, z_i) di, \tag{4}$$

where $\omega_i \geq 0$ is the Pareto weight on individual i. We denote by $g_i = \omega_i \cdot U_{ic}$ the social marginal welfare weight on individual i. With utility linear in consumption, we have $g_i = \omega_i$. Without loss of generality, we further normalize the weights to sum to one over the population so that $\int_i \omega_i di = 1$. We first consider linear taxes and then turn to nonlinear taxes.

2.2.1 Optimal Linear Capital and Labor Taxation

We start by studying the optimal linear taxes at rates τ_K and τ_L on capital and labor income. Recall that $\bar{r} \equiv r \cdot (1 - \tau_K)$ denotes the net-of-tax return on capital. The individual maximizing choices are such that $a_i'(k_i) = \delta_i - \bar{r}$ and $h_i'(z_i) = 1 - \tau_L$ so that k_i depends positively on $\bar{r}$ and z_i depends positively on $1 - \tau_L$. For budget-balance, tax revenues are rebated lump-sum and the transfer to each individual is $G = \tau_K \cdot rk^m(\bar{r}) + \tau_L \cdot z^m(1 - \tau_L)$ where $z^m(1 - \tau_L) = \int_i z_i di$ is aggregate labor income that depends on $1 - \tau_L$ and $k^m(\bar{r}) = \int_i k_i di$ is aggregate capital which depends on $\bar{r}$. The government chooses τ_K and τ_L to maximize social welfare SWF in (4), with $c_i = (1 - \tau_K) \cdot rk_i + (1 - \tau_L) \cdot z_i + \tau_K \cdot rk^m(\bar{r}) + \tau_L \cdot z^m(1 - \tau_L)$ and $U_i(c_i, k_i, z_i) = c_i + a_i(k_i) - h_i(z_i) + \delta_i \cdot (k_i^{init} - k_i)$.

Let the elasticity of aggregate capital k^m with respect to $\bar{r}$ be denoted by e_K and the elasticity

[12] There is no temptation to increase the tax rate on capital returns unannounced, as individuals adjust instantaneously, so that the gain from such a tax hike goes to zero. If unanticipated wealth levies are allowed then the capital stock can always be expropriated. In our time continuous model, a wealth levy can be approximated by an infinite tax on capital income for an infinitesimal time. If the capital income tax rate is bounded (say at 100%), wealth levies are ruled out. If wealth levies are anticipated, they can be fully avoided in our model with a suitable Dirac quantum consumption just before the wealth levy followed by a corresponding Dirac quantum saving just after the wealth levy.

of aggregate labor income z^m with respect to the net of tax rate $1-\tau_L$ be e_L. Because there are no income effects, we have $e_L > 0$ and $e_K > 0$. Standard optimal tax derivations using the individuals' envelope theorems for the choice k_i yield:

$$\frac{dSWF}{d\tau_K} = rk^m \cdot \left[\int_i \omega_i \cdot \left(1 - \frac{k_i}{k^m}\right) di - \frac{\tau_K}{1-\tau_K} \cdot e_K \right]$$

The social marginal welfare weight on individual i is $g_i = \omega_i$. At the optimal τ_K, we have $dSWF/d\tau_K = 0$, leading to the following proposition.

Proposition 1. ***Optimal linear capital tax.*** *The optimal linear capital tax is given by:*

$$\tau_K = \frac{1-\bar{g}_K}{1-\bar{g}_K + e_K} \quad \textit{with} \quad \bar{g}_K = \frac{\int_i g_i \cdot k_i}{\int_i k_i} \quad \textit{and} \quad e_K = \frac{\bar{r}}{k^m} \cdot \frac{dk^m}{d\bar{r}} > 0. \tag{5}$$

The optimal labor tax can be derived exactly symmetrically:

$$\tau_L = \frac{1-\bar{g}_L}{1-\bar{g}_L + e_L} \quad \text{with} \quad \bar{g}_L = \frac{\int_i g_i \cdot z_i}{\int_i z_i} \quad \text{and} \quad e_L = \frac{1-\tau_L}{z^m} \cdot \frac{dz^m}{d(1-\tau_L)} > 0. \tag{6}$$

Remarks. The optimal capital tax is zero if $\bar{g}_K = 1$ or $e_K = \infty$. $\bar{g}_K = 1$ happens when there are no redistributive concerns along the capital income dimension (g_i is uncorrelated with k_i).

We discuss social preferences embodied in the social welfare weights g_i in Section 3.1. Briefly, as long as wealth is concentrated among individuals with lower social marginal welfare weights (such that g_i is decreasing in k_i and, hence $\bar{g}_K < 1$) the optimal capital tax is strictly positive.

The revenue maximizing tax rates (which arise when $\bar{g}_K = 0$ and $\bar{g}_L = 0$) are

$$\tau_K^R = \frac{1}{1+e_K} \quad \text{and} \quad \tau_L^R = \frac{1}{1+e_L}. \tag{7}$$

2.2.2 Optimal Nonlinear Separable Taxes

We now turn to the nonlinear tax system separable in labor and capital income, characterized by the tax schedules $T_L(z)$ and $T_K(rk)$. The individual's budget constraint is given by:

$$c_i = rk_i - T_K(rk_i) + z_i - T_L(z_i), \tag{8}$$

so that utility is:

$$U_i(c_i, k_i, z_i) = rk_i - T_K(rk_i) + z_i - T_L(z_i) + a_i(k_i) - h_i(z_i) + \delta_i \cdot (k_i^{init} - k_i). \qquad (9)$$

The first-order conditions characterizing the individual's choice of capital and labor income are:

$$a_i'(k_i) = \delta_i - r(1 - T_K'(rk_i)) \quad \text{and} \quad h_i'(z_i) = 1 - T_L'(z_i).$$

We denote the average relative welfare weight on individuals with capital income higher than rk, by $\bar{G}_K(rk)$ and the average relative welfare weight on individuals with labor income higher than z, by $\bar{G}_L(z)$:

$$\bar{G}_K(rk) = \frac{\int_{\{i:rk_i \geq rk\}} g_i di}{P(rk_i \geq rk)} \quad \text{and} \quad \bar{G}_L(z) = \frac{\int_{\{i:z_i \geq z\}} g_i di}{P(z_i \geq z)}. \qquad (10)$$

Let the cumulative distributions of capital and labor income be $H_K(rk)$ and $H_L(z)$. We denote by $h_K(rk)$ and $h_L(z)$ the corresponding densities when the tax system is linearized at points rk and z.[13] We define the local Pareto parameters of the capital and labor income distributions as:

$$\alpha_K(rk) \equiv \frac{rk \cdot h_K(rk)}{1 - H_K(rk)} \quad \text{and} \quad \alpha_L(z) \equiv \frac{z \cdot h_Z(z)}{1 - H_Z(z)}.$$

Clearly, the income distributions and local Pareto parameters depend on the tax system. The local elasticity of k with respect to the net of tax return $r(1 - T_K'(rk))$ at income level rk is denoted by $e_K(rk)$, while the local elasticity of z with respect to $1 - T_L'(z)$ is denoted by $e_L(z)$.

Because wealth and labor choices are separable, due to the lack of income effects and separable preferences, each tax satisfies the standard Mirrlees (1971) formula and can be expressed in terms of elasticities as in Saez (2001), as shown in the next proposition (the proof is in appendix).

Proposition 2. ***Optimal nonlinear capital and labor income taxes.***

[13] To be precise, these are the densities that would arise if the actual nonlinear tax system were replaced by the linearized tax system at points rk and z. They are related to the actual densities but have the advantage of not being locally affected by the nonlinearity of the tax system, and hence are useful to write more parsimonious optimal tax formulas. Saez (2001) introduced this concept to simplify the presentation of optimal nonlinear labor income taxation and provides complete details. Our analysis exactly parallels Saez (2001) analysis.

The optimal nonlinear capital and labor income taxes are:

$$T'_K(rk) = \frac{1-\bar{G}_K(rk)}{1-\bar{G}_K(rk)+\alpha_K(rk)\cdot e_K(rk)} \quad \textit{and} \quad T'_L(z) = \frac{1-\bar{G}_L(z)}{1-\bar{G}_L(z)+\alpha_L(z)\cdot e_L(z)}. \qquad (11)$$

Asymptotic Nonlinear Formula. In Section 4 we show that capital income is very concentrated, with top 1% capital income earners earning more than 60% of total capital income. The asymptotic formula when $rk \to \infty$ in (11) is therefore relevant for most of the tax base.

$$T'_K(\infty) = \frac{1-\bar{G}_K(\infty)}{1-\bar{G}_K(\infty)+\alpha_K(\infty)\cdot e_K(\infty)}. \qquad (12)$$

The revenue maximizing rate obtains if $\bar{G}_K(\infty) = 0$.

Optimal linear tax rate in top bracket. Paralleling Saez (2001) optimal labor income analysis, it is also easy to derive a formula for the optimal linear tax rate in the top bracket above a given capital income threshold:

$$\tau_K^{top} = \frac{1-\bar{g}_K^{top}}{1-\bar{g}_K^{top}+a_K^{top}\cdot e_K^{top}}$$

with $\bar{g}_K^{top}$ the average social marginal welfare weight in the top bracket, e_K^{top} the elasticity in the top bracket, and a_K^{top} the Pareto parameter in the top bracket. The Pareto parameter is defined as $a_K^{top} = \frac{E[k_i|k_i \geq k^{top}]}{E[k_i|k_i \geq k^{top}]-k^{top}}$ where k^{top} is the threshold for the top bracket. As capital income is so concentrated, this formula has wide applicability (see our numerical simulations below).

2.3 Foundations of Wealth in the Utility

Wealth in the utility is needed technically in order to generate finite steady state elasticities, but is also attractive *per se* because it helps generate a non degenerate steady state even with arbitrary heterogeneity in discount rates or returns and allows to better match the data by capturing relevant non consumption benefits of capital.[14] That there must be benefits from

[14]Technically, the standard dynamic model with only utility for consumption leads to a degenerate steady state, where $\delta_i = \delta = \bar{r}$. This precludes heterogeneity in time preferences and implies an infinite elasticity of capital to taxes in the steady-state. Introducing utility for wealth is, however, not the only way and our derived tax formulas – expressed in terms of sufficient statistics – do not depend on it. Indeed, in section 5, we discuss two other assumptions used in the literature to obtain non-degenerate (and more realistic) responses of capital

wealth other than consumption was already recognized by Weber, Keynes, and Smith among others. Max Weber called the phenomenon of individuals valuing wealth *per se* the "capitalist spirit" (Weber, 1958).[15] Keynes (1931) regretted people's "love of money as a possession." In Keynes (1919), he also lamented that "the duty of saving became nine-tenths of virtue and the growth of the cake the object of true religion," the cake being total wealth. Even more important was his observation that saving was seemingly only done for the sake of holding wealth. "Saving was for old age or for your children; but this was only in theory–the virtue of the cake was that it was never to be consumed, neither by you nor by your children after you."

Already Smith (1759) lamented that wealth could lend social status and moral prestige.[16] Wealth can be used by people as a very visible – even ostentatious– signal of one's innate abilities and strengths.[17] Conspicuous consumption is one example of an attempt to signal status and wealth to others, presumably because there are benefits from being perceived as wealthy.[18]

The assumption of wealth in the utility can be micro-founded and justified by the empirical evidence. Wealth in the utility ultimately means relaxing the restrictive assumption that wealth only brings utility benefits through the sheer consumption flow that can be bought with it. While wealth will eventually be consumed by oneself or one's heirs, there are "warm glow" or "joy of ownership" benefits from having it even without spending it, which limit the impatience to consume it.

There is no compelling empirical evidence that a model with only utility for consumption captures microeconomic behavior better than the model with wealth in the utility. Quite the contrary, it has been shown that the standard Bewley models as in Aiyagari (1994) cannot match the empirical wealth distribution. First, precautionary savings in themselves cannot

to taxes: introducing uncertainty, as in Aiyagari (1995), or consumption-dependent discount rates $\delta_i(c_i)$ as in Judd (1985). As argued in this section, a model with only heterogeneity or stochasticity in labor earnings does not fit the data well.

[15] Weber (1958) viewed it as a result of Protestant values promoting saving, frugality, and capital accumulation.

[16] "This disposition to admire, and almost to worship, the rich and the powerful, and to despise, or, at least, to neglect persons of poor and mean condition, [...] is, at the same time, the great and most universal cause of the corruption of our moral sentiments. That wealth and greatness are often regarded with the respect and admiration which are due only to wisdom and virtue; and that the contempt, of which vice and folly are the only proper objects, is often most unjustly bestowed upon poverty and weakness, has been the complaint of moralists in all ages."

[17] Social status concerns due to wealth may lead to externalities and to corrective taxation, which could be an interesting extension for future research.

[18] Christophera and Schlenker (2000) show in a randomized experiment, that people perceived to be wealthier are also perceived to be more able and talented (see also Dittmar (1992)).

rationalize high wealth holdings at the top without "the capitalist spirit" (Carroll, 1997, 2000; Quadrini, 1999). Second, it is difficult to generate a saving behavior that makes the distribution of wealth much more concentrated than that of labor earnings (Benhabib and Bisin, 2016).[19] Third, as we show in Section 4, there is an important two-dimensional heterogeneity in capital and labor income: even conditional on labor income, capital income is unequally distributed, which means a second dimension of heterogeneity, in addition to differences in labor earnings ability, is required.

We next discuss formally three possible microfoundations for wealth in the utility.

2.3.1 Bequest Motive

The wealth in the utility specification can arise from bequest motives. With a warm-glow bequest motive, if an agent dies at date T, his utility is:

$$V_i(T) = \int_0^T u_i(c_i(t))e^{-\rho_i t}dt + e^{-\rho_i T}\phi_i(k_i(T)), \tag{13}$$

where ρ_i is the discount rate of agent i and $\phi_i(k_i(T))$ is the warm glow utility from the bequest $k_i(T)$ left at time T. If the death time T is stochastic and follows a Poisson process with rate p_i for agent i, then, as in the "perpetual youth" model of Yaari (1965) and Blanchard (1985), the expected utility can be rewritten in infinite horizon with:

$$V_i = \int_0^\infty e^{-(\rho_i+p_i)t} \cdot \left[u_i(c_i(t)) + p_i \cdot \phi_i(k_i(t))\right] dt. \tag{14}$$

This amounts to our wealth in the utility formulation with $\delta_i = \rho_i + p_i$ and $a_i(k_i(t)) = p_i \cdot \phi_i(k_i(t))$.

On the empirical side, De Nardi (2004) shows quantitatively that a model with a bequest motive can both explain large wealth holdings at the top and better match the lifecycle profiles of savings. Cagetti and De Nardi (2007) combine a bequest motive with a model of entrepreneurship, also discussed next.

[19] That households want to keep wealth for purposes other than consumption is also suggested by behavior in retirement: very little wealth is annuitized, especially among the very wealthy, many assets are still available at death, and indeed, wealthy households do not appear to be rapidly de-accumulating wealth closer to their death.

2.3.2 Entrepreneurship

Wealth in the utility can also arise from a model of entrepreneurship. Entrepreneurship has been used as a key explanatory factor for the shape of the wealth distribution according to Quadrini (1999, 2000).[20]

In this model, there is a utility flow from running a business, which captures the non-pecuniary private benefits net of the effort or disutility costs of being an entrepreneur. Non-pecuniary benefits or costs from entrepreneurship have been shown to be substantial and important explanations for occupational choice (Hamilton, 2000; Hurst and Pugsley, 2010). Entrepreneur i receives a return on their capital r_i (Section 3.3.2 deals formally with heterogeneous returns).[21] For instance, if $a_i(k) = \eta_i k_i^\gamma/\gamma$ with $\gamma < 1$, and there is a linear tax on capital τ_K, entrepreneur i would chose a capital level such that: $r_i(1 - \tau_K) = \delta_i - \eta_i k_i^{\gamma-1}$.

More generally, the wealth in the utility specification can apply to agents managing a wealth portfolio. This is an activity which entails not only a financial return, but also potential non-pecuniary benefits and/or time and effort costs.

2.3.3 Service Flows From Wealth

Capital is embodied in tangible or financial assets, which yield service flows. One salient example is housing, which yields a stream of utility in terms of housing services. But even financial assets provide utility in the form of security or potential liquidity beyond and above their financial return.

In that sense, our model resembles "money in the utility" models. Money is a special asset that yields zero nominal return and high liquidity. Other capital assets have different liquidities and returns. As explained in Poterba and Rotemberg (1987), wealth held in the form of different assets is akin to other durables in that it provides services (e.g., security or liquidity) even when it is not consumed. Whether those services from wealth enter utility the exact same way as other durable goods is an empirical question that would merit careful estimation of utility functions. Poterba and Rotemberg (1987) argue that "many goods provide different 'types' of utility" and that to single out wealth services as being "unworthy of inclusion in a consumer's utility seems arbitrary at best."

[20]A useful extension for future research would be to have stochastic returns to capital.

[21]It is possible for $a_i(k)$ to be on net negative in which case we need $\delta_i < r_i$.

The utility flows from assets are widely documented in the finance literature as being needed to better fit the financial data. Examples of papers which model housing capital as both an asset with returns and as a consumption good providing utility flows are Piazzesi et al. (2007), Stokey (2009), and Kiyotaki et al. (2011). The latter specifically assigns a different utility to renting a house and owning a house (e.g., the owner can modify the house to fit their own taste, which yields utility), which is exactly in the spirit of our specification.

In Section 3.3.3, we explicitly consider differentiated taxation of various types of capital assets.

3 Policy Issues

We now put this newly gained simplicity to use, and consider how our framework can shed light on several salient issues in the public debate about capital taxation.

3.1 Ethical Considerations

We start by discussing four ethical standpoints which are often encountered in the public debate, and what level of capital tax they would imply. We can do this because our approach in terms of sufficient statistics is very amenable to the use of *generalized social welfare weights* g_i as in Saez and Stantcheva (2016), which can better capture the normative considerations which are relevant for capital taxation.[22]

Inequality in wealth deemed unfair. Inequality in wealth is viewed by some as unfair as it arises from preferences for wealth, higher patience, or higher returns on capital. Higher patience could for instance be considered a skill that allows some individuals to save more and be better off in the long-run, much in the same way that a higher earning ability allows people to earn more and be better off in the traditional optimal labor income tax model. Higher returns on capital could be perceived as "luck." In that case, redistributing from wealth lovers to non-

[22] The generalized social welfare weights are given by $g_i = g(c_i, k_i, z_i; x_i^b, x_i^s)$ where x_i^b is a vector of characteristics which enter both utility and the weights, while x_i^s is a vector of characteristics that only enters the weights. This allows to introduce a gap between individual preferences and social considerations. Hence, it allows for a wider range of normative considerations to be taken into consideration than with standard welfare weights.

wealth lovers could be deemed socially desirable.[23] Social welfare weights g_i are then decreasing in k_i. For linear taxes, then, $\bar{g}_K < 1$ and $\tau_K > 0$.

Inequality in wealth deemed fair. Conversely, some may consider inequality in wealth fair and irrelevant for redistribution. In this case, social welfare weights do not depend on k_i and are uncorrelated with k_i. People supporting this view may argue that higher wealth comes from a higher taste for savings (rather than consuming). It is through the sacrifice of earlier consumption that an individual has accumulated wealth. There is no compelling reason to redistribute "from the ant to the grasshopper" because the grasshopper had the same opportunity to save. In this case, if we further assume that wealth k_i is uncorrelated with other characteristics affecting social welfare weights (see discussion just below), then $\bar{g}_K = 1$ and $\tau_K = 0$.

Wealth as a tag. Wealth can be a marker and tag for a characteristic that society cares about, but that taxes cannot directly condition on. In this case, g_i may not depend on k_i directly (as discussed in the previous paragraph), but is correlated with k_i, leading to $\bar{g}_K \neq 1$. For instance, society may care about equality of opportunity and may want to compensate people from poorer backgrounds for their difficult start in life. Even if society does not care about tastes for wealth and wealth per se, higher wealth could be a tag for a richer family background. For example and following Saez and Stantcheva (2016), if $g_i = 1$ for people from a low background and is zero for others, then $\bar{G}_K(rk)$, the average social welfare weight on those with capital income above rk will be the representation index of those from a low background among individuals with capital income above rk. If people with high capital income come disproportionately from wealthy backgrounds, then $\bar{G}_K(rk)$ is less than one, leading to a positive nonlinear capital income tax rate using formula (11).

Similarly, wealth can be a tag for earnings ability. Suppose there is inequality in both capital and labor income, but that the government only cares about the latter, so that g_i only depends on z_i and $T_L(z_i)$. If capital and labor income are uncorrelated, then $\bar{g}_K = 1$ and the optimal τ_K is zero. If they are positively correlated, then $\bar{k} < 1$ and hence $\tau_K > 0$: in this case, high wealth individuals also have higher labor income on average, and wealth acts as a form of tag.[24]

[23] The case for this argument may be even stronger if wealth comes from inheritances.

[24] Gordon and Kopczuk (2014) propose an analysis both theoretical and empirical along this line.

Horizontal equity concerns. Horizontal equity concerns mean that society does not want to treat differently people with the same “ability to pay.” The key issue, which involves non-trivial value judgements, is to define “ability to pay”. It could be total income, capital income, labor income, or even the consumption of some particular goods. For instance, should ability to pay be measured by labor income only?

On the affirmative side are those who criticize the “double taxation” of income, first in the form of earned labor income and then in the form of an additional tax on capital income earned on savings out of labor income. In addition “equality of opportunity” type of arguments for savings (as opposed to equality of outcomes, in analogy to labor taxation) state that conditional on a given labor income, everybody has the same opportunities to save (assuming everybody starts with zero wealth). This is the view that the grasshopper and the ant, with the same labor income, simply made different choices the consequences of which they have to bear.

On the negative side, an increase in returns on assets more generally would benefit savers and, conditional on a given labor income, individuals with a strong preference for wealth could end up with much higher incomes if the rate of return on capital is high. Indeed, in conceptual debates about the desirability of taxing capital income in the law and economics literature on taxation, proponents of the capital tax tend to use high rate of return scenarios (e.g., Warren (1980)) while opponents tend to use low rate of return scenarios (e.g., Weisbach and Bankman (2006)).

Overall, the most natural concept seems total income $y = z + rk$. A higher return on capital r is an advantage for wealth lovers, but this advantage is taken into account by the comprehensive income concept. With strong horizontal equity preferences, this justifies the comprehensive income tax (barring a Pareto improvement of providing a component specific tax break) (see Online Appendix A.5).

3.2 Choosing the Tax Base

3.2.1 Comprehensive Income Tax System $T(z + rk)$

An important policy question is how the tax rate should be set if all income – whether stemming from capital or labor– were to be treated the same way for tax purposes. In many countries, most “ordinary” capital income, such as interest from a standard savings account, is taxed jointly with labor income by the individual income tax. Within our framework, we can easily solve for the optimal nonlinear tax on comprehensive income $y \equiv rk + z$, of the form $T_Y(y)$, i.e., for the optimal system within the class of tax systems that treat capital and labor income perfectly symmetrically. We then discuss when such a tax system is optimal. In this case, the optimal tax formula turns out to take the same form as in Mirrlees (1971) and Saez (2001).

Define the average welfare weight on individuals with total income higher than y as:

$$\bar{G}_Y(y) = \frac{\int_{\{i:y_i \geq y\}} g_i di}{P(y_i \geq y)}. \tag{15}$$

Let $H_Y(y)$ be the cumulative distribution of the total income distribution and $h_Y(y)$ the associated density (assuming again a linearized tax system at point y). Let $\alpha_Y(y) \equiv \frac{y h_Y(y)}{1-H_Y(y)}$ be the local Pareto parameter for the distribution of total income y and $e_Y(y)$ is the elasticity of total income to the net of tax rate $1 - T'_Y(y)$ at point y.

Using the envelope theorem, we obtain a standard optimal tax formula on full income.

Proposition 3. ***Optimal tax on comprehensive income.***

(i) The optimal nonlinear *tax on comprehensive income $y = rk + z$ is given by:*

$$T'_Y(y) = \frac{1 - \bar{G}_Y(y)}{1 - \bar{G}_Y(y) + \alpha_Y(y) \cdot e_Y(y)}.$$

(ii) The optimal linear *tax on comprehensive income is:*

$$\tau_Y = \frac{1 - \bar{g}_Y}{1 - \bar{g}_Y + e_Y}. \tag{16}$$

$$\textit{with} \quad \bar{g}_Y \equiv \frac{\int_i g_i y_i}{y^m} = \frac{z^m \bar{g}_L + rk^m \bar{g}_K}{z^m + rk^m} \quad \textit{and} \quad e_Y \equiv \frac{dy^m}{d(1-\tau_Y)} \frac{(1-\tau_Y)}{y^m} = \frac{z^m e_L + rk^m e_K}{z^m + rk^m}. \tag{17}$$

A tax system based on comprehensive income may be optimal for equity reasons (discussed in Section 3.1) or for efficiency reasons, due to the existence of income shifting opportunities between the capital and labor income bases, which we discuss next.

3.2.2 Income Shifting

A salient issue in the policy debate is the possibility of shifting income between the labor and capital bases (see e.g., Saez et al. (2012) for a survey of the empirical literature).

To model this, suppose that individuals can shift an amount of labor income x from the labor to the capital tax base at a utility cost $d_i(x)$, increasing and convex in x. Hence, if reported labor income at time t is $z_i^R(t)$, we have $x_i(t) = z_i(t) - z_i^R(t)$. The aggregate shifted amount at time t is $x^m(t) \equiv \int_i x_i(t)di$. We consider linear taxes in this section.

We can easily show that in this case again, the dynamic and static problems are equivalent. The discounted normalized utility of individual i,

$$V_i(\{c_i(t), k_i(t), z_i(t), x_i(t)\}_{t\geq 0}) = \delta_i \cdot \int_0^\infty [c_i(t) + a_i(k_i(t)) - h_i(z_i(t)) - d_i(x_i(t))]e^{-\delta_i t}dt,$$

under the budget constraint:

$$\dot{k}_i(t) = \bar{r}k_i(t) + (1-\tau_L)z_i(t) - c_i(t) + (\tau_L - \tau_K)x_i(t) + \tau_L(z^m(t) - x^m(t)) + \tau_K(rk^m(t) + x^m(t)),$$

is equivalent to the static model:

$$U_i(c, k, z, x) = c + a_i(k) - h_i(z) - d_i(x) + \delta_i \cdot (k^i_{init} - k),$$

subject to the static budget constraint $c = \bar{r}k + (1-\tau_L)z + (\tau_L - \tau_K)x + \tau_L(z^m - x^m) + \tau_K(rk^m + x^m)$. This static model of tax shifting was analyzed in Piketty et al. (2014), as well as in the Piketty and Saez (2013a) survey of the static optimal income tax literature. Our analysis shows that this static model is actually consistent with a dynamic model of savings. The individual's choice is characterized by the following conditions:

$$h_i'(z_i) = 1 - \tau_L \quad \text{and} \quad a_i'(k_i) = \delta_i - \bar{r},$$

$$d_i'(x_i) = \tau_L - \tau_K \quad \text{and} \quad c_i = \bar{r}k_i + (1-\tau_L)z_i + (\tau_L - \tau_K)x_i + \tau_L(z^m - x^m) + \tau_K(rk^m + x^m).$$

Hence, labor income z_i is a function of the net-of-tax rate $1-\tau_L$, capital income rk_i is a function of the net-of-tax return $\bar{r}$, and shifted income x_i is a function of the tax differential $\Delta\tau \equiv \tau_L - \tau_K$.

In the same way that we previously defined the distributional factors for capital and labor income in (5) and (6), we can define the distributional factor for shifted income as: $\bar{g}_X = \int_i \omega_i x_i / z^m$. As long as the distributional factor $\bar{g}_X$ is small enough (in a way made precise in the proof in the Appendix) so that allowing income shifting is not an attractive way of redistributing income, we have the following results.

Proposition 4. ***Optimal Labor and Capital Taxes with Income Shifting.***

i. If $e_K > e_L \cdot \frac{1-\bar{g}_K}{1-\bar{g}_L}$*, then* $\frac{1-\bar{g}_L}{1-\bar{g}_L+e_L} \geq \tau_L > \tau_K \geq \frac{1-\bar{g}_K}{1-\bar{g}_K+e_K}$ *and conversely, if* $e_K < e_L \cdot \frac{1-\bar{g}_K}{1-\bar{g}_L}$*, then* $\frac{1-\bar{g}_L}{1-\bar{g}_L+e_L} \leq \tau_L < \tau_K \leq \frac{1-\bar{g}_K}{1-\bar{g}_K+e_K}$.

ii. If there is no shifting, the linear tax rates are set according to their usual formulas in (5) and (6).

iii. If shifting is infinitely elastic, then the tax differential $\Delta\tau$ *goes to 0 and* $\tau_K = \tau_L = \tau_Y = \frac{1-\bar{g}_Y}{1-\bar{g}_Y+e_Y}$ *where* $\bar{g}_Y = \frac{z^m \bar{g}_L + rk^m \bar{g}_K}{z^m + rk^m}$ *is the distributional factor of total income, and* $e_Y = \frac{z^m e_L + rk^m e_K}{z^m + rk^m}$ *is the elasticity of total income.*

Thus as long as there is shifting with a finite elasticity the labor and capital taxes are compressed toward each other, away from their optimal values with no shifting. With an infinite shifting elasticity, the optimum is to set a comprehensive tax on full income $y = rk + z$, as solved for in (16). Strong shifting opportunities, with elasticities tending to infinity, provide a justification for a tax based on total comprehensive income which is orthogonal to the social ethical considerations discussed in Section 3.1.

3.2.3 Consumption taxation

Can a consumption tax achieve more redistribution than a wealth tax and be more progressive than a tax on labor income? Our simple model allows us to cleanly assess the role of and the scope for a consumption tax.

Let us define real wealth as wealth expressed in terms of purchasing power, or, equivalently, wealth as normalized by the price of consumption. It seems natural that individuals should care about real wealth rather than nominal wealth for the real economic power or status that it confers. As long as individuals care about real wealth, a consumption tax is equivalent to a tax on labor income augmented with a tax on initial wealth as in the standard model with no utility for wealth (see e.g., Auerbach and Kotlikoff (1987); Kaplow (1994); Auerbach (2009)). Hence the consumption tax cannot achieve a more equal steady state than the labor tax. In the simplest case with a linear consumption tax, it is immediate to see this equivalence.[25]

If the tax exclusive rate is t_C so that the implied price of consumption is $1+t_C$ the equivalent tax inclusive rate is τ_C which is such that $1-\tau_C = 1/(1+t_C)$. Real wealth is $k^r = k\cdot(1-\tau_C)$ and flow utility is $u_i = c + a_i(k^r) - h_i(z)$. The budget constraint of the individual becomes $\dot{k} = [\bar{r}k + z - T_L(z)] - c/(1-\tau_C) + G$, where $G = \tau_L z^m + \tau_K r k^m + t_C c^m$ is the lump-sum transfer rebate of tax revenue. The budget constraint can be rewritten in terms of real wealth as: $\dot{k}_r = \bar{r}k^r + (z - T_L(z))\cdot(1-\tau_C) + G\cdot(1-\tau_C) - c$.

In real terms, the consumption tax τ_C then just adds a layer of taxes on labor income, leaving $\bar{r}$ unchanged. For the individual, the steady state (i.e., the static model) $(\bar{r}, T_L, \tau_C)$ is equivalent to $(\bar{r}, \hat{T}_L, \tau_C = 0)$ with $\hat{T}_L$ such that $z - \hat{T}_L(z) = (z - T_L(z))\cdot(1-\tau_C)$.

The difference between these two tax systems is that consumption taxation also taxes initial wealth by reducing its real value from k_i^{init} to $k_i^{r,init} = (1-\tau_C)\cdot k_i^{init}$. This means that a consumption tax does successfully tax initial wealth, but has no long term effect on the distribution of real wealth. If the government undoes this initial wealth redistribution by giving a lump-sum transfer $\tau_C \cdot k_i^{init}/(1-\tau_C)$ to an individual i with initial wealth holdings k_i^{init}, the equivalence between a consumption tax system $(\bar{r}, T_L, \tau_C)$ and a modified labor tax system with no consumption tax $(\bar{r}, \hat{T}_L, \tau_C = 0)$ becomes fully complete in the dynamic consumer problem, the steady-state of the consumer, and the intertemporal government budget. Hence we have:

Proposition 5. ***Equivalence of consumption taxes and labor taxes.*** *A linear consumption at inclusive rate τ_C is equivalent to a tax on labor income combined with a tax on initial wealth.*

To refute a common fallacy on the redistributive power of consumption taxes, suppose that there is no initial wealth (and, hence, no need for a compensating transfer if a consumption tax were to be introduced) and that labor income is inelastic and uniform across individuals. Differences in wealth then only arise from differences in tastes for wealth. It is clear that a pure labor income tax achieves no redistribution in this setting. It just taxes the inelastic and

[25] With a progressive consumption tax, the equivalence is less immediate, but nevertheless present and we consider this case in Online Appendix A.6.

equal labor income and rebates it back as an equal lump-sum transfer to all individuals. If there were a consumption tax in this setting, those with higher preferences for wealth would end up having higher income, higher consumption, and pay higher taxes than those with lower preferences for wealth in steady state. But recall that the consumption tax is fully equivalent to the labor income tax in this setting and that the labor income tax achieves no redistribution. Thus, while wealth lovers look like they pay higher taxes in the steady state on their higher consumption, this is because they paid less taxes while building up their wealth at instant 0. This initial wealth accumulation is what gives them higher steady state consumption in the first place. Wealth lovers build up more nominal wealth with consumption taxation so that their real wealth is the same as under the equivalent labor income tax (and no consumption tax).

It is hence important to draw a distinction between the observed cross-section and the lifetime distribution of resources. In our simple model, in the cross-sectional steady-state, the consumption tax looks redistributive, when, in reality, it is not.

3.3 Extensions

3.3.1 Jointness in preferences for labor and capital

There could be jointness in the preferences for work and wealth, which introduces cross-elasticities between the capital and labor taxes. It is indeed reasonable to think that work incentives could be affected by wealth.

The discounted utility is:

$$V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}) = \delta_i \int_0^{\infty} [c_i(t) + v_i(k_i(t), z_i(t))] e^{-\delta_i t} dt, \tag{18}$$

with $v_i(k, z)$ increasing concavely in k and decreasing concavely in z. With linear taxes τ_K and τ_L, the budget constraint of individual i is:

$$\frac{dk_i(t)}{dt} = \bar{r} k_i(t) + (1 - \tau_L) \cdot z_i(t) + r\tau_K k^m(t) + \tau_L z^m(t) - c_i(t).$$

The choice $(c_i(t), k_i(t), z_i(t))$ for individual i at any time $t > 0$ is such that:

$$-v_{iz}(k_i(t), z_i(t)) = 1 - \tau_L, \qquad v_{ik}(k_i(t), z_i(t)) = \delta_i - \bar{r},$$

$$\text{and} \quad c_i(t) = \bar{r}k_i(t) + (1-\tau_L)\cdot z_i(t) + r\tau_K k^m(t) + \tau_L z^m(t).$$

The dynamic model is again equivalent to the static specification:

$$U_i(c_i, k_i, z_i) = c_i + v_i(k_i, z_i) + \delta_i(k_i^{init} - k_i).$$

Denote by $e_{L,(1-\tau_K)} \equiv \frac{1-\tau_K}{z^m} \cdot \frac{dz^m}{d(1-\tau_K)}$ the cross-elasticity of average labor income to the net-of-tax return and by $e_{K,(1-\tau_L)} \equiv \frac{1-\tau_L}{rk^m} \cdot \frac{d(rk^m)}{d(1-\tau_L)}$ the cross-elasticity of average capital income to the net-of-tax labor tax rate.

Proposition 6. ***Optimal labor and capital taxes with joint preferences.*** *With joint preferences, the optimal linear capital tax (respectively, labor tax) taking the labor tax (respectively, the capital tax) as given is:*

$$\tau_K = \frac{1 - \bar{g}_K - \tau_L \frac{z^m}{rk^m} e_{L,(1-\tau_K)}}{1 - \bar{g}_K + e_K} \quad \textit{and} \quad \tau_L = \frac{1 - \bar{g}_L - \tau_K \frac{rk^m}{z^m} e_{K,(1-\tau_L)}}{1 - \bar{g}_L + e_L}. \tag{19}$$

The formula for each tax applies even if the other tax is not optimally set. The effects of jointness in preferences on the optimal labor and capital taxes depend on the complementarity or substitutability of preferences for capital and labor. If having more capital decreases the cost of work, then $e_{L,(1-\tau_K)} > 0$ and, at any given τ_L, the capital tax should optimally be set lower. It is possible to combine the optimal formulas for τ_L, τ_K in (19) to solve for the jointly optimal τ_K, τ_L.[26]

3.3.2 Heterogeneous Returns to Capital

In practice, individuals may have very different returns on their wealth. Financially savvy people may be able to hold optimized portfolios with higher returns for instance. Higher wealth individuals empirically seem to reap a higher return, potentially because of smarter investments or economies of scale in financial management (Piketty, 2014). Entrepreneurs investing their capital in a business may have different abilities for running their business and generating returns.

[26] As is the case with optimal tax formulas in general, the elasticities and distributional factors $\bar{g}_K, \bar{g}_L$ are implicit functions of τ_L, τ_K. Therefore this would still be an implicit formula for optimal tax rates.

With heterogeneous returns to capital, the full dynamic model with utility as in (1) subject to the budget constraint in (2), where r is replaced by a heterogeneous return r_i is again equivalent to the same static model as above, with the following budget constraint: $c_i = r_i(1-\tau_K)k_i + (1-\tau_L)z_i + \tau_K \int_i r_i k_i(\bar{r}_i) + \tau_L z^m(1-\tau_L)$.

At the optimal τ_K, we have $dSWF/d\tau_K = 0$, so that:

$$\tau_K = \frac{1-\bar{g}_{rK}}{1-\bar{g}_{rK}+e_{rK}} \quad \text{with} \quad \bar{g}_{rK} = \frac{\int_i g_i \cdot r_i k_i}{\int_i r_i k_i} \quad \text{and} \quad e_{rK} = \frac{(1-\tau_K)}{\int_i r_i k_i} \cdot \frac{d\int_i r_i k_i}{d(1-\tau_K)} > 0.$$

Heterogeneous returns do not affect the formula in terms of sufficient statistics, $\bar{g}_{rK}$ and e_{rK}. However, they may affect our ethical judgments on taxes, especially if there is a systematic correlation (as discussed in Piketty (2014)) between wealth and the return on wealth.

Different returns on capital could be perceived as unfair: for a given amount of sacrificed consumption, some individuals reap higher returns, much like for a given amount of sacrificed leisure, some individuals reap a higher labor income in the standard labor tax model. Redistribution across individuals with different returns may then be perceived as desirable, even conditional on total capital income.[27]

3.3.3 Different Types of Capital Assets

Another issue which would be very difficult to handle in standard dynamic capital tax models is that, in practice, there is not just one single type of capital, but rather different assets, with different liquidity and payoff patterns. Moreover, individuals may have heterogeneous tastes for different assets. Our model is flexible enough to incorporate different types of capital assets and heterogeneous preferences for them. Thanks to the direct utility component for wealth here, we can rationalize why people would hold assets with different returns above and beyond the standard risk-return trade-off considerations. For instance, a home can yield direct utility benefits. Government bonds or shares in one's own company may also have an individual-specific value, if people care about the national or company-specific contribution that their capital makes.

Consider J assets with different returns denoted generically by r^j, taxes τ_K^j, and net-of-tax

[27] Put differently, someone with a high r_i (a "luck" shock) should be deemed less deserving than someone with a high k_j (a higher consumption sacrifice) conditional on $r_i k_i = r_j k_j$.

return $\bar{r}^j$. IindividualIndividual i holds a level k_i^j of asset j, with initial level $k_i^{init,j}$. For simplicity, assume exogenous and uniform labor income z. The static utility equivalent for individual i can feature joint preferences in the assets:

$$U_i = c_i + a_i(k_i^1, .., k_i^J) + \delta_i \cdot \sum_{j=1}^{J} (k_i^{init,j} - k_i^j),$$

with the budget constraint:

$$c_i = \sum_{j=1}^{J} \bar{r}^j k_i^j + z + \sum_{j=1}^{J} \tau_K^j r^j k^{m,j}.$$

It is straightforward to derive the tax rates on each asset, analogous to the formula for capital and labor taxes with joint preferences in Section 3.3.1:

Proposition 7. ***Different types of capital with heterogeneous, joint preferences.*** *The optimal tax on capital asset j, given all other tax rates τ_K^s for $s \neq j$ (not necessarily optimally set) is given by:*

$$\tau_K^j = \frac{1 - \bar{g}_K^j - \sum_{s \neq j} \tau_K^s \frac{k^{m,s}}{k^{m,j}} e_{K^s,(1-\tau_K^j)}}{1 - \bar{g}_K^j + e_K^j} \tag{20}$$

$$\text{with} \quad \bar{g}_K^j = \frac{\int_i g_i \cdot k_i^j}{\int_i k_i^j}, \quad e_K^j = \frac{\bar{r}^j}{k^{m,j}} \cdot \frac{dk^{m,j}}{d\bar{r}^j} > 0, \quad \text{and} \quad e_{K^s,(1-\tau_K^j)} = \frac{\bar{r}^j}{k^{m,s}} \cdot \frac{dk^{m,s}}{d\bar{r}^j}. \tag{21}$$

The tax on each type of capital asset is first determined by the two standard considerations of equity and efficiency. Indeed with no cross-elasticities (which arises with separability $a_i(k_i^1, ...k_i^J) = \sum_{j=1}^J a_i^j(k_i^j)$) the formulas are simply:

$$\tau_K^j = \frac{1 - \bar{g}_K^j}{1 - \bar{g}_K^j + e_K^j}.$$

Assets with higher elastiticities (e_K^j) should be taxed less. Those with a higher redistributive impact, i.e., for which holdings are concentrated among high welfare weight individuals ($\bar{g}_K^j$ high) should be taxed less, all else equal. Conversely, assets equally distributed ($\bar{g}_K^j \approx 1$) should not be taxed much for redistributive purposes. Society may have very different value judgements regarding different assets, embodied in very different weights $\bar{g}_K^j$, leading to different optimal tax rates.

Second, the efficiency cost of taxing asset j depends on its cross-elasticities with other assets and hence its fiscal spillovers to the other assets' tax bases. If the asset is complementary to many other assets the efficiency cost of taxing it may be much larger than the own-price elasticity.

In addition, if the government cannot freely optimize the tax rate on some asset s, then, when asset j and asset s are complements ($e_{K^s,(1-\tau_K^j)} > 0$), the higher existing tax on asset s would push towards a lower optimal tax on asset j.

3.4 Economic Growth and the Aggregate Capital Stock

3.4.1 Economic Growth

How would economic growth affect the optimal capital tax rate? Suppose that there is technological progress at an exogenous rate $g > 0$, leading to economic growth, so that all per capita variables grow at rate $g > 0$. We can perform the normalization and denote normalized variables with a tilda: $\tilde{z}(t) = z(t)e^{-gt}, \tilde{k}(t) = k(t)e^{-gt}, \tilde{c}(t) = c(t)e^{-gt}$. To sustain a balanced growth path with quasi-linear utility, the sub-utility functions need to take the form $h_{ti}(z(t)) = e^{gt} \cdot h_i(\tilde{z}(t))$ and $a_{ti}(k(t)) = e^{gt} \cdot a_i(\tilde{k}(t))$. We also assume that $T_t(z(t), rk(t)) = e^{gt} \cdot T(\tilde{z}(t), r\tilde{k}(t))$.

The discounted normalized utility should now be written as:

$$
\begin{aligned}
V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}) &= \delta_i \cdot \int_0^\infty [c_i(t) + a_{ti}(k_i(t)) - h_{ti}(z_i(t))]e^{-\delta_i t}dt \\
&= \delta_i \cdot \int_0^\infty [\tilde{c}_i(t) + a_i(\tilde{k}_i(t)) - h_i(\tilde{z}_i(t))]e^{-(\delta_i - g)t}dt.
\end{aligned}
$$

The budget constraint of individual i is:

$$
\dot{k}_i(t) = rk_i(t) + z_i(t) - T(z_i(t), rk_i(t)) - c_i(t) \quad \text{i.e.} \quad \dot{\tilde{k}}_i(t) = (r-g)\tilde{k}_i(t) + \tilde{z}_i(t) - T(\tilde{z}_i(t), r\tilde{k}_i(t)) - \tilde{c}_i(t).
$$

Hence, this problem is mathematically equivalent to our earlier problem. Similar derivations show that the normalized solution $(\tilde{c}_i, \tilde{k}_i, \tilde{z}_i)$ for individual i at any time $t > 0$ is such that:

$$h_i'(\tilde{z}_i) = 1 - T_L'(\tilde{z}_i, r\tilde{k}_i) \quad \text{and} \quad a_i'(\tilde{k}_i) = \delta_i - r(1 - T_K'(\tilde{z}_i, r\tilde{k}_i)) \quad \text{and} \quad \tilde{c}_i = (r-g)\tilde{k}_i + \tilde{z}_i - T(\tilde{z}_i, r\tilde{k}_i).$$

The actual levels of $(c_i(t), k_i(t), z_i(t))$ are then simply equal to: $(\tilde{c}_i \cdot e^{gt}, \tilde{k}_i \cdot e^{gt}, \tilde{z}_i \cdot e^{gt})$.

Again, $(\tilde{k}_i, \tilde{z}_i)$ immediately jumps to its steady-state value through an instantaneous Dirac quantum jump in consumption and wealth at date 0. We have:

$$\begin{aligned} V_i(\{\tilde{c}_i, \tilde{k}_i, \tilde{z}_i\}_{t \geq 0}) &= \frac{\delta_i}{\delta_i - g} \cdot \left[\tilde{c}_i + a_i(\tilde{k}_i) - h_i(\tilde{z}_i) + (\delta_i - g) \cdot (k_i^{init} - \tilde{k}_i)\right] \\ = \frac{\delta_i}{\delta_i - g} &\cdot \left[(r-g)\tilde{k}_i + \tilde{z}_i - T(\tilde{z}_i, r\tilde{k}_i) + a_i(\tilde{k}_i) - h_i(\tilde{z}_i)\right] + \delta_i \cdot (k_i^{init} - \tilde{k}_i) \end{aligned}$$

Therefore, with growth, maintaining normalized wealth $\tilde{k}_i$ requires saving $g \cdot \tilde{k}_i$ in perpetuity, hereby lowering consumption by $g \cdot \tilde{k}_i$.

Intuitively, with economic growth, maintaining a given level of normalized wealth requires higher savings and hence reduced consumption. Suppose the economy moves from $g = 0$ to $g > 0$ at time t_0. At time t_0, there is no jump in wealth as normalized wealth is not affected by g. The equation for V_i above shows that wealth lovers (who choose a high $\tilde{k}_i$) gain relatively less than non wealth lovers (who choose for example $\tilde{k}_i = 0$). Economic growth benefits those with no capital more than wealth lovers owning capital.

Let us consider linear taxes on capital for simplicity, with again $\bar{r} = r(1 - \tau_K)$. If $\bar{r} < g$, then wealth lovers would hold more wealth, but have lower consumption than those with less wealth. Conversely, if $\bar{r} > g$, then wealth lovers would hold more wealth and also have higher consumption. In a world in which society disregards wealth per se and cares mostly about consumption (i.e., social welfare weights are based on consumption c only), $\bar{\tau}_K = 1 - g/r$ may be a natural upper bound on the capital tax. This discussion connects with the famous r vs. g discussion at the heart of Piketty (2014).

3.4.2 The aggregate capital stock and an endogenous return to capital

An often discussed policy issue is that, in practice, the return to capital may not be exogenously given by r and may instead endogenously depend on an aggregate production function $F(K, L)$ where $K = \int_i k_i di$ is aggregate capital and $L = \int_i l_i di$ is aggregate labor, with l_i the effective labor supplied by individual i. Earnings are equal to $z_i = w \cdot l_i$ with $w = F_L$ the wage per unit of effective labor. $r = F_K$ is the marginal return to capital.

A direct application of the Diamond and Mirrlees (1971) theory implies that the optimal tax formulas for capital and labor would be unchanged with an aggregate production function. In other words, optimal tax rates depend solely on the supply side elasticities and general equilibrium price effects are irrelevant. The intuition is simple. Consider for instance increasing τ_K. This reduces the capital stock through supply side responses, which in turns increases the pre-tax rate of return r and reduces the pre-tax wage rate w through general equilibrium price effects. This effectively shifts part of the capital tax increase from capital to labor. Relative to the exogenous factor price case, it is as if the government had increased the tax rate on capital by less and increased at the same time the tax rate on labor by some amount. However, this tax incidence shifting can be offset by increasing the capital tax rate by more in the first place and correspondingly reduce the tax rate on labor income. If there are no profits (for example with constant returns to scale in the production function), this tax offset has zero fiscal cost hence the general equilibrium price effects can be freely offset by the tax system and hence ignored in optimal tax analysis.[28]

Thanks to the Diamond-Mirrlees theory, the question of how to tax capital holdings of different individuals can be treated separately from the question about the optimal aggregate capital stock. It has been made repeatedly in the optimal capital tax literature, for example by Atkinson and Sandmo (1980); King (1980) in the OLG life-cycle model of savings, and more recently by Piketty and Saez (2013b) in a model of bequests.

4 Numerical Application to U.S. Taxation

In this section, we give empirical content to the optimal tax rates derived in Section 2. One of the advantages of our method is that the sufficient statistics that appear in the optimal tax formula provide a clear link to the data. We use IRS tax data for 2007 on labor and capital income distributions.[29] We follow the conventions of Piketty and Saez (2003) to define income and percentile groups. The individual unit is the tax unit defined as a single person

[28]With pure profits, this result also carries through if the government can tax pure profits 100%.

[29]We choose 2007 as this is the most recent year of publicly available micro-level US tax data available before the Great Recession. By September 2016, the most recent year available was 2010.

with dependents if any or a married couple with dependents if any. Capital income is defined as all capital income components reported on individual tax returns, and includes dividends, realized capital gains, taxable interest income, estate and trust income, rents and royalties, net profits from businesses (including S-corporations, partnerships, farms, and sole proprietorships). Labor income is defined as market income reported on tax returns minus capital income defined above. It includes wages and salaries, private pension distributions, and other income.[30] We recognize that the tax based income components we use to classify capital and labor incomes do not perfectly correspond to economic capital and labor incomes.[31] Yet, any tax system that taxes capital and labor separately has to use the existing tax based income components. For simplicity, any negative income is set at zero. In aggregate, capital income represents 26% of total income and labor income represents 74% of total income (see Figure 2). As our theory boils down to a static model, it is directly suited for thinking through optimal taxation of annual labor and capital income, as actual income tax systems operate.

4.1 Empirical Facts on Capital and Labor Income

We present here three key facts about the distributions of labor and capital income and highlight their implications for optimal taxation.

i. Capital income is more unequally distributed than labor income.

The distributions of both labor and capital income (and thus of total income) exhibit great inequalities but capital income is much more concentrated than labor income as shown in the Lorenz curves in Figure 1. The top 1% people as ranked by capital income earn 63% of all capital income, while the bottom 80% earn essentially zero capital income.

This fact implies, first, that the Pareto parameter of the capital income distribution will be much smaller than that of the labor income distribution. Second, the top capital tax on top capital earners will be the one that applies to the bulk of capital income and that, hence, the asymptotic capital tax rate we derived in (12) plays an important role.

ii. At the top, total income is mostly capital income.

[30] Our definition of capital income is broad (and correspondingly, our definition of labor income is narrow), as business profits are actually a mix of labor and capital income.

[31] See Piketty et al. (2016) for an attempt to reconstruct the economic capital and labor incomes starting from tax data.

At the top of the income distribution total income comes mostly from capital income. Figure 2 shows capital and labor income as a fraction of total income for the full population (P0-P100) and for several subgroups as ranked by total income. At the top of the income distribution, capital makes up close to 80% of total income.

This fact implies, first, that the Pareto tail parameter of the capital income distribution will be roughly equal to that of the total income distribution. Second, the social welfare weights on top income individuals should be set based on considerations of fairness and deservingness of capital income.

iii. Two-dimensional heterogeneity in both labor and capital income.

There is an important two-dimensional heterogeneity in labor and capital income. Conditional on labor income, capital income continues to exhibit a lot of inequality. Figure 3 plots the Lorenz curves for capital income (the cumulative share of capital income owned by those below each percentile of the capital income distribution), but conditional on being in four groups according to labor income: all individuals, the bottom 50% by labor income, the top 10% by labor income and the top 1% by labor income. Even conditional on labor income, there is still a very large concentration of capital income.

This two-dimensional heterogeneity means that, even if we were to perfectly redistribute labor income, there would still be a lot of capital income inequality and hence substantial total income inequality as well. The differences in capital income conditional on the same labor income could arise from any of the heterogeneities in preferences (discount rate δ_i or value for wealth $a_i(k)$) or bequests described above. They would all violate the assumptions of the Atkinson-Stiglitz theorem and typically justify the need for a capital or wealth tax in addition to an optimal nonlinear labor income tax.

4.2 Optimal Separable Tax Schedules

4.2.1 Methodology

We first start by considering the optimal separable tax schedules for capital and labor income of the form $T_L(z_i)$ and $T_K(rk_i)$, making use our sufficient statistics non-linear formulas derived in Section 2.2.2.

We assume constant elasticities for labor and capital income, denoted by, respectively, e_L

and e_K.[32] We will consider a range of plausible elasticity parameters as there is a paucity of empirical estimates on the capital elasticity.

Starting from the micro-level IRS tax data, we invert individuals' choices of labor and capital income, given the current U.S. tax system to obtain the implicit latent types which are consistent with these observed choices and these constant elasticities. The distribution of types is hence such that, given the constant behavioral elasticities and the actual U.S. tax schedule, the capital and labor income distributions match the empirical ones (Saez (2001) developed this methodology in the case of optimal labor income taxation). We then fit non-parametrically the distribution of latent types. We repeat the same procedure for total income.

At the top, the distributions of labor, capital, and total income exhibit constant hazard rates and approximate a Pareto distribution with tail parameters denoted by, respectively, a_L, a_K, and a_Y. The empirical Pareto parameters are plotted in Figure 4 for labor, capital, and total income. For labor income the Pareto parameter is around $a_L = 1.6$, for capital income it is $a_K = 1.38$, and for total income it is $a_Y = 1.4$ (given that the tail of total income is mostly capital income).

To capture social preferences for redistribution, we assign exogenous weights g_i which decline in observed disposable income at the current tax system such that the weight for individual i in the data is equal to $g_i = 1/((z_i + rk_i)(1 - \tau^{US}) + R^{US})$ where $\tau^{US} = 25\%$ and R^{US} mimic the U.S. average tax rate on total income and demogrant. Such weights decline to zero as income goes to infinity, implying that optimal top rates are given by the asymptotic revenue maximizing tax rates derived earlier.

4.2.2 Results

Panels (a) and (b) in Figure 5 show, respectively, the optimal marginal labor income tax as a function of labor income and the optimal marginal capital income tax as a function of capital income, each for three different values of the elasticity parameters, namely 0.25, 0.5, and 1. We

[32] For labor income, as is well known, this requires a disutility of work of the form $h_i(z) = z_i^0 \cdot (z/z_i^0)^{1+1/e_L}/(1 + 1/e_L)$ where z_i^0 is exogenous potential earnings equal to actual earnings when the marginal labor income tax rate is zero. Similarly, for capital income, this requires a utility of wealth of the form $a_i(k) = \delta_i \cdot k - r \cdot k_i^0 \cdot (k/k_i^0)^{1+1/e_K}/(1 + 1/e_K)$ where k_i^0 is exogenous potential wealth equal to actual steady state wealth when the marginal capital income tax rate is zero. This disutility of wealth function has to depend on the discount rate δ_i and the rate of return r. It is first increasing and then decreasing in wealth k. However, in equilibrium, the individual always chooses k_i in the increasing portion of the $a_i(k)$ function.

use a range of possible elasticities given the uncertainty coming out of the empirical literature (see Saez et al. (2012) for a recent survey).

The optimal labor and capital income tax rates both follow closely the shape of the empirical Pareto parameter from Figure 4. The optimal labor income tax rate hence takes the familiar shape as in Saez (2001) and is lower when the elasticity of labor income to the net of tax rate is higher.

The capital income tax schedule is new. The asymptotic nonlinear tax rate, which approximates the linear top tax rate, as explained in Section 2.2.2 covers the vast majority of the capital income tax base because capital income is so concentrated. Above the top 1%, the optimal marginal tax rate on capital income is essentially constant, so that the nonlinear tax schedule at the top is very well approximated by a linear tax rate.[33] Naturally, the level of that optimal linear top tax rate depends inversely on the elasticity of capital income to the net of tax return.

Because capital income is more concentrated than labor income, the Pareto parameter for capital income is lower than for labor income, leading to a higher top tax rate for capital income than for labor income when the elasticities e_L and e_K are the same. Therefore e_K would need to be significantly higher than e_L to justify imposing the same top tax rate on capital and labor incomes.

4.3 Optimal Comprehensive Tax Schedule

We then turn to exploiting the optimal tax on comprehensive income, $T_Y(y)$, with $y = z + rk$, making use of the nonlinear formulas derived in Section 3.2.1 in terms of sufficient statistics. We repeat the same procedure outlined above for labor and capital income, assuming that the elasticity of total income e_Y is constant. We again consider three possible values. Panel (c) in Figure 5 plots the optimal marginal tax rate $T_Y'(y)$ as a function of total income y.

The optimal marginal tax rate on total income has a shape similar to that on labor income. Often in numerical applications of the Mirrlees (1971) labor income tax model, total income is used for the calculations. We can here rigorously compare the resulting two schedules. The asymptotic top tax rate on total income is closest to the asymptotic top tax rate on capital

[33] The constant tax rate is driven by the constant elasticity and the social welfare weights which become small and do not vary much above the top 1%.

income from panel (b) as capital income dominates labor income among top incomes.

5 General Model with Concave Utility of Consumption

In this section, we generalize the results from the previous simple model to the case with an arbitrary concave utility for consumption. After introducing the generalized model, we discuss three possible ways to do optimal tax analysis: (1) considering unanticipated tax reforms, (2) considering anticipated tax reforms, (3) focusing on the long-run steady-state. We will show that in all three cases, the formulas from Sections 2 and 3 still apply in this generalized model, as long as the elasticity of the tax base is appropriately defined.[34] The type of reform considered (anticipated, unanticipated, or steady state) is itself a policy choice and embodies an implicit normative judgement. In the simple model of Section 2, all three approaches are equivalent so that determining the correct normative framework is straightforward. This is a key strength of the simple framework, in addition to its tractability. With concave utilities the three approaches deliver different elasticities and hence quantitatively different optimal tax rates. We point out the advantages and drawbacks of each of the three approaches, and argue that none is perfect.

Individual maximization. In the generalized model with concave utility for consumption and wealth in the utility, the instantaneous utility of individual i is $u_i(c_i(t), k_i(t), z_i(t))$, which is increasing and concave in consumption c and wealth k, and decreasing and convex in labor earnings z with arbitrary heterogeneity across individuals. We assume a uniform discount factor δ across individuals to simplify the presentation. In Appendix A.2, we provide and discuss all formulas with heterogeneous discount rates δ_i. With a general tax system $T_t(z, rk)$, individual i chooses $(c_i(t), k_i(t), z_i(t))_{t\geq 0}$ to maximize

$$V_i = \delta \int_{t=0}^{\infty} u_i(c_i(t), k_i(t), z_i(t))e^{-\delta t}dt \quad \text{s.t.} \quad \frac{dk_i(t)}{dt} = rk_i(t) + z_i(t) - T_t(z_i(t), rk_i(t)) - c_i(t).$$

Routine calculations show that, with basic regularity assumptions and assuming that the tax system converges to $T(z, rk)$, this individual maximization converges to a steady state (c_i, k_i, z_i)

[34] We will also compare our results to those of earlier models, showing that the unifying tax formulas we obtain are widely applicable.

characterized by:

$$u_{ik}/u_{ic} = \delta - r(1 - T_K'), \qquad u_{ic} \cdot (1 - T_L') = -u_{iz}, \quad \text{and} \quad c_i = rk_i + z_i - T(z_i, rk_i), \tag{22}$$

where u_{ic}, u_{ik}, u_{iz} denote the partial derivatives of $u_i(c, k, z)$ and T_K', T_L' the marginal tax rates on capital income and labor income all evaluated at the steady state (c_i, k_i, z_i).

Conditional on labor income, wealth is heterogenous across individuals due to differences in the taste for wealth (embodied in the utility function u_i). The steady state wealth k_i and earnings z_i for individual i naturally depend upon the tax system with standard income and substitution effects. Relative to the simpler model in Section 2, individuals care about consumption smoothing. Hence the convergence to the steady state is no longer instantaneous and there are transitional dynamics. We denote by k^m and z^m average (and also aggregate as population is normalized to one) wealth and labor earnings in steady state.

Simplifying assumptions. We make three simplifying assumptions for pedagogical purposes.

First, we assume that the government always chooses a period by period neutral budget constraint so that for all t:

$$\int_i T_t(z_i(t), rk_i(t))di = 0.$$

Second, we consider the case of linear taxation with time invariant tax rates on capital and labor income τ_K, τ_L with a budget balancing lump-sum rebate $G(t)$. Hence, $T_t(z, rk) = \tau_L z + \tau_K rk - G(t)$. The first assumption on per period budget balancing implies that $G(t) = \tau_L z^m(t) + \tau_K rk^m(t)$ where $k^m(t)$ and $z^m(t)$ are average wealth and earnings at time t.

Third, we assume that at time 0, the economy is already in steady state with its initial tax system $T(rk, z) = \tau_L z + \tau_K rk - G$. As a result, G is also time invariant. This also implies that $c_i(t), z_i(t), k_i(t), k^m(t), z^m(t)$ are all time invariant and equal to their steady state values c_i, z_i, k_i, k^m, z^m. Starting from this steady state, we will consider small budget neutral tax reforms. At the optimum, such tax reforms should have zero first order effect on welfare. These first order conditions give us optimal tax formulas for τ_L and τ_K.

The results largely carry over without making these three assumptions but notation becomes more cumbersome. We discuss some of these extensions, in particular nonlinear taxation (in Appendix A.4), starting away from the steady state (in Appendix A.2.2), and time varying tax rates (in Section 5.4).

Steady-state elasticities. In the steady state defined above, aggregating across individuals, average capital $k^m = \int_i k_i$ and average earnings $z^m = \int_i z_i$ will be functions of $1 - \tau_K$ and $1 - \tau_L$ (as the lump-sum rebate G is also a function of τ_K, τ_L through budget balance). In particular, k^m has a finite elasticity with respect to the net-of-tax return $\bar{r} = r(1 - \tau_K)$ which we denote again by e_K as in Section 2.

$$\textbf{Steady state elasticity:} \quad e_K = \frac{\bar{r}}{k^m}\frac{dk^m}{d\bar{r}} \tag{23}$$

Note that this steady state elasticity takes into account the fact that a change in $\bar{r}$ through a change in τ_K also affects the steady state lump-sum grant G, creating further effects on wealth accumulation (and labor supply) decisions. Hence, e_K is a *policy elasticity* (Hendren, 2016) that mixes substitution and income effects. Importantly, utility for wealth remains the key ingredient making the elasticity e_K finite. With no utility for wealth, the steady state is such that $\delta = r(1 - \tau_K)$ so that if $\delta < r(1 - \tau_K)$, individuals are patient and accumulate an ever growing capital stock (and, conversely, accumulate growing debt if $\delta > r(1 - \tau_K)$), and capital supply is infinitely elastic in steady state.[35] Transitional dynamics are irrelevant for the steady-state elasticity. There is also a finite cross-elasticity of k^m with respect to $1 - \tau_L$ denoted by $e_{K,1-\tau_L} = ((1 - \tau_L)/k^m) \cdot dk^m/d(1 - \tau_L)$.

Aggregate labor earnings z^m also has finite elasticities with respect to both $1 - \tau_L$ and $\bar{r}$. In particular, the cross-elasticity $e_{L,1-\tau_K} = (z_m/\bar{r})(dz^m/d\bar{r})$ captures how labor earnings respond to changes in $\bar{r}$. In the generalized model, cross-elasticities arise not only because of potential jointness of (k, z) in utility as in Section 3.3.1 but also through income effects since the marginal utility of consumption u_{ic} affects labor supply decisions as seen in (22).

5.1 Optimal Taxes with Unanticipated Tax Reforms

Optimal tax framework. We start with a pre-existing time invariant linear tax system and an economy that has converged to steady state as of time 0. The government can change tax

[35] Besides wealth in the utility, other ways to make the steady-state elasticity finite is to introduce stochastic earnings as in Aiyagari (1995) or endogenous discount factors that depend on the consumption levels as in Judd (1985). We relate our results to these models below and in Appendix A.3.

policy at time 0 once and for all without individuals anticipating the tax policy change. The pre-existing tax system is optimal if a small budget neutral reform of the initial tax system has zero first order effect on welfare. Formally, the government chooses the new time invariant and period-by-period budget balanced tax system (τ_K, τ_L) to maximize

$$SWF = \int_i \omega_i V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}) di, \tag{24}$$

$$\text{where} \quad V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}) = \delta \cdot \int_{t=0}^{\infty} u_i(c_i(t), k_i(t), z_i(t)) e^{-\delta t} dt,$$

and $\omega_i \geq 0$ are exogenous Pareto weights. We define the social marginal welfare weight on person i as $g_i = \omega_i u_{ic}(c_i(0), k_i(0), z_i(0))$. We assume without loss of generality (by normalizing the Pareto weights ω_i) that they sum to one: $\int_i g_i di = 1$.

Unanticipated elasticities. Starting from a steady state (τ_K, τ_L), we consider a small reform $d\tau_K$ that takes place at time 0 and is unanticipated. To meet period-by-period budget balance, the lump-sum grant $G(t) = \tau_K r k^m(t) + \tau_L z^m(t)$ adjusts automatically by $dG(t) = rk^m(t)d\tau_K + \tau_K r dk^m(t) + \tau_L dz^m(t)$. We will derive conditions such that the small reform has zero first order effect on welfare, which effectively implies that the initial tax rate τ_K is optimal.[36]

Let $e_K^u(t)$ be the elasticity of aggregate capital in period t, $k^m(t)$, with respect to the net of tax rate $\bar{r}$, i.e.: $e_K^u(t) = (\bar{r}/k^m(t)) \cdot (dk^m(t)/d\bar{r})$. The superscript u signifies that the reform was unanticipated (as of time 0). Again, $e_K^u(t)$ is a policy elasticity that incorporates the income effects of $dG(t)$ on $k^m(t)$. $e_K^u(t)$ varies with time because of transitional dynamics due to consumption smoothing concerns. Over time, it converges to the steady state elasticity e_K defined in (23). In contrast to Section 2, the convergence is not immediate because individuals smooth consumption and, hence, adjust their wealth slowly. Therefore, under regularity assumptions, $e_K^u(t)$ starts at zero at $t = 0$, builds up with time t, and converges to $e_K > 0$ for t large. We define the unanticipated elasticity of capital with respect to $\bar{r}$ as the time discounted average of

[36] It is also possible to start from an arbitrary tax system (τ_{K0}, τ_{L0}) and away from the steady state, and then derive the optimal unanticipated new tax system (τ_K, τ_L) implemented at time 0 that maximizes social welfare. We do this in Appendix A.2.2. The formulas are similar, but require keeping track across time of all variables that converge slowly to the new steady-state, demanding more cumbersome notations to define the time discounted average of all variables.

the elasticities $e_K^u(t)$ in each period as follows:

$$\textbf{Unanticipated elasticity for reform at time } 0\textbf{:} \quad e_K^u = \delta \int_{t=0}^{\infty} e_K^u(t) e^{-\delta t} dt. \tag{25}$$

Typically the unanticipated elasticity e_K^u is smaller than the steady-state elasticity e_K as the unanticipated response $e_K^u(t)$ grows overtime toward the steady-state elasticity e_K. Economically, adjusting one's wealth requires changing consumption levels temporarily. As individuals value consumption smoothing with concave utility, they do not want abrupt changes in consumption and hence they adjust their wealth slowly. Unanticipated labor earnings elasticities and cross elasticities can be defined in the same way. We denote by $e_{L,1-\tau_K}^u(t)$ the time t elasticity of earnings $z^m(t)$ with respect to $\bar{r}$ and by $e_{L,1-\tau_K}^u$ the time discounted average of $e_{L,1-\tau_K}^u(t)$.

Optimal tax solution. Using the envelope theorem (i.e., that behavioral responses $dk_i(t)$ can be ignored when computing the change in individual welfare dV_i) and the fact that $u_{ic}(t)$ and $k_i(t)$ are time invariant when starting from a steady state, we can compute the welfare impact of the small tax change on individual i utility as $dV_i = \delta u_{ic} \int_{t=0}^{\infty} [dG(t) - rk_i d\tau_K] e^{-\delta t} dt$. Therefore, using the expression for $dG(t)$ from above, we have:

$$dV_i = u_{ic} \cdot \delta \int_{t=0}^{\infty} [rk^m d\tau_K + \tau_K r dk^m(t) + \tau_L dz^m(t) - rk_i d\tau_K] e^{-\delta t} dt.$$

From this expression, we can introduce the elasticities $e_K^u(t)$ and $e_{L,1-\tau_K}^u(t)$, aggregate across individuals using Pareto weights ω_i, and obtain the following expression for the change in total social welfare (24):

$$dSWF = rd\tau_K \int_i g_i \cdot \delta \int_{t=0}^{\infty} \left[k^m - k_i - \frac{\tau_K k^m}{1-\tau_K} e_K^u(t) - \frac{\tau_L z^m / r}{1-\tau_K} e_{L,1-\tau_K}^u(t) \right] e^{-\delta t} dt$$

$$= rk^m d\tau_K \left[1 - \int_i g_i \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} e_K^u - \frac{\tau_L}{1-\tau_K} \frac{z^m}{rk^m} e_{L,1-\tau_K}^u \right].$$

The optimal tax rate is such that $dSWF = 0$ which yields the following optimal linear capital tax rate formula.

Proposition 8. ***Optimal linear taxes with unanticipated responses.***

$$\tau_K = \frac{1 - \bar{g}_K - \tau_L \frac{z^m}{rk^m} e^u_{L,1-\tau_K}}{1 - \bar{g}_K + e^u_K} \quad \textit{with} \quad e^u_K = \delta \int_{t=0}^{\infty} e^u_K(t) \cdot e^{-\delta t} dt, \tag{26}$$
$$e^u_{L,1-\tau_K} = \delta \int_{t=0}^{\infty} e^u_{L,1-\tau_K}(t) \cdot e^{-\delta t} dt, \quad \textit{and} \quad \bar{g}_K = \int_i g_i \cdot k_i / k^m.$$

A symmetric formula holds for the optimal labor income tax rate τ_L as in Proposition 6 (derivation omitted). The formula is qualitatively exactly the same as in Section 2 by simply replacing the steady state elasticities by the unanticipated elasticities. As typically $e^u_K < e_K$ due to slowly building responses, we expect the optimal tax rate to be higher than in the steady state approach. Intuitively and very simply, the government can tax capital more because it can tax existing capital that cannot respond.[37]

It is also possible to consider the optimal nonlinear tax system. A particularly simple and relevant case is to consider a linear tax rate on labor income and a nonlinear capital income tax with a constant tax rate τ_K for capital income above rk^{top}. We consider this case in online Appendix A.4.[38]

Discussion. If the responses of capital to tax changes are very fast, then e^u_K is very close to the steady state elasticity e_K. In this case, even the quantitative implications of this approach will be similar to those of the simpler model of Section 2. Empirically, policy makers in general worry about capital adjustments happening very quickly following tax changes by, for instance, capital flights abroad (Johannesen, 2014).[39] Saez et al. (2012), surveying the literature on taxable income elasticities, argue that the long-term responses, although particularly important in the case of a dynamic decision such as capital are understudied. Saez (2017) shows strong short term retiming response of capital income (particularly of realized capital gains and dividends)

[37] With heterogeneous discount rates δ_i, the same optimal tax formulas apply but the discounted elasticities need to be redefined such that $e^u_K = \int_i \delta_i \cdot g_i \int_{t=0}^{\infty} e^u_K(t) \cdot e^{-\delta_i t} dt$. Full derivations are in Appendix A.2.1.

[38] Note that we can also generalize the other results from Section 3. The optimal tax on total income $y_i = rk_i + z_i$ takes the same form as in Proposition 3 with the long-run elasticity e_Y replaced by the total elasticity of the income tax base, taking into account the transitional adjustments, $\bar{e}_Y = \delta \int_0^{\infty} e_Y(t) \cdot e^{-\delta t} dt$. Similarly it is straightforward to generalize the results in subsections 3.3.2 and 3.3.3. Regarding the latter, with transitional dynamics, the government will be more tempted to tax more heavily assets which are slower to adjust (holding fixed the long-run elasticity e^j_K and the distributional factor $\bar{g}^j_K$).

[39] Johannesen (2014) shows that the introduction of a withholding tax for EU individuals with Swiss bank accounts led immediately, within two quarters of the reform, to a drop of 30-40% in deposits. Empirical evidence on the short-run versus long-run responses of capital to taxes is very difficult to come by.

following the 2013 US federal income top tax rate increase but much smaller medium term responses. Companies can modify their dividend payouts quickly to changes in dividend taxation for the sake of their shareholders (Chetty and Saez, 2005; Alstadsaeter and Fjaerli, 2009).[40]

If responses are slow on the other hand, then $e^u_K < e_K$. In the short-run, the equity-efficiency trade-off for capital taxation looks more favorable if individuals are not able to adjust their capital quickly. As a result, and considering formula (26), the government can tax more by taking advantage of these slow responses in the short-run.

However, exploiting such slow responses does not seem very appealing from a normative perspective. A well-designed policy cannot or should not endlessly exploit short-run adjustment costs. If nothing else, this will create a commitment problem for the government as it will always look appealing to unexpectedly increase taxes on existing capital temporarily.[41]

5.2 Optimal Taxes with Anticipated Tax Reforms

Traditionally, the literature has tried to resolve the problem of taxing existing capital by considering anticipated tax reforms. Such reforms are supposed to give existing capital the chance to adjust before the new tax policy takes effect. We consider again an economy with time invariant (τ_L, τ_K) in steady state as of time 0. The government maximizes again discounted utilities exactly as in (24). We now assume that the government pre-announces a small budget balanced tax change that takes place $T > 0$ years from now. As the optimum, such a reform should have no first order effect on welfare.[42]

Anticipated elasticities. Consider a small reform $d\tau_K$ that takes place at time $T > 0$ and is pre-announced at time zero. All individuals anticipate that the tax rate will go up starting at time T. Intuitively, individuals want a lower wealth stock after time T but do not like abrupt consumption changes due to consumption smoothing considerations. Hence they start responding immediately as of time 0 by reducing their wealth stock slowly. We denote by $e^a_K(t) = (\bar{r}/k^m(t)) \cdot (dk^m(t)/d\bar{r})$ the elasticity of aggregate capital in period t. The superscript

[40] These authors find that the introduction of the Norwegian shareholder income tax led to immediate effects on payouts, emphasizing that capital income can react very quickly and flexibly to tax changes.

[41] A large literature has analyzed the problem of commitment for capital taxation (see the early work by Kydland and Prescott (1977), Fischer (1980), and, more recently, Farhi et al. (2012)).

[42] When T becomes large, this approach is equivalent to finding the optimal long-run tax rates as in Chamley (1986) and Judd (1985) (see our discussion below).

a denotes that the tax change is anticipated to take place at T ($e^a_K(t)$ formally depends on T as well). $e^a_K(t)$ converges again to the steady state elasticity e_K as $t \to \infty$. We define the anticipated elasticity of capital with respect to $\bar{r}$ as the time discounted average of the $e^a_K(t)$.

$$\textbf{Anticipated elasticity for reform at time } T\textbf{:} \quad e^a_K = \delta \int_{t=0}^{\infty} e^a_K(t) e^{-\delta(t-T)} dt. \qquad (27)$$

We discount elasticities from time T as the reform starts in time T, but responses start at time $t = 0$ because the reform is anticipated. Anticipated labor earnings elasticities and cross elasticities can be defined in the same way.

The anticipated elasticity e^a_K is typically larger than the unanticipated e^u_K because individuals start responding in advance of the reform. The two elasticities coincide when $T = 0$ and it is expected that e^a_K grows with T as individuals have more and more time to respond in advance of the reform.

In a wide range of models, e^a_K grows to infinity as T grows to infinity. Piketty and Saez (2013b) showed that this is the case in the standard model with no utility of wealth as in Chamley-Judd (as well as in the endogenous discount rate model of Judd (1985)). This is also the case in our model with utility of wealth as we shall see (and as shown formally in Appendix A.2.2). The intuition for this result is simple. If T is large, then individuals have an opportunity to respond a very long time in advance. As time is discounted at rate $\delta > 0$, early responses loom large relative to the discounted capital tax bases for $t \geq T$, which are directly affected by the reform. Increasing the tax rate in the distant future makes the government incur a lot of revenue losses before the tax change happens in exchange for mechanical gains from the tax increases that are far into the future and hence heavily discounted.

Optimal tax solution. The derivation of the optimal tax formula closely parallels the derivation in the case of unanticipated tax changes and hence delivers the exact same tax formulas by simply replacing the unanticipated elasticities with the anticipated elasticities.

Proposition 9. ***Optimal linear taxes with anticipated responses.***

$$\tau_K = \frac{1 - \bar{g}_K - \tau_L \frac{z^m}{rk^m} e^a_{L,1-\tau_K}}{1 - \bar{g}_K + e^a_K} \quad with \quad e^a_K = \delta \int_{t=0}^{\infty} e^a_K(t) \cdot e^{-\delta(t-T)} dt, \tag{28}$$

$$e^a_{L,1-\tau_K} = \delta \int_{t=0}^{\infty} e^a_{L,1-\tau_K}(t) \cdot e^{-\delta(t-T)} dt, \quad and \quad \bar{g}_K = \int_i g_i \cdot k_i / k^m.$$

If $T \to \infty$, then $e^a_K \to \infty$ so that $\tau_K \to 0$.

Note that the value of the optimal tax rate will be different for each possible reform date T. With heterogeneous discount rates, the same formula applies, replacing $\bar{g}_K = \int_i g_i e^{-\delta_i T} \cdot k_i / (k^m \int_i e^{-\delta_i T})$, $e^a_K = \int_i \delta_i g_i e^{-\delta_i T} \int_0^\infty e^a_K(t) \cdot e^{-\delta_i(t-T)} dt$, and $e^a_{L,1-\tau_K} = \int_i \delta_i g_i e^{-\delta_i T} \int_0^\infty e^a_{L,1-\tau_K}(t) \cdot e^{-\delta_i(t-T)} dt$, which all depend on the reform time T. Appendix A.2.2 provides all formal derivations.

Discussion. Hence, in this case, the optimal formula takes again the same shape, with the extra implication that the relevant elasticity becomes infinite when the tax reform is anticipated infinitely in the advance. While this leads to a zero optimal capital tax rate, it occurs in a particularly unrealistic policy setting, namely if the reform is announced infinitely in advance with perfect certainty. It is also a particularly fragile result since adding uncertainty about future earnings makes the elasticity finite as shown in Piketty and Saez (2013b) and discussed below. To our knowledge, there is also no empirical evidence that pre-announcing changes in capital income tax rates leads to changes in consumption immediately because of consumption smoothing responses. There is evidence of avoidance responses in anticipation of tax changes (see e.g., Saez et al. (2012); Saez (2017)) but they clearly take place through the tax avoidance mechanisms of our simple model from Section 2 rather than consumption smoothing. Furthermore, with heterogeneous discount rates, as shown above, $\bar{g}_K = \int_i g_i e^{-\delta_i T} k_i / (k^m \int_i e^{-\delta_i T})$, so that the anticipated approach would effectively load all welfare weights on the most patient individuals and put zero welfare weight on all others (Piketty and Saez (2013b) discuss this issue in detail). This is obviously not an attractive normative property.[43] Therefore, our view is that the anticipated reform approach is not an appealing way to resolve the problem of taxing

[43] An anticipated reform with a finite $T < \infty$ either does not resolve the problem of taxing existing capital (i.e., it will still exploit sluggish responses) or runs into the same unrealistic predictions of very large elasticities, as well as normative difficulties on how to weight agents with heterogeneous discount rates.

existing capital.

5.3 Optimal Taxes in the Utility Based Steady State Approach

The unanticipated and anticipated approaches both have problems. The unanticipated approach makes optimal tax rates too high because it exploits sluggish responses and hence effectively taxes existing capital. The anticipated approach solution lacks realism and generates anticipated responses that do not seem relevant empirically yet get infinitely large in the model.[44] Therefore, in this section, we propose an alternative, new, and non-standard solution to the optimal capital tax problem. Our goal is to neutralize the ability of the government to exploit sluggish responses. However, instead of using anticipation (which leads to the aforementioned undesirable features), we want the government to explicitly recognize the long-run steady state behavioral responses as the normatively relevant ones. We also want the government to respect individual savings choices. This is why we call our approach "utility based steady state approach." We are going to see that this goal leads to an obvious solution, namely to use *the standard optimal tax formulas with the steady-state elasticities*. The challenge is that the goal cannot be formalized in a standard way as a fully consistent maximization problem, which is why this solution is non-standard and new.

Optimal tax problem. Let us assume again that we start from a steady state with tax system (τ_L, τ_K) and we consider a small reform $d\tau_K$ at time 0. As we have seen in section 5.1, the actual response to this tax change is sluggish. The simplest way to formalize the idea that the government is not allowed to exploit such sluggish responses is to assume that the government considers that steady state responses start immediately in terms of revenue implications. To be specific, the real change in taxes collected at time t is $dG(t) = rk^m d\tau_K + \tau_K r dk^m(t) + \tau_L dz^m(t)$ yet the government considers that the budgetary effect at time t is $dG = rk^m d\tau_K + \tau_K r dk^m + \tau_L dz^m$. In other words, the government absorbs the difference between $dG(t)$ and dG. For example, in the case of a tax increase $d\tau_K > 0$, because of transitional dynamics, responses are smaller at first so that the real $dG(t)$ is actually bigger than the steady-state dG but we just assume that

[44] It is also normatively unappealing as it puts very low weight on impatient agents.

the government "burns" the surplus.[45] Hence, normatively, the government ignores the gains it can make by exploiting slow responses. Formally the government problem can be written as finding the tax system (τ_K, τ_L) that maximize SWF as in (24) but assuming that the lump-sum grant $G(t)$ is equal to the steady-state lump-sum grant: $G = r\tau_K k^m + \tau_L z^m$ instead of the actual lump-sum grant $r\tau_K k^m(t) + \tau_L z^m(t)$. All derivations are in Appendix A.2.3.

Optimal tax solution. This problem can be solved exactly paralleling the derivations from sections 5.1 and 5.2 by simply replacing the unanticipated or anticipated elasticities with the steady state elasticities defined above.

Proposition 10. ***Optimal linear capital tax in the utility based steady state approach***

$$\tau_K = \frac{1 - \bar{g}_K - \tau_L \frac{z^m}{rk^m} e_{L,1-\tau_K}}{1 - \bar{g}_K + e_K} \quad \textit{with} \quad \bar{g}_K = \int_i g_i \cdot k_i / k^m. \tag{29}$$

A symmetric equation holds for the optimal labor income tax rate τ_L as in Proposition 6. Hence, the same tax formulas hold by simply using the steady state elasticities. One advantage of this approach is that it is fully robust to introducing heterogeneity in discount rates across individuals as heterogeneity in discount rates is normatively irrelevant in the steady state.

Importantly, all of the applications from the simple model in Sections 2 and 3 carry over with the steady-state approach with relatively slight modifications. These modifications relate to the concave utility function, which introduces income effects and cross-elasticities (but in either case, there are no transitional dynamics). In fact, with exogenous labor income, all capital tax formulas from Sections 2 and 3 carry over one for one. With endogenous labor income, we would simply have to augment the capital tax formulas with the cross-elasticity of labor income with respect to the net of tax rate $\bar{r}$.[46]

Link with steady state welfare maximization. The optimal tax from Proposition 10 is related but not identical to choosing the budget balanced tax system (τ_L, τ_K) that maximizes

[45] Conversely, for a $d\tau_K < 0$, we assume that the government can "make up" the deficit due to the fact that revenue increasing behavioral responses build up only slowly toward the full steady state response.

[46] The optimal labor tax formulas in turn have to be augmented with the cross elasticity of capital income with respect to the labor tax. The policy elasticities e_K, e_L, $e_{K,1-\tau_L}$ and $e_{L,1-\tau_K}$ captured only substitution effects in Sections 2 and 3, while they mix substitution and income effects here. This does not matter for our sufficient statistics formulas expressed in terms of policy elasticities.

steady state welfare $SWF = \int_i \omega_i \cdot u_i(c_i, k_i, z_i)di$. In the steady-state maximization case, the lump-sum grant is always given by $G = r\tau_K k^m + \tau_L z^m$. Solving this problem is straightforward but does not generate the same formula as in Proposition 10 because the steady-state maximization objective is paternalistic.

To see this, suppose we consider a small tax reform $d\tau_K$ that triggers a wealth response dk_i for individual i in steady state. Using the first order conditions (22), the infinitesimal change in wealth dk_i has an effect on individual i steady state instantaneous utility $u_i = u_i(r(1-\tau_K)k_i + (1-\tau_L)z_i + G, k_i, z_i)$ equal to $du_i = (u_{ic}r(1-\tau_K) + u_{ik})dk_i = u_{ic}\delta dk_i$ where the last equality comes from the steady state condition $u_{ik}/u_{ic} = \delta - r(1-\tau_K)$ from (22). Hence, the behavioral change in dk_i does have a first order impact on the individual utility. Intuitively, increasing wealth looks good because the steady state "forgets" that accumulating wealth required to sacrifice consumption in the past. This artificially creates a positive welfare effect of wealth accumulation that will tend to lower the optimal capital income tax. This "internality," i.e., the omission of the cost to the individual of past sacrificed consumption, is not normatively appealing in this case where our principle is that the government should respect individuals' savings decisions.[47]

The simplest way to fix this issue of paternalism is to deliberately ignore the effect of dk_i on individual welfare by stating that any behavioral response triggered by a tax reform should have zero first order effect on individual welfare through the envelope theorem. Intuitively, this means that the government respects individual savings' decisions. With this "forced" envelope theorem assumption, the optimal tax derivation can be done entirely in the steady-state without worrying about dynamic considerations and the optimal steady-state tax formula is given exactly as in Proposition 10.

To see this, consider tax change $d\tau_K$ (say starting at time 0). The economy converges to a new steady state. Using the forced envelope theorem assumption, the direct welfare effect of $d\tau_K$ on individual i utility (keeping the lump-sum rebate G constant) is $-u_{ic}rk_i d\tau_K$. The higher capital tax hurts each person to the extent of her capital income, independently of how strongly she responds to the tax change. In the new steady state, the change in the aggregate

[47] Piketty and Saez (2013b) use such a steady-state objective in their benchmark analysis of inheritance taxation. The steady-state objective is easier to justify in the inheritance tax context as each time period represents a new generation of distinct individuals. As a result, leaving a larger bequest does generate a positive "externality" on the next generation of individuals.

k^m and z^m are governed by the steady-state elasticities, and they in turn affect the lump-sum rebate G by dG further affecting individual i utility by $u_{ic}dG$. The total effect on individual i utility is therefore $du_i = u_{ic}[-rk_i d\tau_K + dG]$. Aggregating the money metric effects $-rk_i d\tau_K + dG$ across individuals using marginal social welfare weights g_i immediately leads to the conventional optimal tax formula as in Proposition 10.

There are thus two ways to obtain the optimal tax formula from Proposition 10. First, maximize dynamic social welfare SWF as in (24) but assuming that the lump-sum grant $G(t)$ is equal to the steady-state lump-sum grant, or, second, maximize steady state welfare $\int_i \omega_i \cdot u_i(c_i, k_i, z_i)di$ using the forced envelope theorem.

Discussion. The solution proposed in this section is normatively appealing but is unfortunately not conceptually consistent with a fully spelled out dynamic model. The scenario proposed above of assuming that the lump-sum grant responds immediately as in steady state is a well closed problem. The direct steady-state derivation with the "forced envelope" theorem is simple and intuitive but cannot–to the best of our attempts–be presented as a well defined formal maximization problem.

Another way to defend this non-standard steady-state solution is consider the analogy with labor income taxation. The Mirrlees (1971) model of optimal labor income taxation can be narrowly interpreted as a labor supply model with the elasticity of hours of work to taxes. It can also be interpreted more broadly as a model of earnings supply incorporating long-run responses of human capital accumulation or occupational choice. For labor too there is a short-run elasticity in which hours are adjusted, and a long-run, potentially larger, elasticity based on skill choice or occupational choice. The same formulas – which we routinely use– carry over simply substituting the short-run labor supply elasticity by the long-run elasticity of earnings e_L in the standard optimal (linear) tax rate formula $\tau_L = (1 - \bar{g}_L)/(1 - \bar{g}_L + e_L)$.

If one wanted to rigorously micro-found the long-run dynamic of labor supply responses, one would face exactly the same difficulties we are discussing here. Indeed, following Jones et al. (1993), there is a literature on optimal dynamic labor taxation that parallels the optimal dynamic capital taxation of Chamley and Judd. It also finds that long-run optimal tax rates on labor incomes should be zero because anticipated elasticities for human capital accumulation are also infinite in the limit (see our discussion on anticipation elasticities). Yet, this literature has not had much traction in altering views about long-run optimal labor income taxation nor in actual policy debates. Hence, to the extent the reader believes the conventional static Mirrlees (1971) model is helpful to understand optimal long-run labor taxation, the reader also has to keep an open mind to the approach proposed in this section for optimal long-run capital taxation.

5.4 Connection with Earlier Models

We now discuss the connection with earlier models. Our formulas can be applied using each model's specific elasticity values. The latter are determined both by the model's primitives (e.g., wealth-in-the-utility or standard concave utility, uncertainty in labor earnings, etc.) and by the type of reform considered (anticipated, with different anticipation horizons T, unanticipated, etc.). Conditional on the elasticities, the primitives of the model are largely irrelevant. Table 1 provides a systematic comparison and summary of all the cases.

The Chamley-Judd model. In the Chamley-Judd model (Chamley, 1986; Judd, 1985), individuals have a standard utility $u(c_{it}, z_{it})$ and there is no uncertainty. Chamley and Judd assume that the government chooses and announces the optimal time variant tax system (τ_{Lt}, τ_{Kt}) at time zero. Hence, effectively, individuals anticipate the tax rate at time t exactly t years in advance. Initially, the taxes are unanticipated and hence tax rates on capital are large. However, as t grows large, the tax rates are anticipated long in the distance, and the tax rate τ_{Kt} converges to zero. In our formula (28), in the long distance future, T goes to infinity so that the anticipated elasticity becomes infinite and the optimal long run tax rate is correspondingly zero, which is another form of restating the famous zero tax result of Chamley and Judd.

Without utility of wealth, as we have noted, the steady state is degenerate unless $\delta = \bar{r}$, which means that in the steady state, any change in the capital income tax rate leads to an infinite response. Hence, $e_K = \infty$ and the optimal capital tax in the steady state approach of Section 5.3 is also zero.[48]

[48]This optimal zero steady state result in the Chamley-Judd model was pointed out by Piketty (2000), p. 444.

The Judd endogenous discount factor model. In Judd (1985), the discount rate $\delta_i = \delta_i(c_i)$ depends smoothly and negatively on consumption. Utility is:

$$V_i(\{c_i(t), z_i(t)\}_{t\geq 0}) = \int_0^\infty u_i(c_i(t), z_i(t)) e^{-\int_0^t \delta_i(c_i(s))ds} dt.$$

In this model, the steady state elasticity is finite. Indeed, the key point Judd (1985) wanted to make with this model is to show that the zero capital tax rate was not the consequence of an infinite steady state elasticity of capital supply. However, the anticipated elasticity also grows to infinity as $T \to \infty$ (as shown in Piketty and Saez (2013b)) and hence the zero capital tax result with anticipated tax reforms is ultimately again the consequence of an infinite elasticity.

In Appendix A.3.1, we derive the optimal linear tax formula starting from a steady state and considering an unanticipated reform, which is the same as in (28), except that $\bar{g}_K$ is redefined to take into account that the welfare impact of taxes now also goes through the discount factor $\delta_i(c_i)$ which depends negatively on consumption:

$$g_i = \frac{\omega_i \frac{1}{\delta_i(c_i)} \left(u_{ic} + \frac{\delta_i'(c_i)}{\delta_i(c_i)} u_i \right)}{\int_i \omega_i \frac{1}{\delta_i(c_i)} \left(u_{ic} + \frac{\delta_i'(c_i)}{\delta_i(c_i)} u_i \right)} \quad \text{and} \quad e_K^u = \int_i g_i \delta_i(c_i) \int_{t=0}^\infty e^{-\delta_i(c_i)t} e_K(t) dt.$$

Again, the faster capital adjusts, the closer e_K^u is to the long-run elasticity e_K. As with wealth-in-the-utility, the steady state of this model is non-degenerate, with $\delta_i(c_i(t)) = \bar{r}$ and generates a finite long-term elasticity e_K and hence a positive optimal tax rate using the steady state approach with the same formula as in equation (29) of Proposition 10.

The Aiyagari model with uncertainty. In the Aiyagari (1995) model, time is discrete and earnings are stochastic. The economy converges to an ergodic steady state that depends on the long-run tax system but is independent of the distribution of initial wealth. In this model, all elasticities including the steady state elasticity and the anticipated elasticity with $T \to \infty$ are finite as shown in Piketty and Saez (2013b). As a result, optimal capital tax rates are positive with all three approaches, although the quantitative elasticities and optimal tax rates will be different. Importantly, adding uncertainty in earnings does not affect the formulas expressed in terms of elasticities. Adding utility for wealth in the Aiyagari model would not change the qualitative elasticities either.

Therefore, adding uncertainty can be seen as a way to obtain finite elasticities and derive formulas expressed in sufficient statistics. Our approach in this paper has been instead to use wealth in the utility which also generates finite steady state elasticities and a simpler way to derive optimal tax rates.[49]

In our view, the fact that the formulas expressed in terms of distributional terms and elasticities remain the same across such different models is a good illustration of the value of the sufficient statistics approach to optimal policy.

6 Conclusion

In this paper we propose a tractable new model for capital taxation, which creates a link to the policy debate and empirical analysis. We first presented a simple model with linear utility for consumption and concave utility for wealth which generates immediate adjustments of capital in response to taxes, a non-degenerate, smooth response of capital to taxes, and allows for arbitrary heterogeneity in preferences for capital, work, and discount rates.

We derive formulas for optimal linear and nonlinear capital income taxation which are expressed in terms of the elasticity of capital with respect to the net-of-tax rate of return, the shape of the capital income distribution, and the social welfare weights at each capital income level. We put the simplicity of this model to use by considering a range of policy issues such as the cases with joint-preferences and cross-elasticities between capital and labor, economic growth, heterogeneous returns to capital across individuals, different types of capital assets and heterogeneous tastes for each of them, or optimal taxes on comprehensive income.

We make use of our sufficient statistics formulas to numerically simulate optimal taxes based on U.S. tax data. Given how concentrated the distribution of capital is, the asymptotic tax rate for capital applies for the majority of capital income in the economy and should be higher than the top tax rate on labor income if the supply elasticities of labor and capital with respect to tax rates are the same. The theoretical framework we provide points to the key elasticities that need to be estimated in future work. These include the cross-elasticities between capital and labor and the elasticities and cross-elasticities of different types of capital assets, which it may

[49] In particular, deriving nonlinear optimal tax formulas is not tractable in the model with uncertainty while it is relatively simple in our model with certainty.

be optimal to tax differently.

Our approach is very amenable to incorporating alternative justice and fairness principles for capital taxation in an operational way. We discuss a range of ethical considerations regarding capital taxation that are salient in the policy debate. As long as, conditional on labor income, social marginal welfare weights depend directly on wealth (which is the case if wealth is perceived as unfairly distributed for many possible reasons) or are correlated with wealth (as in the case of the use of wealth as a tag), there is scope for capital taxation from an equity perspective. In future work, it would be very valuable to better understand society's equity considerations when it comes to capital taxation.

We show how our results extend to a model with a general, concave utility function as long as the elasticities of the capital tax base are appropriately adjusted to take into account transitional dynamics and potentially slow adjustments. The qualitative lessons from the simpler model carry over. However, the relevant quantitative elasticities depend on the normative framework chosen and we have tried to explain as clearly as possible the advantages and the drawbacks of each approach. The faster the adjustment of capital to taxes, the closer the quantitative results are across the three approaches and to those of the simpler model. The unanticipated reform approach has realism appeal but is normatively unattractive because it allows the government to heavily tax the existing capital stock and exploit sluggish responses. The anticipated reform approach, especially for very long-run tax rates, lacks realism both in terms of policy process and in terms of the infinite elasticities in generates in models without uncertainty. The utility based steady state approach has normative and simplicity appeal but we have not been able to make it fully consistent with a standard dynamic model. Our conjecture is that this is not possible. The issue of how to tax the existing capital stock and which of the three types of policy approaches to consider is irrelevant in the simpler model with linear utility from consumption, which is one of its key strengths in addition to its tractability. Our view is that concave utility of consumption and the consumption smoothing effects it creates introduce dilemmas which obscure the optimal capital tax problem and, yet, are probably second order for the normative and practical problem of taxing capital income that is mostly concentrated at the top.

References

Aiyagari, R. (1994). Uninsured idiosyncratic risk and aggregate savings. *Quarterly Journal of Economics 109*, 659–684.

Aiyagari, R. (1995). Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting. *Journal of Political Economy 103*(6), 1158–1175.

Alstadsaeter, A. and E. Fjaerli (2009). Neutral taxation of shareholder income? corporate responses to an announced dividend tax. *International Tax and Public Finance 16*(4), 571–604.

Atkinson, A. and A. Sandmo (1980). Welfare implications of the taxation of savings. *Economic Journal 90*, 529–549.

Atkinson, A. B. and J. E. Stiglitz (1976). The design of tax structure: direct versus indirect taxation. *Journal of Public Economics 6*(1), 55–75.

Auerbach, A. (2009). The choice between income and consumption taxes: A primer. In A. Auerbach and D. Shaviro (Eds.), *Institutional Foundations of Public Finance: Economic and Legal Perspectives*. Cambridge: Harvard University Press.

Auerbach, A. J. and L. J. Kotlikoff (1987). *Dynamic fiscal policy*, Volume 11. Cambridge University Press Cambridge.

Benhabib, J. and A. Bisin (2016). Skewed wealth distributions: Theory and empirics. *NYU Working Paper*.

Blanchard, O. (1985). Debt, Deficits, and Infinite Horizons. *Journal of Political Economy 93*, 223–247.

Cagetti, M. and M. De Nardi (2007). Entrepreneurship, Frictions, and Wealth. *Journal of Political Economy 114*, 835–870.

Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics 112*, 1–56.

Carroll, C. D. (2000). Why do the rich save so much? In J. B. Slemrod (Ed.), *Does Atlas Shrug? The Economic Consequences of Taxing the Rich.* Harvard University Press.

Chamley, C. (1986). Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica 80*(3), 607–622.

Chetty, R. and E. Saez (2005). Dividend taxes and corporate behavior: Evidence from the 2003 dividend tax cut. *Quarterly Journal of Economics 120*(3), 791–833.

Christophera, A. N. and B. R. Schlenker (2000). The impact of perceived material wealth and perceiver personality on first impressions. *Journal of Economic Psychology 21*(1), 1–19.

De Nardi, M. (2004). Wealth Inequality and Intergenerational Links. *Review of Economic Studies 71*, 743–768.

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

Diamond, P. A. and J. A. Mirrlees (1971). Optimal taxation and public production i: Production efficiency. *The American Economic Review 61*(1), 8–27.

Dittmar, H. (1992). Perceived material wealth and first impressions. *British Journal of Social Psychology 31*(4), 379–391.

Farhi, E., C. Sleet, I. Werning, and S. Yeltekin (2012). Nonlinear capital taxation without commitment. *Review of Economic Studies 79*(4), 1469–1493.

Fischer, S. (1980). Dynamic Inconsistency, Cooperation, and the Benevolent Dissembling Government. *Journal of Economic Dynamics and Control 2*, 93–107.

Golosov, M., A. Tsyvinski, and N. Kocherlakota (2003). Optimal indirect and capital taxation. *Review of Economic Studies 70*(3), 569–588.

Golosov, M., A. Tsyvinski, and I. Werning (2006). New dynamic public finance: a user's guide. In *NBER Macroeconomics Annual 2006, Volume 21*, pp. 317–388. MIT Press.

Gordon, R. and W. Kopczuk (2014). The Choice of Personal Income Tax Base. *Journal of Public Economics 118, 97-110.*

Hamilton, B. H. (2000). Does entrepreneurship pay? an empirical analysis of the returns to self?employment. *Journal of Political Economy 108*(3), 604–631.

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy 30*(1), 51–89.

Hurst, E. and B. W. Pugsley (2010). Non pecuniary benefits of small business ownership. *Chicago Booth Working Paper*.

Johannesen, N. (2014). Tax evasion and swiss bank deposits. *Journal of Public Economics 111*, 46–62.

Jones, L. E., R. E. Manuelli, and P. E. Rossi (1993). Optimal taxation in models of endogenous growth. *Journal of Political economy 101*(3), 485–517.

Judd, K. L. (1985). Redistributive taxation in a simple perfect foresight model. *Journal of public Economics 28*(1), 59–83.

Kaplow, L. (1994). Taxation and risk taking: A general equilibrium perspective. *National Tax Association 47*(4), 789–798.

Keynes, J. M. (1919). *The Economic Consequences of the Peace.* London: Macmillan.

Keynes, J. M. (1931). *Essays in Persuasion.* London: Macmillan.

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

Kiyotaki, N., A. Michaelides, and K. Nikolov (2011). Winners and losers in housing markets. *Journal of Money, Credit and Banking 43*(2-3), 255–296.

Kydland, F. and E. C. Prescott (1977). Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy 85*, 473–491.

Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies 38*, 175–208.

Novokmet, F., T. Piketty, and G. Zucman (2017). From soviets to oligarchs: Inequality and property in russia, 1905-2016. Working Paper 23712, National Bureau of Economic Research.

Piazzesi, M., M. Schneider, and S. Tuzel (2007). Housing, consumption and asset pricing. *?Journal of Financial Economics 83*, 531–569.

Piketty, T. (2000). Theories of persistent inequality and intergenerational mobility. *Handbook of income distribution 1*, 429–476.

Piketty, T. (2011). On the long-run evolution of inheritance: France 1820-2050. *Quarterly journal of economics 61*(3), 1071–1131.

Piketty, T. (2014). *Capital in the Twenty-first Century.* Harvard University Press.

Piketty, T. and E. Saez (2003). Income inequality in the united states, 1913–1998. *The Quarterly Journal of Economics 118*(1), 1–41.

Piketty, T. and E. Saez (2013a). Optimal labor income taxation. In M. F. Alan J. Auerbach, Raj Chetty and E. Saez (Eds.), *Handbook of Public Economics, Vol. 5*, pp. 391–474. Amsterdam: North-Holland.

Piketty, T. and E. Saez (2013b). A theory of optimal inheritance taxation. *Econometrica 82*(4), 1241–1272.

Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy 6*(1), 230–71.

Piketty, T., E. Saez, and G. Zucman (2016). Distributional national accounts: Methods and estimates for the united states. NBER Working Paper 22945, NBER.

Poterba, J. and J. Rotemberg (1987). Money in the utility function: An empirical implementation. In W. Barnett and K. Singleton (Eds.), *New Approaches to Monetary Economics.*, pp. 219–240. Cambridge University Press.

Quadrini, V. (1999). The Importance of Entrepreneurship for Wealth Concentration and Mobility. *Review of Income and Wealth 45*(1), 1–19.

Quadrini, V. (2000). Entrepreneurship, Savings and Social Mobility. *Review of Economic Dynamics 3*, 1–40.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*(1), 205–229.

Saez, E. (2017). Taxing the rich more: Preliminary evidence from the 2013 tax increase. *Tax Policy and the Economy 31*(1), 71–120.

Saez, E., J. Slemrod, and S. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 3–50.

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review 106*(1), 24–45.

Saez, E. and G. Zucman (2016). Wealth inequality in the united states since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics 131*(2), 519–578.

Smith, A. (1759). *The Theory of Moral Sentiments.*

Stokey, N. (2009). Moving costs, nondurable consumption and portfolio choice. *?Journal of Economic Theory 144*, 2419?2439.

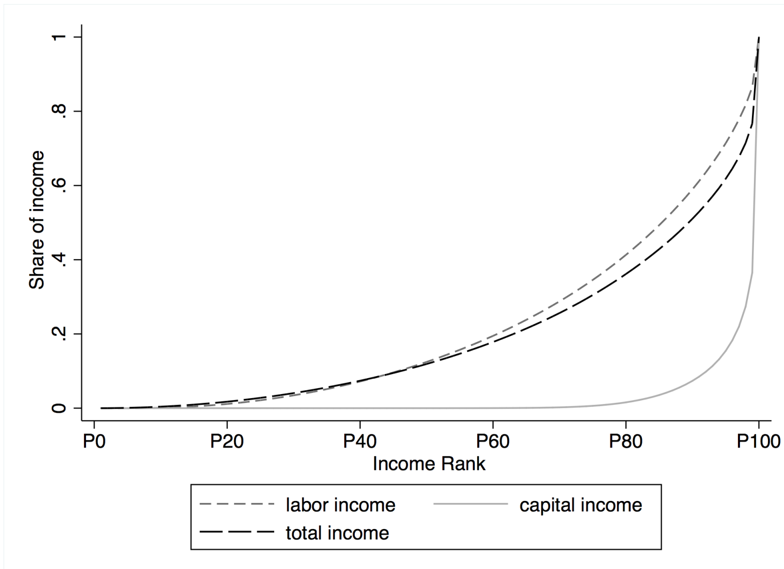
Warren, A. C. (1980). Would a consumption tax be fairer than an income tax? *The Yale Law Journal Volume 89*(Number 6).

Weber, M. (1958). *The Protestant Ethic and the Spirit of Capitalism.* Charles Scribner and Sons, New York.

Weisbach, D. A. and J. Bankman (2006). The Superiority of an Ideal Consumption Tax over an Ideal Income Tax. *58 Stanford Law Review 1413*.

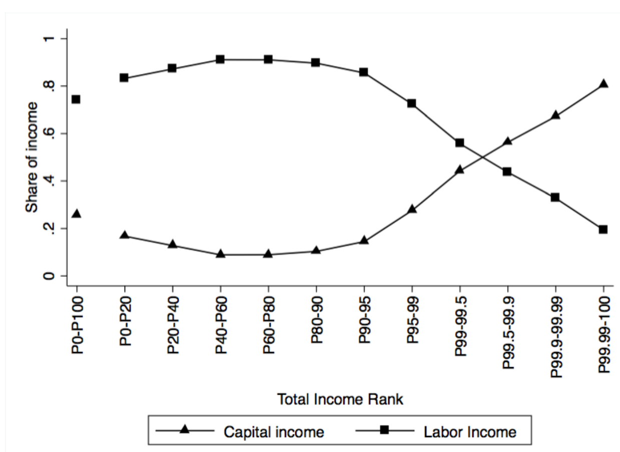
Yaari, M. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies 32*, 137–150.

Figure 1: Lorenz Curves for capital, labor, and total income



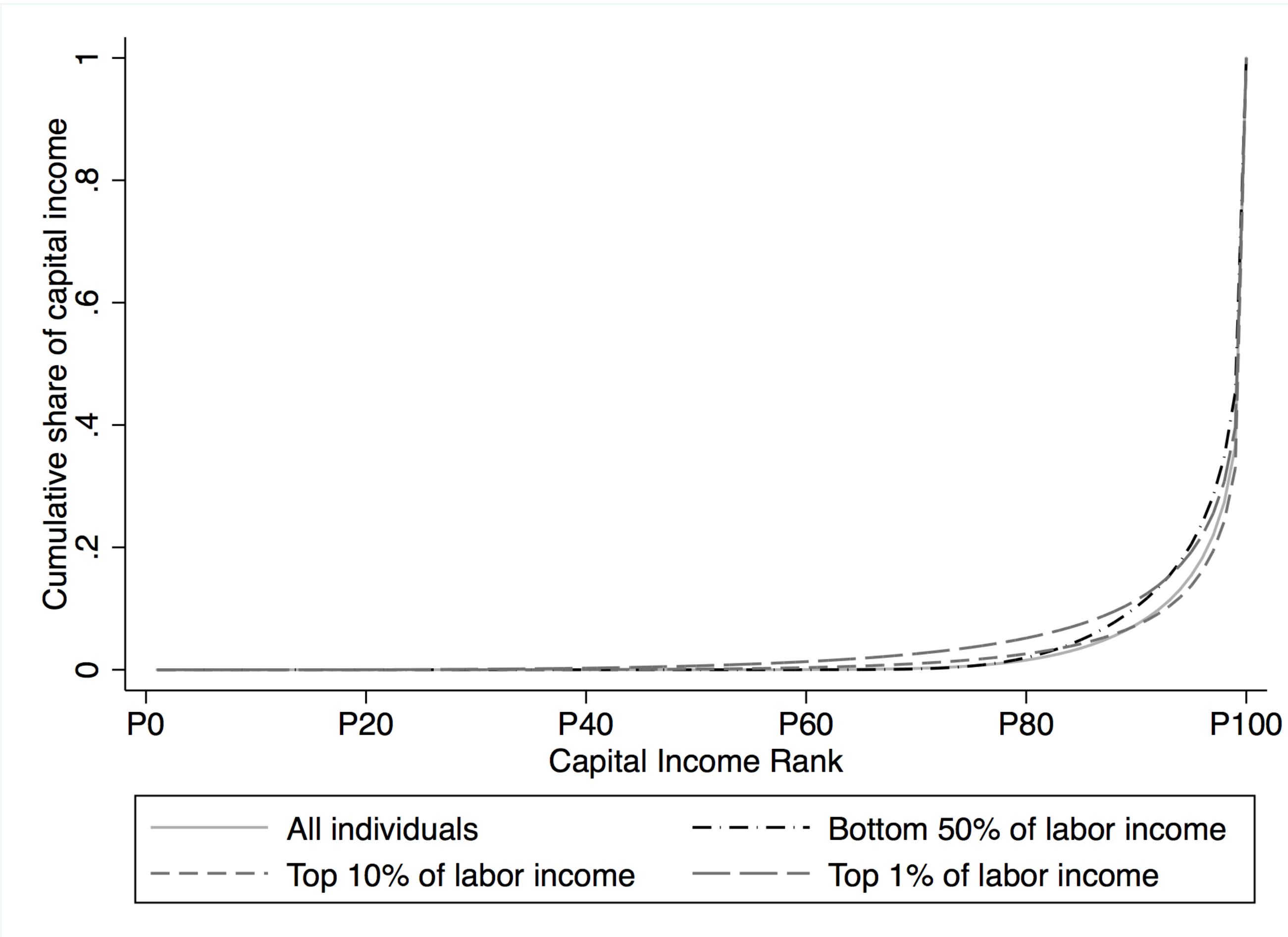
Notes: Computations based on IRS tax return data for year 2007. The figure represents the Lorenz curves for labor income, capital income, and total income (capital + labor income). The Lorenz curve is the cumulative share of income owned by those below each income percentile (x-axis). The distributions of both labor and capital income (and, thus, of total income) exhibit great inequalities, but capital income is much more concentrated than labor income.

Figure 2: Capital and labor incomes as a share of total income



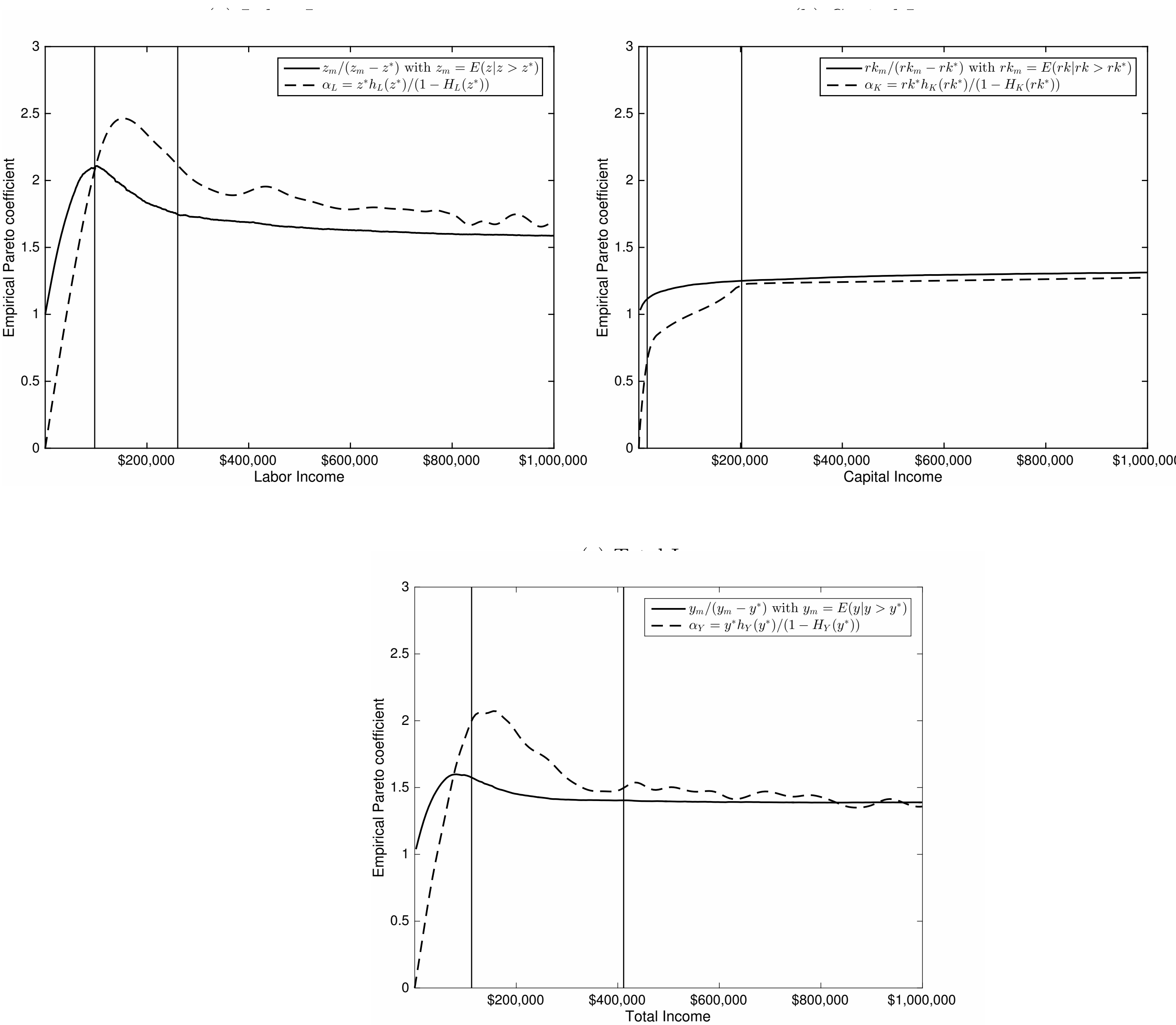
Notes: Computations based on IRS tax return data for year 2007. The figure shows the composition of total income within several groups, ranked by total income, and marked on the horizontal axis. The first observation represents the overall population P0-P100. P0-P20 denotes the bottom 20% tax units, etc. At the top of the income distribution, most of total income comes from capital income.

Figure 3: Two-dimensional heterogeneity:
Lorenz curves for capital, conditional in labor income



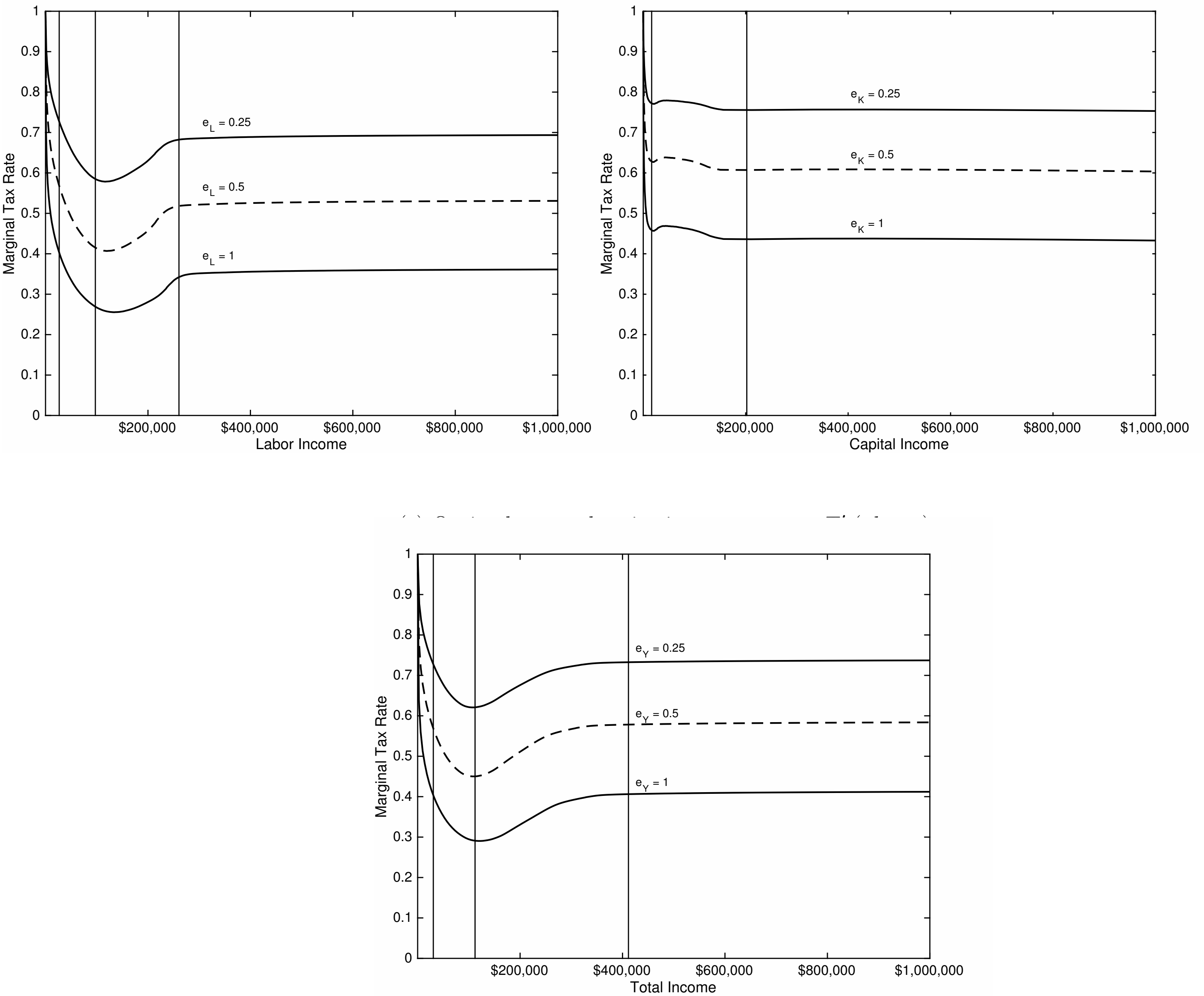
Notes: Computations based on IRS tax return data for year 2007. The figure depicts the Lorenz curves for capital income (the Lorenz curve is the cumulative share of capital income owned by those below each percentile of the capital income distribution), for four groups defined by labor income: all individuals, the bottom 50% by labor income, the top 10% by labor income and the top 1% by labor income. Even conditional on labor income, there are large inequalities in capital income. Put differently, there is a lot of two-dimensional heterogeneity in both labor and capital income.

Figure 4: Empirical Pareto parameters



Notes: Computations based on IRS tax return data for year 2007. The figure depicts the empirical Pareto parameters for the labor income distribution (panel (a)), the capital income distribution (panel (b)) and the total income distribution (panel (c)). For labor income, we compute the top bracket Pareto parameter $z_m/(z_m - z^*)$ relevant for the optimal linear tax rate above z^* and the local Pareto parameter $\alpha(z^*) = z^* h_L(z^*)/(1 - H_L(z^*))$ where $h_L(z)$ is the density and $H_L(z)$ the cumulated distribution, which is relevant for the optimal nonlinear $T'_L(z^*)$. The x-axis depicts z^*. The vertical lines depict the 90th and 99th percentiles of each distribution. We repeat the same for capital income rk and total income $y = rk + z$. At the top, all three distributions are very well approximated by Pareto distributions with constant tail parameters of around $a_L = 1.6$ for labor, $a_K = 1.38$ for capital, and $a_Y = 1.4$ for total income.

Figure 5: Optimal Marginal Tax Rates



Notes: Computations based on IRS tax return data for year 2007. Optimal marginal tax rates based on the formulas in Section 2.2.2. Panel (a) plots the optimal marginal tax rate on labor income. Panel (b) plots the optimal marginal tax rate on capital income. Panel (c) plots the optimal marginal tax rate on total income. In each panel, optimal marginal tax rates are plotted for three different elasticity values: 0.25, .5, and 1. In each panel, the three vertical lines represent, respectively, the median, the top 10% and the top 1% thresholds of the 2007 the labor, capital, and total income distributions in the U.S. (the median capital income is zero).

Table 1: Comparison of Elasticities and Taxes in Capital Taxation Models

Utility (1)	Transitional Dynamics? (2)	Uncertainty or Certainty? (3)	Reform anticipated or unanticipated? (4)	Model (5)	e_K^a (6)	e_K^u (7)	e_K (8)	Optimal τ_K (9)
Wealth in the Utility	No	Certainty	Either	Sections 2 and 3	$= e_K$	$= e_K$	$< \infty$	> 0
	Yes	Certainty	Anticipated at T	Section 5.2	∞ if $T \to \infty$		$< \infty$	0 if $T \to \infty$
			Unanticipated	Section 5.1		$< e_K$	$< \infty$	> 0
Standard	Yes	Uncertainty	Anticipated at T	Aiyagari (1995)	$< \infty$		$< \infty$	> 0
			Unanticipated	Section 5.1		$< e_K$	$< \infty$	> 0
	Yes	Certainty	Anticipated at T	Chamley-Judd	∞ if $T \to \infty$		∞	0 if $T \to \infty$
			Unanticipated	Section 5.1		$< e_K$	∞	> 0
Endogenous $\delta(c_i)$	Yes	Certainty	Anticipated at T	Judd (1985)	∞ if $T \to \infty$		$< \infty$	0 if $T \to \infty$
			Unanticipated	Section 5.1		$< e_K$	$< \infty$	> 0

Notes: This table presents a comparison of supply elasticities of capital with respect to the net-of-tax rate of return and optimal capital income tax rates across various models. Column (1) indicates the type of utility function. Column (2) indicates whether there are transitional dynamics (which is equivalent to whether the utility is linear vs. concave in consumption). Column (3) indicates whether there is uncertainty in future labor incomes and preferences. Column (4) indicates whether the tax reform determining the optimal tax rate is anticipated (at time T) or unanticipated (at time 0). Column (5) indicates the Section in the paper covering the model or whether an existing paper in the literature covers it. Columns (6)-(8) describe the magnitude of the three elasticities: e_K the long-run steady state elasticity from (23), e_K^u the unanticipated elasticity defined in (25), and e_K^a the anticipated elasticity defined in (27). Column (9) describes the sign and magnitude of the optimal linear tax rate τ_K on capital income. It is assumed that $\bar{g}_K < 1$ so that taxing capital income is desirable (absent any behavioral response). Adding wealth in the utility to the Aiyagari model does not change the predictions.

Online Appendix – Not For Publication

A.1 Proofs for Section 2

Proof of Proposition 2.

We derive the optimal capital tax. The optimal labor tax is derived exactly in the same way.

Consider a small reform $\delta T_K(rk)$ in which the marginal tax rate is increased by $\delta\tau_K$ in a small band from capital income rk to $rk+d(rk)$, but left unchanged anywhere else. This reform has a mechanical revenue effect, a behavioral effect, and a welfare effect.

The mechanical revenue effect above capital income rk is

$$d(rk)\delta\tau_K[1-H_K(rk)].$$

The behavioral effect comes only from taxpayers with capital income in the range $[rk, rk+d(rk)]$. Thanks to the linear utility (i.e., no income effects), taxpayers above rk do not respond to the tax rates since they do not face a change in their marginal tax rate. Taxpayers in the small band have a behavioral response to the higher marginal tax rate. They each reduce their capital income by $\delta(rk) = -e_K\delta\tau_K/(1-T_K'(rk))$ where e_K is the elasticity of capital income rk with respect to the net-of-tax return $r(1-T_K'(rk))$. As there are $h_K(rk)d(rk)$ taxpayers affected by the change in marginal tax rates, the resulting loss in tax revenue is equal to:

$$-d(rk)\delta\tau_K \cdot h_K(rk)e_K(rk)rk\frac{T_K'(rk)}{(1-T_K'(rk))},$$

with $e_K(rk)$, as defined in the text, the average elasticity of capital income in the small band.

The change in tax revenue is rebated lump-sum to all taxpayers. The value of this lump-sum transfer to society is $\int_i g_i = 1$ due to the absence of income effects (the lump-sum rebate also does not change any behavior with linear utility).

By definition of the average social marginal welfare weight above rk, $\bar{G}_K(rk)$, in (10), the welfare effect on the tax payers above rk who pay more tax $\delta\tau_K \cdot d(rk)$ is:

$$-\delta\tau_K \cdot d(rk)\int_{i:rk_i\geq rk} g_i = -\delta\tau_K \cdot d(rk)(1-H_K(rk))\bar{G}_K(rk).$$

At the optimum, the sum of the mechanical revenue effect, the behavioral effect, and the welfare effect needs to be zero, which requires that:

$$d(rk)\delta\tau_K \cdot \left[1 - H_K(rk) - h_K(rk) \cdot e_K(rk) \cdot rk \cdot \frac{T'_K(rk)}{1 - T'_K(rk)}\right] \\ - d(rk)\delta\tau_K \cdot (1 - H_K(rk)) \cdot \bar{G}_K(rk) = 0.$$

We can divide everything by $d(rk)\delta\tau_K$ and re-arrange to obtain:

$$\frac{T'_K(rk)}{1 - T'_K(rk)} = \frac{1}{e_K(rk)} \cdot \frac{1 - H_K(rk)}{rk \cdot h_K(rk)} \cdot (1 - \bar{G}_K(rk)).$$

Using the definition of the local Pareto parameter $\alpha_K(rk) = rkh_K(rk)/(1 - H_K(rk))$, we obtain the capital tax formula in the proposition. The optimal marginal labor tax formula is derived in the same way, replacing capital income rk with labor income z.

Proof of Proposition 3.

The derivation of the optimal tax on comprehensive income follows exactly the proof of Proposition 2, replacing capital income rk with total income y.

Proof of Proposition 4.

The government maximizes:

$$SWF = \int_i \omega_i U_i(c_i, k_i, z_i, x_i)$$

$$\text{with} \quad U_i(c_i, k_i, z_i, x_i) = \bar{r}k_i + (1 - \tau_L)z_i + (\tau_L - \tau_K)x_i + \tau_L(z^m - x^m) \\ + \tau_K(rk^m + x^m) + a_i(k_i) - h_i(z_i) - d_i(x_i) + \delta_i \cdot (k_i^{init} - k_i).$$

The first order conditions with respect to τ_L and τ_K are:

$$\int_i \omega_i(z^m - x^m - (z_i - x_i)) - \tau_L \frac{dz^m}{d(1 - \tau_L)} - (\tau_L - \tau_K)\frac{dx^m}{d\tau_L} = 0,$$

$$\int_i \omega_i(rk^m + x^m - (rk_i + x_i)) - \tau_K r \frac{dk^m}{d(1 - \tau_K)} - (\tau_L - \tau_K)\frac{dx^m}{d\tau_K} = 0.$$

Since x_i depends only on $\tau_L - \tau_K$, we have that: $\frac{dx^m}{d\tau_L} = -\frac{dx^m}{d\tau_K} = \frac{dx^m}{d(\tau_L - \tau_K)}$. Let $\Delta\tau \equiv \tau_L - \tau_K$. The FOCs can be rewritten as:

$$\frac{z^m - x^m - \int_i \omega_i(z_i - x_i)}{\frac{dz^m}{d(1-\tau_L)}} - \Delta\tau \frac{\frac{dx^m}{d(\tau_L - \tau_K)}}{\frac{dz^m}{d(1-\tau_L)}} = \tau_L,$$

$$\frac{rk^m + x^m - \int_i \omega_i(rk_i + x_i)}{r\frac{dk^m}{d(1-\tau_K)}} + \Delta\tau \frac{\frac{dx^m}{d(\tau_L-\tau_K)}}{r\frac{dk^m}{d(1-\tau_K)}} = \tau_K.$$

Let us simplify notation a bit and denote:

$$z' \equiv \frac{dz^m}{d(1-\tau_L)} \quad k' \equiv \frac{dk^m}{d(1-\tau_K)} \quad x' \equiv \frac{dx^m}{d(\tau_L-\tau_K)}.$$

Taking the difference of those two equations, we can express $\Delta\tau$ as

$$\Delta\tau\left(1 + x'\left(\frac{1}{z'} + \frac{1}{rk'}\right)\right) = \frac{z^m - x^m - \int_i \omega_i(z_i - x_i)}{z'} - \frac{rk^m + x^m - \int_i \omega_i(rk_i + x_i)}{rk'}, \qquad \text{(A1)}$$

Since $\left(1 + x'\left(\frac{1}{z'} + \frac{1}{rk'}\right)\right) > 0$, the sign of $\Delta\tau$ is that of the right-hand side of the above expression.

$$\Delta\tau > 0 \Leftrightarrow \frac{z^m - x^m - \int_i \omega_i(z_i - x_i)}{z'} > \frac{rk^m + x^m - \int_i \omega_i(rk_i + x_i)}{rk'}.$$

Define the distributional factor of shifted income, by analogy to the distributional factors $\bar{g}_K$ and $\bar{g}_L$ for capital and labor income.

$$\bar{g}_X = \frac{\int_i \omega_i x_i}{z^m}.$$

The right-hand side of (A1) can be rewritten as:

$$RHS = \frac{1 - \frac{x^m}{z^m} - \bar{g}_L + \bar{g}_X}{\frac{e_L}{1-\tau_L}} - \frac{1 + \frac{x^m}{rk^m} - \bar{g}_K - \bar{g}_X\frac{z^m}{rk^m}}{\frac{e_K}{1-\tau_K}}.$$

Hence:

$$\Delta\tau > 0 \Leftrightarrow \frac{1 - \frac{x^m}{z^m} - \bar{g}_L + \bar{g}_X}{\frac{e_L}{1-\tau_L}} > \frac{1 + \frac{x^m}{rk^m} - \bar{g}_K - \bar{g}_X\frac{z^m}{rk^m}}{\frac{e_K}{1-\tau_K}}.$$

Suppose that $\bar{g}_X$ is small enough – otherwise, encouraging "shifting" may be good for distributional reasons. Formally, for $x^m > 0$,

$$\frac{x^m}{rk^m} - \bar{g}_X\frac{z^m}{rk^m} > 0 \quad \text{and} \quad \frac{x^m}{z^m} - \bar{g}_X > 0.$$

Conversely, for $x^m < 0$, we have $\bar{g}_X < 0$, and we assume that $\bar{g}_X$ is small relative to x^m in absolute value.

$$\frac{x^m}{rk^m} - \bar{g}_X\frac{z^m}{rk^m} < 0 \quad \text{and} \quad \frac{x^m}{z^m} - \bar{g}_X < 0.$$

We can then write:

$$\Delta\tau > 0 \Leftrightarrow e_K > e_L \cdot \frac{1-\tau_K}{1-\tau_L} \cdot \frac{\left(1+\frac{x^m}{rk^m} - \bar{g}_K - \bar{g}_X \frac{z^m}{rk^m}\right)}{\left(1-\frac{x^m}{z^m} - \bar{g}_L + \bar{g}_X\right)}.$$

If $\Delta\tau = 0$, there is no shifting and hence $x_i = 0$ for all i and $x^m = 0$, and hence $\bar{g}_X = 0$. Therefore,

$$\text{If} \quad \Delta\tau = 0: \quad e_K = e_L \frac{(1-\bar{g}_K)}{(1-\bar{g}_L)}.$$

If $\Delta\tau > 0$, then $x^m > 0$ and $e_K > e_L \frac{(1-\bar{g}_K)}{(1-\bar{g}_L)}$.
Conversely, if $\Delta\tau < 0$, then $x^m < 0$ and $e_K < e_L \frac{(1-\bar{g}_K)}{(1-\bar{g}_L)}$.
Thus:

$$\Delta\tau \lesseqqgtr 0 \Leftrightarrow e_K \lesseqqgtr e_L \frac{(1-\bar{g}_K)}{(1-\bar{g}_L)}.$$

We can now rewrite the FOCs as:

$$z^m(1 - \frac{x^m}{z^m} - \bar{g}_L + \bar{g}_X) - \Delta\tau x' = z^m e_L \frac{\tau_L}{1-\tau_L},$$

$$rk^m(1 + \frac{x^m}{rk^m} - \bar{g}_K - \bar{g}_X \frac{z^m}{rk^m}) + \Delta\tau x' = rk^m e_K \frac{\tau_K}{1-\tau_K}.$$

We distinguish three cases:

- If $e_K > e_L \frac{(1-\bar{g}_K)}{(1-\bar{g}_L)}$, then $\Delta\tau > 0$ and

$$e_L \frac{\tau_L}{1-\tau_L} < 1 - \frac{x^m}{z^m} - \bar{g}_L + \bar{g}_X < 1 - \bar{g}_L.$$

  and in this case:

$$e_K \frac{\tau_K}{1-\tau_K} > (1 + \frac{x^m}{rk^m} - \bar{g}_K - \bar{g}_X \frac{z^m}{rk^m}) > 1 - \bar{g}_K.$$

  So that the optimal tax rates with shifting are bracketed by their revenue maximizing rates.

- If there is no shifting, $x \equiv 0$ then revenue maximizing rates apply.

- If x' is very large (very sensitive shifting to any tax differential), then from equation (A1), we have that $\Delta\tau \approx 0$ and hence $\tau_L \approx \tau_K$. Summing the FOCs and using this equality yields $\tau_L = \tau_K = \tau_Y$ where τ_Y is the optimal linear tax rate on comprehensive income derived in Proposition 3.

Proof of Proposition 5.

Let us compare the following two regimes considered in the text:

Regime 1 – Consumption tax regime: $(\bar{r}, T_L, \tau_C)$, with an initial lump-sum transfer $\tau_C \cdot k_i^{init}/(1-\tau_C)$ to wealth holders with initial wealth k_i^{init}.

Regime 2 – No consumption tax regime: $(\bar{r}, \hat{T}_L, \tau_C = 0)$ with $(z - \hat{T}_L(z)) = (z - T_L(z)) \cdot (1-\tau_C)$. Let $\tilde{k}_i$ denote the steady state wealth choice under this regime.

We will show that these regimes are equivalent in the steady state, in the consumer's dynamic optimization problem, and in the government's revenue raised, as claimed in the text.

Steady-state equivalence:

The budget constraint in regime 1 is: $\dot{k} = [\bar{r}k + z - T_L(z)] - c/(1-\tau_C) + G$, where $G = \tau_L z^m + \tau_K r k^m + t_C c^m$ is the lump-sum transfer rebate of tax revenue. The budget constraint can be rewritten in terms of real wealth as: $\dot{k}^r = \bar{r}k^r + (z - T_L(z)) \cdot (1-\tau_C) + G \cdot (1-\tau_C) - c$.

Utility is:

$$u_i = c_i + a_i(k_i^r) - h_i(z_i).$$

The first-order conditions of the individual are:

$$(1 - T_L'(z_i)) \cdot (1-\tau_C) = h_i'(z_i), \quad a_i'(k_i^r) = \delta_i - \bar{r}.$$

Given that $(1-\hat{T}_L'(z_i)) = (1-T_L'(z_i)) \cdot (1-\tau_C)$ for all z_i, the steady-state choices of labor income and real capital of the individual are unaffected. Using the steady state budget constraint, real consumption c_i is also not affected as long as the real lump-sum transfer $G \cdot (1-\tau_C)$ is not affected, which we prove right below. The link between the two capital levels is: $\tilde{k}_i = (1-\tau_C) \cdot k_i$ (since real steady state wealth is unaffected).

Equivalence of the dynamic consumer optimization problem.

The law of motion in real-wealth equivalent, $\dot{k}^r = \bar{r}k^r + (z - T_L(z)) \cdot (1-\tau_C) + G \cdot (1-\tau_C) - c$, is the same in regime 1 and regime 2 as long as the real lump-sum transfer $(1-\tau_C) \cdot G$ is the same, which we show below. The initial wealth after the lump-sum transfer $\tau_C \cdot k_{init}^i/(1-\tau_C)$ from the government becomes $k_i^{init} + \tau_C \cdot k_i^{init}/(1-\tau_C) = k_i^{init}/(1-\tau_C)$, so that real wealth after the transfer is k_i^{init}, the same it was in the tax regime without a consumption tax.

Equivalence of government revenue.

In regime 1, there is first the initial cost of providing the lump-sum $\tau_C \cdot \int_i k_i^{init}/(1-\tau_C)$ to all initial wealth holders. At the same time, the initial consumption change is taxed, which yields: $\tau_C \cdot \int_i (k_i^{init} - k_i)/(1-\tau_C)$.

In real terms, this is worth:

$$A = -\tau_C \cdot \int_i k_i.$$

The nominal tax flow per period under this regime is (which is also equal to the lump-sum transfer per-period in nominal terms is G:

$$G = \frac{\tau_C}{1-\tau_C} \int_i c_i + \int_i T_L(z_i) + \int_i \tau_K r k_i.$$

We can express consumption under this regime as:

$$c_i = (z_i - T_L(z_i))(1-\tau_C) + \bar{r}(1-\tau_C)k_i + G(1-\tau_C).$$

and aggregate consumption as:

$$\int_i c_i = (1-\tau_C) \int_i (z_i - T_L(z_i)) + \bar{r}(1-\tau_C) \int_i k_i + G(1-\tau_C).$$

Solving for G using the definition of G and the expression for aggregate consumption yields:

$$G = \int_i T_L(z_i) + \frac{\tau_C}{1-\tau_C} \left(\int_i z_i + \bar{r} \int_i k_i \right) + \frac{1}{1-\tau_C} \int_i \tau_K r k_i.$$

In real terms, revenue is:

$$(1-\tau_C) \cdot G = (1-\tau_C) \int_i T_L(z_i) + \tau_C \int_i z_i + \tau_C \bar{r} \int_i k_i + \int_i \tau_K r k_i.$$

In Regime 2, the (real) revenue is:

$$\int_i \hat{T}_L(z_i) + \int_i \tau_K r \tilde{k}_i.$$

Using the map between the labor income taxes: $(z - \hat{T}_L(z)) = (z - T_L(z)) \cdot (1-\tau_C)$, we obtain that the real revenue in Regime 2 is:

$$\int_i (\tau_C z + T_L(z) \cdot (1-\tau_C)) + \int_i \tau_K r \tilde{k}_i.$$

The difference between the per-period real revenue in regime 1 and that in regime 2 is hence: $\tau_C \int_i r k_i$. Recall that the initial change in revenue in regime 1 was $A = -\tau_C \cdot \int_i k_i$, which, converted into a per-period equivalent is exactly $A \cdot r = -\tau_C \cdot \int_i r k_i$ and cancels out perfectly the change in per-period revenue between the two regimes.

Proof of Proposition 6.

Let G be government revenue which is rebated lump-sum. The change in revenue from a change in the capital income tax $d\tau_K$ is:

$$dG = rk^m \left[1 - \frac{\tau_K}{1-\tau_K} \cdot e_K - \frac{\tau_L}{1-\tau_K} e_{L,(1-\tau_K)} \frac{z^m}{rk^m}\right] \cdot d\tau_K.$$

Hence the change in social welfare is:

$$\frac{dSWF}{d\tau_K} = \int_i g_i \left(-rk_i + \frac{dG}{d\tau_K}\right) = \left(\int_i g_i\right) \cdot \left(-\frac{\int_i g_i r k_i}{\int_i g_i} + \frac{dG}{d\tau_K}\right).$$

Setting this to zero and using the definition of $\bar{g}_K = \frac{\int_i g_i k_i}{\int_i g_i k^m}$, yields:

$$\tau_K = \frac{1 - \bar{g}_K - \tau_L e_{L,(1-\tau_K)} \frac{z^m}{rk^m}}{1 - \bar{g}_K + e_K},$$

which is the optimal capital tax formula with joint preferences and cross-elasticities. The optimal labor tax formula with cross elasticities can be derived exactly symmetrically.

A.2 Proofs for Section 5 and Extensions

A.2.1 Unanticipated Reforms

Proof of Proposition 8

The steady state is characterized by: $u_{ik}/u_{ic} = \delta_i - r(1 - T'_K)$, $u_{ic} \cdot (1 - T'_L) = -u_{iz}$ and $c_i = rk_i + z_i - T(z_i, rk_i)$. With linear taxes, this simplifies to: $u_{ik}/u_{ic} = \delta_i - \bar{r}$, $u_{ic} \cdot (1 - \tau_L) = -u_{iz}$ and $c_i = \bar{r}k_i + z_i(1 - \tau_L)$.

Let us assume that the economy has converged to the steady state. Consider a small reform $d\tau_K$ that takes place at time 0 and is unanticipated. Let us denote by $e^u_K(t)$ the elasticity of aggregate $k^m(t)$ with respect to the net of tax rate $r(1 - \tau_K)$. Using the envelope theorem (i.e., behavioral responses dk_{ti} can be ignored when computing dV_i), the effect on the welfare of individual i is:

$$dV_i = d\tau_K \cdot \delta_i \Bigg[\int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) r k^m(t) \cdot e^{-\delta_i t} - \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) r k_i(t) \cdot e^{-\delta_i t}$$
$$- \frac{\tau_K}{1-\tau_K} \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) r k^m(t) e_K^u(t) \cdot e^{-\delta_i t} dt$$
$$- \frac{\tau_L}{1-\tau_K} \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) e_{L,1-\tau_K}^u(t) z^m(t) e^{-\delta_i t} dt \Bigg].$$

In the steady state, $k^m(t)$, $z^m(t)$, $c_i(t)$, $z_i(t)$, and $k_i(t)$ are time-constant, so that the change in individual i's utility is:

$$dV_i = d\tau_K \cdot r k^m \Bigg[u_{ic}(c_i, k_i . z_i) - u_{ic}(c_i, k_i, z_i) \frac{k_i}{k^m}$$
$$- \frac{\tau_K}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \int_0^\infty e_K^u(t) \cdot e^{-\delta_i t} dt - \frac{\tau_L}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \frac{z^m}{r k^m} \int_0^\infty e_{L,1-\tau_K}^u(t) \cdot e^{-\delta_i t} dt \Bigg].$$

and the change in social welfare is:

$$dSWF = \int_i \omega_i dV_i = \int_i d\tau_K \cdot r k^m \omega_i \Bigg[u_{ic}(c_i, k_i, z_i) - u_{ic}(c_i, k_i, z_i) \frac{k_i}{k^m}$$
$$- \frac{\tau_K}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \int_0^\infty e_K^u(t) \cdot e^{-\delta_i t} dt - \frac{\tau_L}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \frac{z^m}{r k^m} \int_0^\infty e_{L,1-\tau_K}^u(t) \cdot e^{-\delta_i t} dt \Bigg].$$

Using the normalization of social welfare weights: $\int_i \omega_i u_{ic} = 1$ and $g_i = \omega_i u_{ic}$.

$$dSWF \propto 1 - \int_i \frac{g_i k_i}{k^m} - \frac{\tau_K}{1-\tau_K} \int_i \delta_i g_i \int_0^\infty e_K^u(t) e^{-\delta_i t} dt - \frac{\tau_L}{1-\tau_K} \frac{z^m}{r k^m} \int_i \delta_i g_i \int_0^\infty e_{L,1-\tau_K}^u(t) e^{-\delta_i t} dt,$$

which yields:

$$\tau_K = \frac{1 - \bar{g}_K - \tau_L \frac{z^m}{r k^m} e_{L,(1-\tau_K)}^u}{1 - \bar{g}_K + e_K^u} \quad \text{with} \quad e_K^u = \int_i g_i \delta_i \int_{t=0}^\infty e_K^u(t) \cdot e^{-\delta_i t} dt$$
$$e_{L,(1-\tau_K)}^u = \int_i g_i \delta_i \int_{t=0}^\infty e_{L,(1-\tau_K)}^u(t) \cdot e^{-\delta_i t} dt \quad \text{and} \quad \bar{g}_K = \int_i g_i \cdot k_i / k^m.$$

With homogeneous discount rate $\delta_i = \delta$, we obtain the formula in the text.

Note that the steady state condition can be rewritten as: $(\delta_i - \bar{r}) u_{ci}(\bar{r} k_i + z_i, k_i, z_i) = u_{ki}(\bar{r} k_i + z_i, k_i, z_i)$ which is a smooth function of k_i and z_i, as long as the function $u_i(c_i, k_i, z_i)$ is smooth and concave in consumption and capital. Hence, the responses of consumption and capital to the net-of-tax return $\bar{r}$ are smooth and non-degenerate. The steady state elasticities are thus finite.

A.2.2 Anticipated Reforms

Proof of Proposition 9

Consider anticipated reform to the capital income tax $d\tau_K$ at time $T > 0$. Capital and labor already start adjusting in anticipation of the reform before time T. The change in the utility of individual i is:

$$dV_i = d\tau_K \cdot \delta_i \Bigg[\int_T^{\infty} u_{ic}(c_i(t), k_i(t), z_i(t)) r k^m(t) \cdot e^{-\delta_i t} dt - \int_T^{\infty} u_{ic}(c_i(t), k_i(t), z_i(t)) r k_i(t) \cdot e^{-\delta_i t} dt \\ - \frac{\tau_L}{1-\tau_K} \int_0^{\infty} u_{ic}(c_i(t), k_i(t), z_i(t)) e^a_{L,1-\tau_K}(t) z^m(t) e^{-\delta_i t} dt \\ - \frac{\tau_K}{1-\tau_K} \int_0^{\infty} u_{ic}(c_i(t), k_i(t), z_i(t)) r k^m(t) e^a_K(t) \cdot e^{-\delta_i t} dt \Bigg].$$

In the steady state, $k^m(t)$, $z^m(t)$, $c_i(t)$, $k_i(t)$, and $z_i(t)$ are time-constant, hence we have:

$$dV_i = d\tau_K r k^m e^{-\delta_i T} \cdot \Bigg[u_{ic}(c_i, k_i, z_i) - u_{ic}(c_i, k_i, z_i) \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \int_0^{\infty} e^a_K(t) \cdot e^{-\delta_i (t-T)} dt \\ - \frac{\tau_L}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \frac{z^m}{r k^m} \int_0^{\infty} e^a_{L,1-\tau_K}(t) \cdot e^{-\delta_i (t-T)} dt \Bigg].$$

With homogeneous discount rates, and using that $\int_i g_i = \int_i u_{ci} \omega_i = 1$, we can write $dSWF = \int_i \omega_i dV_i$:

$$dSWF \propto 1 - \int_i g_i \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} \delta \int_{t>0} e^a_K(t) \cdot e^{-\delta(t-T)} dt - \frac{\tau_L}{1-\tau_K} \frac{z^m}{r k^m} \delta \int_{t>0} e^a_{L,1-\tau_K}(t) \cdot e^{-\delta(t-T)} dt$$

which yields the formula in the text.

Anticipated reform with heterogeneous discount rates

With heterogeneous discount rates, the change in social welfare is:

$$dSWF \propto \int_i \omega_i e^{-\delta_i T} \cdot \Bigg[u_{ic}(c_i, k_i, z_i) - u_{ic}(c_i, k_i, z_i) \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \int_0^{\infty} e^a_K(t) \cdot e^{-\delta_i (t-T)} dt \\ - \frac{\tau_L}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \frac{z^m}{r k^m} \int_0^{\infty} e^a_{L,1-\tau_K}(t) \cdot e^{-\delta_i (t-T)} dt \Bigg]$$

Define the normalized social welfare weight $g_i(T) \equiv \frac{\omega_i u_{ci} e^{-\delta_i T}}{\int_i \omega_i u_{ci} e^{-\delta_i T}}$ which depends on the time of the reform and rewrite:

$$dSWF \propto 1 - \int_i g_i(T) \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} \int_i \delta_i g_i(T) \int_0^\infty e^a_K(t) \cdot e^{-\delta_i(t-T)} dt$$
$$- \frac{\tau_L}{1-\tau_K} \frac{z^m}{rk^m} \int_i \delta_i g_i(T) \int_0^\infty e^a_{L,1-\tau_K}(t) \cdot e^{-\delta_i(t-T)} dt \Big]$$

which yields the same optimal tax formula as in the text, but with

$$\bar{g}_K = \int_i g_i(T) \cdot k_i/k^m, \quad e^a_K = \int_i \delta_i g_i(T) \int_0^\infty e^a_K(t) \cdot e^{-\delta_i(t-T)} dt$$
$$\text{and} \quad e^a_{L,1-\tau_K} = \int_i \delta_i g_i(T) \int_0^\infty e^a_{L,1-\tau_K}(t) \cdot e^{-\delta_i(t-T)} dt$$

Relative to the case with homogeneous discount rates, the social welfare weights $g_i(T)$ now depend on the time of the reform T, and on each person's discount rate δ_i. Agents who are more impatient (larger δ_i) get discounted more heavily in the anticipated social welfare objective. This effect is starker the longer in advance the reform is announced (the larger T). In the limit, only the most patient agents with the lowest δ_i will be counted in the social welfare objective.

Finite and infinite anticipation elasticities

Anticipation elasticities are infinite with wealth in the utility and certainty, but finite with uncertainty (with or without wealth in the utility).

First, the anticipation elasticity to a reform $d\tau_K$ for $t \geq T$ is infinite when there is full certainty, even with wealth in the utility. The proof is as in Piketty and Saez (2013b) for the Chamley-Judd model (without wealth in the utility).

With full certainty, the first-order condition of the agent with respect to capital always holds:

$$u_{ci,t} = (1+\bar{r})/(1+\delta_i) u_{ci,t+1} + 1/(1+\delta_i) u_{ki,t+1}$$

Suppose we start from a situation in a well-defined steady state: $(\delta_i - \bar{r}) u_{ci} = u_{ki}$ where we have perfect consumption smoothing.

The intertemporal budget constraint is:

$$\sum_{t \geq 0} \left(\frac{1}{1+r} \right)^t c_{ti} + \lim_{t \to \infty} k_{ti} = \sum_{t \geq 0} \left(\frac{1}{1+r} \right)^t z_{ti} + k_{0i}$$

Consumption smoothing implies:

$$u_{ci}(\bar{r}k_i + z_i, k_i) = \lambda$$

for the multiplier λ on the budget constraint. Hence, $k_i^\infty = \lim_{t\to\infty} k_{ti} > 0$. Given that there is perfect consumption smoothing, using the budget constraint to solve for consumption yields:

$$c = \left(1 - \frac{1}{1+r}\right)\left(\sum_{t\geq 0}\left(\frac{1}{1+r}\right)^t z_{ti} + k_{0i} - k_i^\infty\right) \tag{A2}$$

Consider what happens if the capital tax rate increases by $d\tau_K > 0$ for $t \geq T$. The present discounted value of all resources, denoted by Y_i for agent i is:

$$Y_i = k_{i0} + \sum_{t=1}^{T}\left(\frac{1}{1+r}\right)^t z_{ti} + \sum_{t\geq T}\left(\frac{1}{1+\bar{r}}\right)^t z_{ti}$$

The change in resources evaluated at $\tau_K = 0$ is:

$$dY_i = \left(\frac{1}{(1+r)}\right)^T \sum_{t\geq T} t\left(\frac{1}{(1+r)}\right)^{t-T+1} z_{ti}d\tau_K \propto \left(\frac{1}{(1+r)}\right)^T d\tau_K$$

Hence, consumption pre-reform will shift down by a factor proportional to $\left(\frac{1}{(1+r)}\right)^T d\tau_K$. From the aggregated budget constraint we have that:

$$k_t^m = (1+r)^t k_0^m - c_0^m(1 + (1+r) + (1+r)^2 + ... + (1+r)^{t-1}) + (z_{t-1}^m + .. + (1+r)^{t-1}z_0^m)$$

Therefore, the change in the aggregate capital stock is:

$$dk_t^m = -dc_0^m\left(\frac{(1+r)^{t-1}-1}{r}\right)$$

Recall that the change in consumption (from (A2)) is proportional to $\left(\frac{1}{(1+r)}\right)^T d\tau_K$. Hence:

$$dk_t^m \propto -\left(\frac{1}{(1+r)}\right)^T\left(\frac{(1+r)^{t-1}-1}{r}\right)d\tau_K = -(1+r)^{-T}\left(\frac{(1+r)^{t-1}-1}{r}\right)d\tau_K$$

Hence:

$$e_{Kt} \propto k_t^m (1+r)^{-T} \left(\frac{(1+r)^{t-1} - 1}{r} \right) d\tau_K$$

The anticipation elasticity e_K^{ante} is defined as:

$$e_K^{ante} = \frac{\delta}{1+\delta} \sum_{t<T} \left(\frac{1}{1+\delta} \right)^{t-T} e_{Kt} \propto \frac{\delta}{1+\delta} \sum_{t<T} \left(\frac{1}{1+\delta} \right)^{t-T} k_t^m (1+r)^{-T} \left(\frac{(1+r)^{t-1} - 1}{r} \right) d\tau_K$$

Since we have $\delta > r$, $\lim_{T\to\infty} \left(\frac{1+\delta}{1+r}\right)^T = \infty$, which makes the sum above (to which the anticipation elasticity is proportional) converge to infinity when T goes to infinity.

If the economy is away from the steady state:

If the economy is not in steady state, the spirit of the formula still holds, but it is no longer possible to treat marginal utilities and aggregate variables as constant. In that case, not just the elasticities, but also the weighting factors multiplying them and the distributional characteristic take into account the full transition path. For conciseness, let $u_{ci}(t) \equiv u_{ic}(c_i(t), k_i(t), z_i(t))$. The change in social welfare is (with each term of the formula marked in underbraces):

$$dSWF \propto 1 - \underbrace{\int_i \omega_i \delta_i \int_T^\infty \frac{u_{ic}(t)}{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) r k^m(t) \cdot e^{-\delta_i t} dt} r k_i(t) \cdot e^{-\delta_i t} dt}_{\bar{g}_K}$$

$$- \frac{\tau_L}{1-\tau_K} \underbrace{\int_i \omega_i \delta_i \int_0^\infty u_{ic}(t) e^a_{L,1-\tau_K}(t) \frac{z^m(t)}{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) z^m(t) \cdot e^{-\delta_i t} dt} e^{-\delta_i t} dt}_{e^a_{L,(1-\tau_K)}} \cdot \underbrace{\frac{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) z^m(t) \cdot e^{-\delta_i t} dt}{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) r k^m(t) \cdot e^{-\delta_i t} dt}}_{z^m/(rk^m)}$$

$$- \frac{\tau_K}{1-\tau_K} \underbrace{\int_i \omega_i \delta_i \int_0^\infty \frac{u_{ic}(t)}{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) r k^m(t) \cdot e^{-\delta_i t} dt} r k^m(t) e^a_K(t) \cdot e^{-\delta_i t} dt}_{e^a_K}$$

For an unanticipated reform starting from an arbitrary tax system and away from the steady state, set $T = 0$ in the above expression for $dSWF$ and replace the anticipated elasticities by their unanticipated counterparts $e^u_{L,(1-\tau_K)}$ and e^u_K.

A.2.3 Utility Based Steady State Approach

Proof of Proposition 10

With the utility based steady state approach, the government maximizes social welfare

$$SWF = \int_i \omega_i V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0})di$$

$$\text{where} \quad V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}) = \delta \cdot \int_{t=0}^{\infty} u_i(c_i(t), k_i(t), z_i(t))e^{-\delta t}dt,$$

subject to the steady state budget constraint, so that the lump-sum grant every period is equal to its steady state value

$$G = r\tau_K k^m + \tau_L z^m$$

Thus, individuals' consumption each period satisfies: $\frac{dk_i(t)}{dt} = rk_i(t)(1-\tau_K) + (1-\tau_L)z_i(t) - c_i(t) + G$.

Consider a small reform $d\tau_K$ that takes place at time 0. It does not matter whether the reform is anticipated or not, as the only relevant behavioral responses for revenues are the steady state responses which affect the steady state lump sum grant G (and, as always, by the envelope theorem, behavioral responses dk_i can be ignored when computing dV_i). Using the envelope theorem, the effect on the welfare of individual i is:

$$dV_i = d\tau_K \cdot \delta_i \Bigg[\int_0^{\infty} u_{ic}(c_i(t), k_i(t), z_i(t)) rk^m \cdot e^{-\delta_i t} - \int_0^{\infty} u_{ic}(c_i(t), k_i(t), z_i(t)) rk_i(t) \cdot e^{-\delta_i t}$$
$$- \frac{\tau_K}{1-\tau_K} \int_0^{\infty} u_{ic}(c_i(t), k_i(t), z_i(t)) rk^m e_K \cdot e^{-\delta_i t} dt - \frac{\tau_L}{1-\tau_K} \int_0^{\infty} u_{ic}(c_i(t), k_i(t), z_i(t)) e_{L,1-\tau_K} z^m e^{-\delta_i t} dt \Bigg].$$

In the steady state, $c_i(t)$, $z_i(t)$, and $k_i(t)$ are time-constant, so that the change in individual i's utility is:

$$dV_i = d\tau_K \cdot rk^m \left[u_{ic}(c_i, k_i . z_i) - u_{ic}(c_i, k_i, z_i)\frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} u_{ic}(c_i, k_i, z_i) e_K - \frac{\tau_L}{1-\tau_K} u_{ic}(c_i, k_i, z_i) \frac{z^m}{rk^m} e_{L,1-\tau_K} \right]$$

and the change in social welfare is:

$$dSWF = \int_i \omega_i dV_i \propto 1 - \int_i \frac{g_i k_i}{k^m} - \frac{\tau_K}{1-\tau_K} e_K - \frac{\tau_L}{1-\tau_K} \frac{z^m}{rk^m} e_{L,1-\tau_K}$$

which yields the formula in the text.

A.3 Comparison to Other Models

A.3.1 Judd (1985) Model

In the Judd (1985) model, individual utility is:

$$V_i(\{c_i(t), z_i(t), k_i(t)\}_{t\geq 0}) = \int_0^\infty u_i(c_i(t), k_i(t), z_i(t)) e^{-\int_0^t \delta_i(c_i(s))ds} dt.$$

The effect on V_i from a small change in the capital tax $d\tau_K$ is now:

$$\begin{aligned} dV_i = d\tau_K \Bigg[\int_{t=0}^\infty \Bigg(u_{ic}(c_i(t), k_i(t), z_i(t)) e^{-\int_0^t \delta_i(c_i(s))ds} + \delta_i'(c_i(t)) \int_t^\infty u_i(s) e^{-\int_0^s \delta_i(c_i(m))dm} ds \Bigg) \\ \times \left(rk^m(t) - rk_i(t) - \frac{\tau_K}{1-\tau_K} rk^m(t) e_K(t) \right) dt \Bigg]. \end{aligned}$$

In the steady state, we can hence write dV_i as:

$$\begin{aligned} & d\tau_K r \left[\int_0^\infty \left(u_{ic} e^{-\delta_i(c_i)t} + \delta_i'(c_i) u_i e^{-\delta_i(c_i)t} \int_t^\infty e^{-\delta_i(c_i)(s-t)} ds \right) \left(k^m(t) - k_i(t) - \frac{\tau_K}{1-\tau_K} k^m(t) e_K(t) \right) dt \right] \\ & = d\tau_K r \Bigg[\left(u_{ic} \int_0^\infty e^{-\delta_i(c_i)t} dt + \delta_i'(c_i) u_i \int_0^\infty e^{-\delta_i(c_i)t} \frac{1}{\delta_i(c_i)} \right) \times [k^m(t) - k_i(t)] \\ & - \int_0^\infty \left(u_{ic} e^{-\delta_i(c_i)t} + \delta_i'(c_i) u_i e^{-\delta_i(c_i)t} \int_0^\infty e^{-\delta_i(c_i)t} ds \right) \frac{\tau_K}{1-\tau_K} k^m(t) e_K(t) \Bigg] \\ & = d\tau_K r k^m \frac{1}{\delta_i(c_i)} \left(u_{ic} + \frac{\delta_i'(c_i)}{\delta_i(c_i)} u_i \right) \left[1 - \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} \delta_i(c_i) \int_0^\infty e^{-\delta_i(c_i)t} e_K(t) \right]. \end{aligned}$$

We can hence see that the formulas from our model apply but with g_i and e_K as redefined in the text.

A.3.2 Aiyagari (1995) Model

Note that all proofs below would be exactly the same as the proofs for wealth-in-the-utility if we reformulated it in discrete time, replacing the standard utility without wealth in the utility, $u_{ti}(c_{ti})$, by $u_{ti}(c_{ti}, k_{ti})$. This is done by letting u'_{ti} denote $\frac{\partial u_{ti}(c_{ti}, k_{ti})}{\partial c_{ti}}$ instead of $\frac{\partial u_{ti}(c_{ti})}{\partial c_{ti}}$.

We apply the envelope theorem, which states that the changes in the capital tax rate $d\tau_K$ only has a direct impact on utility through the direct reduction in consumption that it causes.

Using this, and taking the derivative of the social welfare SWF with respect to $d\tau_K$ yields:

$$
\begin{aligned}
dSWF &= \sum_{t<T} \left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti} \cdot (\tau_K r dk_t^m) + \sum_{t\geq T} \left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti} \cdot (r d\tau_K (k_t^m - k_{ti}) + \tau_K r dk_t^m) \\
&= -d\tau_K \Bigg(\frac{\tau_K}{1-\tau_K} \left[\sum_{t<T} \left(\frac{1}{1+\delta}\right)^t r k_t^m e_{Kt} \int_i \omega_i u'_{ti} + \sum_{t\geq T} \left(\frac{1}{1+\delta}\right)^t r k_t^m e_{Kt} \int_i \omega_i u'_{ti} \right] \\
&+ \sum_{t\geq T} \left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti} \cdot r(k_t^m - k_{ti}) \Bigg) \\
&= -d\tau_K \left(\frac{\tau_K}{1-\tau_K} \left[\sum_{t\geq 0} \left(\frac{1}{1+\delta}\right)^t r k_t^m e_{Kt} \int_i \omega_i u'_{ti} \right] - \sum_{t\geq T} \left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti} \cdot r(k_t^m - k_{ti}) \right).
\end{aligned}
$$

If variables have already converged to their ergodic paths when the anticipation responses start: then all terms in e_{Kt} are zero before the steady state has been reached and hence, we can divide through by $\int_i \omega_i u'_{ti} k_t^m = \int_i g_i k_t^m$ which is constant across t. Thus:

$$
dSWF \propto \frac{\tau_K}{(1-\tau_K)} \left(\frac{\delta}{1+\delta} \sum_{t<T} \left(\frac{1}{1+\delta}\right)^{t-T} e_{Kt} + \frac{\delta}{1+\delta} \sum_{t\geq T} \left(\frac{1}{1+\delta}\right)^{t-T} e_{Kt} \right) - 1 + \frac{\int_i g_i k_{ti}}{\int_i g_i k_t^m}.
$$

Let the distributional factor $\bar{g}_K = \frac{\int_i g_i k_{ti}}{\int_i g_i k_t^m}$. The optimal capital tax in the Aiyagari (1995) model is given by:

$$
\tau_K = \frac{1-\bar{g}_K}{1-\bar{g}_K + e_K}.
$$

with $e_K = \frac{\delta}{1+\delta} \sum_{t>0} \left(\frac{1}{1+\delta}\right)^{t-T} e_{Kt}$. For an unanticipated reform, the formula applies with $T=0$ when the economy is already in the steady state as of time 0.

If variables have not converged to their ergodic paths when the anticipation responses start: we have to take into account the transition of the marginal utilities and the capital stock across time.

$$
dSWF = -d\tau_K \left(\frac{\tau_K}{(1-\tau_K)} \left[\sum_{t<T} \left(\frac{1}{1+\delta}\right)^t r k_t^m e_{Kt} \int_i \omega_i u'_{ti} \right] - \sum_{t\geq T} \left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti} \cdot r(k_t^m - k_{ti}) \right).
$$

Dividing by $\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti}\cdot k_t^m$ yields:

$$dSWF \propto \frac{\tau_K}{(1-\tau_K)}\left[\sum_{t<T}\left(\frac{1}{1+\delta}\right)^t k_t^m e_{Kt}\frac{\int_i \omega_i u'_{ti}}{\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti}\cdot k_t^m}\right]$$
$$-1+\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \frac{\int_i \omega_i u'_{ti}\cdot k_{ti}}{\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti}\cdot k_t^m}.$$

Now we have to redefine the average welfare weight as:

$$\bar{g}_K \equiv \sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \frac{\int_i u'_{ti}\cdot k_{ti}}{\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i u'_{ti}\cdot k_t^m},$$

and the total elasticity as:

$$e_K = \sum_{t\geq 0}\left(\frac{1}{1+\delta}\right)^t k_t^m e_{Kt}\frac{\int_i u'_{ti}}{\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i u'_{ti}\cdot k_t^m}.$$

With these redefined variables, the same formula holds.

A.4 Optimal Nonlinear Taxes in the Generalized Model

Let $e_K^{top}(t)$ be the average elasticity of total capital income of those individuals with capital income above threshold rk^{top}. It is measured at time t following a small reform of the top bracket tax rate $d\tau_K$ taking place at time 0. The elasticity is weighted by capital income. Let $e_{L,1-\tau_K}(t)$ be the elasticity of labor income of those individuals with capital income above threshold rk^{top}.

Proposition A1. ***Optimal top capital tax rate in the steady state.***

Suppose there is a linear tax on labor income τ_L. The optimal top capital tax rate above capital income level rk^{top} takes the form:

$$\tau_K^{top} = \frac{1-\bar{g}_K^{top}-\tau_L\cdot\frac{z^m}{r(k^{m,top}-k^{top})}\cdot\bar{e}_{L,(1-\tau_K)}}{1-\bar{g}_K^{top}+a_K^{top}\cdot\bar{e}_K^{top}},$$

with $\bar{e}_K^{top} \equiv \int_i g_i\delta_i\int_{t=0}^{\infty} e_K^{top}(t)\cdot e^{-\delta_i t}dt$. $\bar{g}_K^{top} = \frac{\int_{i:k_i\geq k^{top}} g_i\cdot(k_i-k^{top})}{\int_{i:k_i\geq k^{top}}(k_i-k^{top})}$ is the average capital income weighted welfare weight in the top capital tax bracket, and $a_K^{top} = \frac{k^{m,top}}{k^{m,top}-k^{top}}$ is the Pareto parameter of the capital income distribution. $\bar{e}_{L,(1-\tau_K)} \equiv \int_i g_i\delta_i\int_{t=0}^{\infty} e_{L,(1-\tau_K)}(t)\cdot e^{-\delta_i t}dt$.

Proof of Proposition A1: We consider the top tax rate τ_K on capital above threshold k^{top}. As r is uniform, this is equivalent to a top tax rate applying above capital income threshold rk^{top}. Let N denote the fraction of individuals above k^{top}. We again use the notation $k^{m,top}$ to denote the average wealth above the top threshold, i.e.:

$$k^{m,top} = \frac{\int_{i:k_i(t)\geq k^{top}} rk_i}{N},$$

Suppose we change the top tax rate on capital by $d\tau_K$. As defined in the text, let $e_K^{top}(t)$ be the elasticity of capital holding of top capital earners (the wealth elasticity of total wealth to the tax rate of those with capital income above rk^{top}). For all individuals above the cutoff, the change in utility is:

$$dV_i = d\tau_K \delta_i \Bigg[\int_0^\infty u_{ic}(c_i(t), k_i(t)) N r (k^{m,top}(t) - k^{top}) e^{-\delta_i t} - \int_0^\infty u_{ic}(c_i(t), k_i(t)) r (k_i(t) - k^{top}) e^{-\delta_i t} - \frac{\tau_K}{1-\tau_K} \int_0^\infty u_{ic}(c_i(t), k_i(t)) N r k^{m,top}(t) e_K^{top}(t) \cdot e^{-\delta_i t} dt \Bigg].$$

Starting from the steady state, capital levels are constant so that:

$$dV_i = u_{ic} r (k^{m,top} - k^{top}) N d\tau_K \left[1 - \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N} - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_0^\infty \delta_i e_K^{top}(t) \cdot e^{-\delta_i t} dt \right],$$

where $a_K^{top} = \frac{k^{m,top}}{(k^{m,top} - k^{top})}$.

For individuals below the cutoff, the change in utility is:

$$dV_i = u_{ic} r (k^{m,top} - k^{top}) N d\tau_K \left[1 - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_0^\infty \delta_i e_K^{top}(t) \cdot e^{-\delta_i t} dt \right].$$

The change in social welfare is such that:

$$dSWF \propto 1 - \int_{i:k_i \geq k^{top}} g_i \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N} - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_i g_i \delta_i \int_0^\infty e_K^{top}(t) \cdot e^{-\delta_i t} dt.$$

Let

$$\bar{g}_K^{top} \equiv \int_{i:k_i \geq k^{top}} g_i \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N} \quad \text{and} \quad e_K^{top} \equiv \int_i g_i \delta_i \int_0^\infty e_K^{top}(t) \cdot e^{-\delta_i t} dt.$$

Then, we obtain the optimal tax rate τ_K such that $dSWF = 0$:

$$\tau_K = \frac{1 - \bar{g}_K^{top}}{1 - \bar{g}_K^{top} + a_K^{top} e_K^{top}}.$$

With endogenous labor, let

$$e_{L,(1-\tau_K)}(t) = \frac{dz^m(t)}{d(1-\tau_K)} \frac{(1-\tau_K)}{z^m(t)} = \frac{dz^m(t)}{d\bar{r}} \frac{\bar{r}}{Nz^m(t)}.$$

be the elasticity of aggregate (average) labor income z^m with respect to the top capital tax rate, normalized by N, in the two bracket tax system.

For all individuals with capital income above the cutoff:

$$\begin{aligned} dV_i = d\tau_K \cdot \delta_i \Bigg[\int_0^\infty & u_{ic}(c_i(t), k_i(t), z_i(t)) N r (k^{m,top}(t) - k^{top}) \cdot e^{-\delta_i t} \\ & - \frac{\tau_L}{1-\tau_K} \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) z^m(t) N e_{L,(1-\tau_K)}(t) \cdot e^{-\delta_i t} \\ & - \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) r (k_i(t) - k^{top}) \cdot e^{-\delta_i t} \\ & - \frac{\tau_K}{1-\tau_K} \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) N r k^{m,top}(t) e_K^{top}(t) \cdot e^{-\delta_i t} dt \Bigg]. \end{aligned}$$

Starting from the steady state, capital and labor income are constant over time:

$$\begin{aligned} dV_i = u_{ic} N r (k^{m,top} - k^{top}) d\tau_K \cdot \Bigg[1 - \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N} \\ - \frac{\tau_L}{1-\tau_K} \frac{z^m}{r(k^{m,top} - k^{top})} \int_0^\infty \delta_i e_{L,(1-\tau_K)}(t) \cdot e^{-\delta_i t} dt - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_0^\infty \delta_i e_K^{top}(t) \cdot e^{-\delta_i t} dt \Bigg]. \end{aligned}$$

The change in social welfare is:

$$\begin{aligned} dSWF = \int_i \omega_i dV_i \propto 1 - \int_{i:rk_i \geq rk^{top}} g_i \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N} \\ - \frac{\tau_L}{1-\tau_K} \frac{z^m}{r(k^{m,top} - k^{top})} \int_i g_i \int_0^\infty \delta_i e_{L,(1-\tau_K)}(t) \cdot e^{-\delta_i t} dt - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_i g_i \int_0^\infty \delta_i e_K^{top}(t) \cdot e^{-\delta_i t} dt. \end{aligned}$$

Define e_K^{top}, $e_{L,(1-\tau_K)}$, and $\bar{g}_K^{top}$ as in the text. The optimal formula in the text is then obtained by rearranging the previous condition.

It is straightforward to generalize this to the case of an anticipated reform by discounting the elasticities using $e^{-\delta(t-T)}$ and using the anticipated elasticities rather than the unanticipated ones.

A.5 Optimal Taxation with Horizontal Equity Concerns.

In this section, we formally consider optimal capital and labor taxation under horizontal equity concerns.

As derived in Section 3.2.1, the optimal revenue-maximizing rates are: $\tau_L^R = \frac{1}{1+e_L}$ and $\tau_K^R = \frac{1}{1+e_K}$. Without loss of generality, we suppose that capital is more elastic so that $\tau_K^R < \tau_L^R$. The optimal linear comprehensive tax on income is, as derived in (16):

$$\tau_Y = \frac{1-\bar{g}_Y}{1-\bar{g}_Y+e_Y} \quad \text{with} \quad \bar{g}_Y = \frac{\int_i g_i \cdot y_i}{\int_i y_i}$$

Suppose that the distribution of capital and labor income is dense enough, so that at every total income level $y = rk + z$, there are agents with $y = rk$ (capital income only) and $y = z$ (labor income only).

Generalized social welfare weights that capture horizontal equity concerns are such that:

(i) If $\tau_L = \tau_K$, then g_i are standard, for instance $g_i = u_{ci}$ for all agents. Any reform that changes taxes should put zero weight on those who after the reform are such that $\tau_L z_i + \tau_K r k_i < \max_j\{\tau_L z_j + \tau_K r k_j | z_j + r k_j = z_i + r k_i\}$, i.e., on those who pay less taxes at a given total income $y = rk_i + z_i$, or, equivalently, have the highest disposable income and consumption at any income. This means that if labor taxes are increased, $g_i = 0$ for those with any positive capital income at each total income level. Conversely, increasing capital taxes will yield $g_i = 0$ for those individuals with some labor income at each total income level.

(ii) If $\tau_L > \tau_K$, then all the social welfare weights are concentrated on those with $\tau_L z_i + \tau_K r k_i > \max_j\{\tau_L z_j + \tau_K r k_j | z_j + r k_j = z_i + r k_i\}$, i.e., on those agents with only labor income. Conversely, if $\tau_L < \tau_K$, all the social welfare weights are on agents with only capital income.

Suppose that, starting from a situation with $\tau_L = \tau_K$ we introduce a small tax break on capital income, $d\tau_K < 0$. Capital income earners now get an unfair advantage and all the weight is concentrated on those with no capital income (equivalently, everyone with $k_i > 0$ receives a weight $g_i = 0$). As a result, a small tax break on capital can only be optimal if it raises tax revenue and, hence, allows to lower the tax rate on labor income as well. This can only occur if $\tau_Y > \tau_K^R$, i.e., the optimal comprehensive tax rate is above the revenue-maximizing rate on capital income.

Proposition A2. ***Optimal labor and capital taxation with horizontal equity concerns.***

(i) If $\tau_Y \leq \tau_K^R$, taxing labor and capital income at the same comprehensive rate $\tau_L = \tau_K = \tau_Y$ is the unique optimum.

(ii) If $\tau_Y > \tau_K^R$, a differential tax system with the capital tax rate set to the revenue maximizing rate $\tau_K = \tau_K^R < \tau_L$ (with both τ_K and τ_L smaller than τ_Y) is the unique optimum.

Proof. Let us consider the two cases in turn.

(i) If $\tau_Y \leq \tau_K^R$.

To see why $\tau_L = \tau_K = \tau^*$ is an equilibrium, suppose that we tried to lower the tax rate on capital income. Then, all the weight will concentrate on people with only labor income, which will then in turn make it optimal to increase the tax on capital again.

This equilibrium is unique. There is no other equilibrium with equal taxes on capital and labor that can raise more revenue with a lower tax rate, by definition of τ_Y as the optimal rate on comprehensive income. There is also no equilibrium with non-equal tax rates on capital and labor. Suppose that we tried to set (without loss of generality) $\tau_K < \tau_L$. Then to raise enough revenue we would require that $\tau_K < \tau_Y < \tau_L$. Since capital owners are now advantaged, all the social welfare weight concentrates on people with only labor income. Since then a fortiori $\tau_K < \tau_K^R$, increasing τ_K would mean that more revenue would be raised, which would allow us to lower τ_L, which is good since all weight is on people with only labor income.

(ii) If $\tau_Y > \tau_K^R$.

In this case, the equilibrium has $\tau_K = \tau_K^R < \tau_Y$ and $\tau_Y > \tau_L > \tau_K^R$. Clearly this is an equilibrium since we cannot decrease τ_L without losing revenue and we cannot raise more revenue through τ_K (since it is already set at the revenue-maximizing rate for the capital tax base). In addition, we cannot decrease τ_K further without increasing τ_L, which is not desirable since it would benefit people capital income earners, who already receive a weight of zero.

This equilibrium is also unique. If we set $\tau_L = \tau_K$ equal, we should set them equal to τ_Y which is the optimal tax rate on comprehensive income. But then, since τ_K is now above its revenue maximizing rate, we could lower both τ_K and τ_L without losing revenues, so this would not be an equilibrium. On the other hand, as long as we set $\tau_K < \tau_L$, capital income earners get zero weight and the only possibility is to go all the way to $\tau_K = \tau_K^R$ since only people with only labor income have a non-zero weight.

$\square$

As a result, horizontal equity concerns will be a force pushing towards the comprehensive income tax system derived in Section 3.2.1. In the text, we provided an efficiency argument in favor of a tax on comprehensive income (based on income shifting opportunities) while the argument here is based on equity considerations. With horizontal equity preferences, deviations from a comprehensive income tax system can only be justified if they raise more revenue and generate a Pareto-improvement, which drastically reduces the scope for them. In Saez and Stantcheva (2016) we argue that this is akin to a generalized Rawlsian principle whereby discrimination against some groups (e.g., capital owners versus labor providers) is only permissible if it makes the group discriminated against better off, i.e., if it generates a Pareto improvement.

A.5.1 Horizontal Equity with Nonlinear Taxation

The same reasoning as for linear taxation with horizontal equity also applies to nonlinear taxes. Starting from a comprehensive tax system $T_Y(z + rk)$ as derived in Section 3.2.1, lowering the tax rate on capital income, conditional on a given total income level, will generate a horizontal inequity and concentrate all social weight on those with no capital income conditional on that total income level. Such a preferential tax break for capital income earners will only be acceptable if it generates more revenue and allows to lower the tax rate on labor income as well. We show this below.

Formally, suppose that we start from the optimal tax on comprehensive income, $T_Y(rk + z)$, as derived in Section 3.2.1 which does not discriminate between capital and labor income conditional on total income. We say that a tax system unambigously favors capital (respectively, labor) at income level $y = rk + z$, if for any (rk, z) such that $y = rk + z$, and any $\varepsilon \in]0, z]$, $T_Y(rk, z) > T(rk + \varepsilon, z - \varepsilon)$ (having more capital income, conditional on a given total income leads to lower taxes). (Note that it may be the case that a tax system favors capital only at some y levels or only at some rk, z ranges..)

Denote a change in the tax by $\delta T(rk, z)$.

A deviation $\delta T(rk, z)$ is said to introduce horizontal inequity, if, starting from a comprehensive tax system $T_Y(rk + z)$, the resulting tax system $T_Y(z + rk) + \delta T(rk.z)$ cannot be expressed as $\tilde{T}_Y(rk + z)$ for some function $\tilde{T}_Y$.

With nonlinear taxes, we can again define the generalized social welfare weights as follows.

i) If there is a comprehensive tax $T_Y(z + rk)$, then everybody has standard weights, such as, for instance, $g_i = u_{ci}$. For any deviation $\delta T(rk, z)$ that introduces horizontal inequity, the weights concentrate on the agents who pay the highest tax at a given total income level, i.e., on those with $T_Y(z_i + rk_i) + \delta T(rk_i, z_i) = \max_j \{T_Y(z_j + rk_j) + \delta T(rk_j, z_j)|z_j + rk_j = rk_i + z_i\}$ (which is equivalent to putting all the weight on the agent(s) with lowest disposable income at any total income level).

Hence, the weights also need to depend on $\delta T(z, rk)$, the direction of the tax reform.

ii) If the tax is such that $T(rk, z)$ cannot be expressed as $\tilde{T}_Y(rk + z)$ for some function $\tilde{T}_Y$, then the weights concentrate on those with

$T(z_i, rk_i) = \max_j \{T(z_j, rk_j) | z_j + rk_j = rk_i + z_i\}$, i.e., on the agents which pay the highest tax (equivalently, have the lowest disposable income) conditional on total income.

Equilibria:

Suppose that, at the comprehensive tax rate, no small reform $\delta T(rk.z)$ that introduces horizontal equity and favors capital (according to our definitions above) can increase total tax revenues, i.e., for all $\delta T(rk, z)$ that favor capital and introduce horizontal inequity, the alternative tax system $\tilde{T}(rk, z) = T(rk + z) + \delta T(rk, z)$ is such that:

$$\int_i T_Y(rk_i(T) + z_i(T))di > \int_i \tilde{T}_Y(rk_i(\tilde{T}) + z_i(\tilde{T}))di$$

where naturally, the choices $z_i(T)$ and $k_i(T)$ depend on the tax system T. Then the unique equilibrium has the comprehensive tax system in place, as derived in 3.2.1. No horizontal inequity can be an equilibrium unless it introduces a Pareto improvement.

Suppose on the other hand that if the revenue maximizing tax rate on capital, $T_K^R(rk)$ were implemented, and a labor income tax $T_L(z)$ was used to complement it, more revenue could be raised than with the tax on comprehensive income $T_Y(rk, z)$ and the tax burden on all agents would be lower than under the comprehensive income tax. Then, the optimum is to set differential taxes on capital and labor income, with the capital tax at its optimal revenue-maximizing schedule. Horizontal inequity is an equilibrium because it generates a Pareto improvement.

A.6 Progressive Consumption Taxes

The progressive consumption tax is defined on an exclusive basis as $t_C(.)$ such that

$$\dot{k} = \bar{r}k + z - [c + t_c(c)]$$

Equivalently, we can again define the inclusive consumption tax $T_C(y)$ on pre-tax resources y devoted to consumption such that $c + t_c(c) = y$ is equivalent to $c = y - T_C(y)$, i.e., $y \to y - T_C(y)$ is the inverse function of $c \to c + t_c(c)$ and hence $1 + t'_C = 1/(1 - T'_C)$.

The case of a progressive consumption tax is most easily explained with inelastic labor income (possibly heterogenous across individuals). Real wealth k^r in the presence of the progressive consumption tax is:

$$k^r(k) = k - \frac{T_C(\bar{r}k + z) - T_C(z)}{\bar{r}}$$

Recall that real wealth is defined as nominal wealth adjusted for the price of consumption.

There are to see why the above is the right expression. First, wealth k provides an income stream $\bar{r}k$ which translates into extra permanent consumption equal to the income minus the tax paid on the extra consumption $\bar{r}k - [T_C(\bar{r}k+z) - T_C(z)]$ which can be capitalized into wealth k^r by dividing by $\bar{r}$. If labor income is heterogeneous across agents, then $k^r(k, z)$ should also be indexed by z. Another way to see this is to ask what the capital k^r would be that would yield the same disposable income as the nominal capital under the consumption tax. Disposable income in terms of real capital k^r is $\bar{r}k^r - T_C(z)$. Disposable income expressed in terms of nominal capital is: $\bar{r}k - T_C(\bar{r}k + z)$. These two must be equal, which yields the expression for k^r above. k^r has three natural properties: with no consumption tax, real and nominal wealth are equal, $dk^r/dk = 1 - T_C'$, i.e., and extra dollar of nominal wealth is worth $1 - T_C'$ in real terms, and $k^r(0) = 0$.

In that case, we have in steady-state

$$c = \bar{r}k + z - T_C(\bar{r}k + z) = \bar{r}k^r + z - T_C(z)$$

and the first order condition for utility maximization is $a_i'(k^r) = \delta - \bar{r}$. Hence, real capital is chosen to satisfy the same condition as nominal capital when there is no consumption tax. Put differently, any consumption tax will be undone by agents in terms of their savings and will have no effect on the real value of their wealth held (and, hence, by definition of the real wealth, on their purchasing power). Hence, the consumption tax is equivalent to a tax on labor income only.

The equivalence is not exact with elastic labor supply, as in that case, the marginal consumption tax depends on the labor choice and the first-order condition for labor income is $h_i'(z) = 1 - T_C'(\bar{r}k + z) + a_i'(k^r)[T_C'(\bar{r}k + z) - T_C'(z)]/\bar{r}$.