NBER WORKING PAPER SERIES

POSITIVE LONG RUN CAPITAL TAXATION: CHAMLEY-JUDD REVISITED

Ludwig Straub
Iván Werning

Working Paper 20441
http://www.nber.org/papers/w20441

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2014

This paper benefited from detailed comments by Fernando Alvarez, Peter Diamond and Stefanie Stantcheva and was very fortunate to count with research assistance from Greg Howard, Lucas Manuelli and Andrés Sarto. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Positive Long Run Capital Taxation: Chamley-Judd Revisited
Ludwig Straub and Iván Werning
NBER Working Paper No. 20441
August 2014
JEL No. H2,H63

ABSTRACT

According to the Chamley-Judd result, capital should not be taxed in the long run. In this paper, we overturn this conclusion, showing that it does not follow from the very models used to derive them. For the model in Judd (1985), we prove that the long run tax on capital is positive and significant, whenever the intertemporal elasticity of substitution is below one. For higher elasticities, the tax converges to zero but may do so at a slow rate, after centuries of high capital taxation. The model in Chamley (1986) imposes an upper bound on capital taxation and we prove that the tax rate may end up at this bound indefinitely. When, instead, the bounds do not bind forever, the long run tax is indeed zero; however, when preferences are recursive but non-additive across time, the zero-capital-tax limit comes accompanied by zero private wealth (zero tax base) or by zero labor taxes (first best). Finally, we explain why the equivalence of a positive capital tax with ever rising consumption taxes does not provide a firm rationale against capital taxation.

Ludwig Straub
Department of Economics, MIT
77 Massachusetts Avenue
Cambridge, MA 02139
straub@mit.edu

Iván Werning
Department of Economics, E17-234
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
iwerning@mit.edu

Positive Long Run Capital Taxation: Chamley-Judd Revisited*

Ludwig Straub
MIT

Iván Werning
MIT

August 2014

According to the Chamley-Judd result, capital should not be taxed in the long run. In this paper, we overturn this conclusion, showing that it does not follow from the very models used to derive them. For the model in Judd (1985), we prove that the long run tax on capital is positive and significant, whenever the intertemporal elasticity of substitution is below one. For higher elasticities, the tax converges to zero but may do so at a slow rate, after centuries of high capital taxation. The model in Chamley (1986) imposes an upper bound on capital taxation and we prove that the tax rate may end up at this bound indefinitely. When, instead, the bounds do not bind forever, the long run tax is indeed zero; however, when preferences are recursive but non-additive across time, the zero-capital-tax limit comes accompanied by zero private wealth (zero tax base) or by zero labor taxes (first best). Finally, we explain why the equivalence of a positive capital tax with ever rising consumption taxes does not provide a firm rationale against capital taxation.

1 Introduction

One of the most startling results in optimal tax theory is the famous finding by Judd (1985) and Chamley (1986). Although working independently, in somewhat different settings and, thus, complementing each other, their conclusions were strikingly similar: capital should go untaxed in any steady state. This implication, dubbed the Chamley-Judd result, is commonly interpreted as applying in the long run, since convergence to a

*This paper benefited from detailed comments by Fernando Alvarez, Peter Diamond, Mike Golosov and Stefanie Stantcheva and was very fortunate to count with research assistance from Greg Howard, Lucas Manuelli and Andrés Sarto. All errors are our own.

steady state is quite naturally taken for granted.[1] The takeaway is that taxes on capital should be zero, at least eventually.

Economic reasoning sometimes holds its surprises. The Chamley-Judd result is not at all obvious and was not anticipated by economists' intuition, despite a large body of work at the time on the incidence of capital taxation and on optimal tax theory more generally. It represented a major watershed from a theoretical standpoint. One may even say that the result is downright puzzling, as witnessed by the fact that economists have since then taken turns putting forth differing intuitions to interpret it, none definitive nor universally accepted to date.

Theoretical wonder aside, a crucial issue is the result's applicability. Many have questioned the model's assumptions, especially that of infinitely-lived agents (e.g. Banks and Diamond, 2010). Still others have set up alternative models, searching for different conclusions. These efforts notwithstanding, opponents and proponents alike acknowledge Chamley-Judd as one of the most important benchmarks in the optimal tax literature.

Here we question the Chamley-Judd results directly, on their own ground and argue that, even within the logic of these models, a zero long-run tax result does not follow. For both the models in Judd (1985) and Chamley (1986), we provide results showing a positive long-run tax when the intertemporal elasticity of substitution is less than or equal to one. We conclude that these models do not actually provide a coherent argument against capital taxation, quite the contrary. We briefly discuss what went wrong with the original results, their interpretations and proofs—which are deceivingly straightforward, but, unfortunately, are just as deceiving as they are straightforward.

Before summarizing our results in greater detail, it is useful to briefly recall the setups in Judd (1985) and Chamley (1986). Start with the similarities. Both papers studied infinitely-lived agents. The models take as given an initial stock of capital and restrict tax instruments to proportional taxes on capital and labor. Lump-sum taxes are limited or excluded, as are capital levies or other forms of capital expropriation. The tax rate is constrained by an upper bound.[2]

Turning to differences, Chamley (1986) focused on a representative agent and assumed

[1] To quote from a few examples, Judd (2002): "[...] setting τ_k equal to zero in the long run [...] various results arguing for zero long-run taxation of capital; see Judd (1985, 1999) for formal statements and analyses." Atkeson et al. (1999): "By formally describing and extending Chamley's (1986) result [...] This approach has produced a substantive lesson for policymakers: In the long run, in a broad class of environments, the optimal tax on capital income is zero." Phelan and Stacchetti (2001): "A celebrated result of Chamley (1986) and Judd (1985) states that with full commitment, the optimal capital tax rate converges to zero in the steady state." Saez (2013): "The influential studies by Chamley (1986) and Judd (1985) show that, in the long-run, optimal linear capital income tax should be zero."

[2] Neither consumption taxes that mimic the effects of an initial wealth expropriation (Coleman II, 2000), nor dividend taxes with capital expenditure (investment) deductions (Abel, 2007) are allowed.

perfect financial markets, with unconstrained government debt. Judd (1985) emphasizes heterogeneity and redistribution in a two-class economy, with workers and capitalists. In addition, the model features financial market imperfections: workers do not save and the government balances its budget, i.e. debt is restricted to zero.[3] As discussed in Judd (1985), it is most remarkable that a zero long-run tax result obtains despite the restriction to budget balance.[4] Although extreme, imperfections of this kind may capture relevant aspects of reality, such as the limited participation in financial markets, the skewed distributions of wealth and the problems one that may accompany high levels of government debt or asset levels..[5]

We begin with the model in Judd (1985). We show that taxes may remain positive even in the long run. When the intertemporal elasticity of substitution (IES) is below one, taxes rise and converge towards a positive limit tax, instead of declining towards zero. This limit tax is significant and drives capital to its lowest feasible level. Indeed, as government spending falls, the lowest feasible capital stock approaches zero, and, the limit tax rate goes to infinity. The long run tax is not zero, far from it: it is large and potentially unbounded.

When the IES is above one, we verify numerically that the solution converges to the zero-tax steady state. However, we show that this convergence may be very slow, sometimes taking centuries for wealth taxes to drop below 1%. Indeed, in the neighborhood of a unitary IES, the speed of convergence is not bounded away from zero. Thus, even when the long-run tax on capital is zero, this property provides a misleading summary of the model's full implications.

We then turn to the representative agent Ramsey model studied by Chamley (1986). As is well appreciated, in this setting, upper bounds on the taxation of capital are imposed in order to avoid the trivial solution involving expropriatory levels of initial taxation of capital. We provide two sets of results. First, we show that if the bounds on capital taxation are not binding in the long run, then the tax is indeed zero. However, we show that for recursive nonadditive utility, this zero tax limit is necessarily accompanied by either zero private wealth converging to zero—in which case the tax base is zero—or the labor tax converging to zero—in which case the first best is achieved. This suggests that

[3] Chamley (1986) offered a few extensions to heterogeneous agents and Judd (1985) also considered another model where workers, just as capitalists, can also save in capital.

[4] Because of the presence of financial restrictions and imperfections, Judd (1985) model does not fit the standard Arrow-Debreu framework, nor the optimal tax theory developed around it such as Diamond and Mirrlees (1971).

[5] Without constraints on debt, capitalists may become highly indebted or not own the capital they manage. This idea that investment requires "skin in the game" is popular in the finance literature and macroeconomic models with financial frictions (see Brunnermeier et al., 2012; Gertler and Kiyotaki, 2010, for surveys).

zero taxes are only attained after high taxation that either obliterates private wealth or allows the government to proceed without any distortionary taxation. Needless to say, these are not the scenarios envisioned by proponents of zero capital taxation.

Second, we show that the upper bounds imposed on the tax rate may bind forever, when the intertemporal elasticity of substitution is below one, implying a positive long-run tax on capital. We prove that this is guaranteed whenever debt is high enough. Importantly, the debt level required is below the peak of the Laffer curve, so this result is not driven by budgetary necessity. Intuitively, elevated levels of debt ensure high labor taxes, making capital taxation relatively attractive, helping to ease a high tax burden.

In a setting very close to that of Chamley (1986), Judd (1999) presents an argument against positive capital taxation without requiring convergence to a steady state. However, we show that these arguments fail, because they involve assumptions on endogenous multipliers that are actually violated at the optimum. We also explain why the observation discussed in that paper that capital taxation amounts to ever increasing consumption taxes, does not provide a rationale against indefinite capital taxation.

To conclude, we present a hybrid model that combines heterogeneity and redistribution, as in Judd (1985), but allowing for government debt, as in Chamley (1986). Capital taxation is especially potent in this setting. When upper bounds are imposed, the optimum indefinite taxation at the bound. This suggests that positive long run taxation may be expected in a wide range of models that are descendants of Chamley (1986) and Judd (1985).

2 Capitalists and Workers: Judd (1985) Revisited

We start with the two-class economy laid out in Judd (1985) without government debt. There are two types of agents, workers and capitalists. Capitalists save and derive all their income from the returns to capital. Workers supply one unit of labor inelastically and live hand to mouth, consuming their entire wage income plus transfers. The government taxes the return to capital to pay for transfers targeted at workers.

Preferences. Capitalists have utility

$$\sum_{t=0}^{\infty} \beta^t U(C_t) \qquad \text{with} \qquad U(C) = \frac{C^{1-\sigma}}{1-\sigma}$$

for $\sigma > 0$ and $\sigma \neq 1$, and $U(C) = \log C$ for $\sigma = 1$. Here $1/\sigma$ denotes the (constant) intertemporal elasticity of substitution. Workers consumption paths are valued according to the utility

$$\sum_{t=0}^{\infty} \beta^t u(c_t)$$

for some concave, continuously differentiable function u. Both agents discount the future with a common discount factor $\beta < 1$. Workers have a constant labor endowment $n = 1$; capitalists do not work. Consumption by workers will be denoted by lowercase c, consumption by capitalists by uppercase C.

Technology. Output is obtained from capital and labor using a neoclassical constant returns production function $F(k_t, n_t)$ satisfying standard conditions.[6] Capital depreciates at rate $\delta > 0$. In equilibrium $n_t = 1$, so define $f(k) = F(k, 1)$. The government consumes a constant flow of goods $g \geq 0$. We normalize both populations to unity and abstract from technological progress and population growth. The resource constraint in period t is then

$$c_t + C_t + g + k_{t+1} \leq f(k_t) + (1 - \delta)k_t.$$

Markets and Taxes. Markets are perfectly competitive, with labor paid wage $F_L(k_t, n_t)$ and the before-tax return on capital

$$R_t^* = f'(k_t) + 1 - \delta.$$

The after-tax return equals R_t and can be parameterized as

$$R_t = (1 - \tau_t)(f'(k_t) - \delta) + 1,$$

where τ_t denotes a tax rate on net returns. This parameterization for R_t is somewhat arbitrary and we often take R_t as a direct policy variable and say that capital is taxed whenever $R_t < R_t^*$ and subsidized if $R_t > R_t^*$.

[6] Increasing in both arguments, strictly concave, continuously differentiable, with the INADA conditions $F_k(k, L) \to \infty$ as $k \to 0$ and $F_k(k, L) \to 0$ as $k \to \infty$.

Agent Behavior. Capitalists solve

$$\max_{\{C_t, a_{t+1}\}} \sum_{t=0}^{\infty} \beta^t U(C_t),$$

$$\text{s.t. } C_t + a_{t+1} = R_t a_t,$$

$$a_{t+1} \geq 0,$$

for some given initial wealth a_0. The associated Euler equation and transversality conditions,

$$U'(C_t) = \beta R_{t+1} U'(C_{t+1}),$$

$$\beta^t U'(C_t) a_{t+1} \to 0,$$

are necessary and sufficient for optimality.

Workers live hand to mouth, their consumption equals their disposable income

$$c_t = f(k_t) - f'(k_t)k_t + T_t,$$

which uses the fact that $F_n = F - F_k k$. Here T_t represent government transfers to workers.

Government Budget Constraint. As in Judd (1985), the government cannot issue bonds and runs a balanced budget. This implies that total wealth equals the capital stock $a_t = k_t$ and that the government budget constraint is

$$g + T_t = (R_t^* - R_t) k_t.$$

Planning Problem. Using the Euler equation to substitute out R_t, the planning problem can be written as[7]

$$\max_{\{c_t, C_t, k_{t+1}\}} \sum_{t=0}^{\infty} \beta^t (u(c_t) + \gamma U(C_t)), \tag{1a}$$

[7] Judd (1985) includes upper bounds on the taxation of capital, which we have omitted because they do not play any important role. As we shall see, positive long run taxation is possible even without these constraints; adding them would only reinforce this conclusion. Upper bounds on taxation play a more crucial role in Chamley (1986). Also, note that Judd (1985) formulates the model in continuous time while we prefer discrete time to facilitate numerical computations. The continuous time model can be understood as a simple limit of our discrete time model.

subject to

$$c_t + C_t + g + k_{t+1} = f(k_t) + (1-\delta)k_t, \tag{1b}$$

$$\beta U'(C_t)(C_t + k_{t+1}) = U'(C_{t-1})k_t, \tag{1c}$$

$$\beta^t U'(C_t)k_{t+1} \to 0. \tag{1d}$$

The government maximizes a weighted sum of utilities with weight γ on capitalists. By varying γ one can trace out points on the constrained Pareto frontier and isolate constrained efficient policies. We sometimes focus on the case with no weight on capitalists $\gamma = 0$, which ensures that desired redistribution runs from capitalists towards workers. Constraint (1b) is simply the resource constraint, or market clearing condition. Constraint (1c) combines the capitalists' first-order condition and budget constraint, (1d) then imposes the transversality condition; together both conditions ensure the optimality of the capitalists' saving decision.

The necessary first-order conditions are

$$\mu_0 = 0, \tag{2a}$$

$$\lambda_t = u'(c_t), \tag{2b}$$

$$\mu_{t+1} = \mu_t \left(\frac{\sigma - 1}{\sigma \kappa_{t+1}} + 1 \right) + \frac{1}{\beta \sigma \kappa_{t+1} \upsilon_t} (1 - \gamma \upsilon_t), \tag{2c}$$

$$\frac{u'(c_{t+1})}{u'(c_t)}(f'(k_{t+1}) + 1 - \delta) = \frac{1}{\beta} + \upsilon_t(\mu_{t+1} - \mu_t), \tag{2d}$$

where $\kappa_t \equiv k_t / C_{t-1}$, $\upsilon_t \equiv U'(C_t)/u'(c_t)$ and the multipliers on constraints (1b) and (1c) are $\lambda_t \beta^t$ and $\mu_t \beta^t$, respectively.

Previous Steady State Results. Judd (1985, pg. 72, Theorem 2) provided a zero-tax result, which we adjust in the following theorem to make explicit the need for the steady state to be interior and for multipliers to converge.

Theorem 1 (Judd, 1985)**.** *Suppose quantities and multipliers converge to an interior steady state, i.e. c_t, C_t, k_{t+1} converge to positive values, and μ_t converges. Then the tax on capital is zero in the limit: $R_t^*/R_t \to 1$.*

The proof is immediate: from equation (2d) we obtain $R_t^* \to 1/\beta$, while from the capitalists' Euler equation we require $R_t \to 1/\beta$. The simplicity of the argument follows from strong assumptions placed on endogenous outcomes. This raises obvious concerns. By adopting assumptions that are close relatives of the conclusions, one may wonder if anything of use has been shown, rather than assumed. We postpone this discussion. Section 3.3 elaborates on this point.

In our rendering of Theorem 1, the requirement that the steady state be interior is important, otherwise, if $c_t \to 0$, one cannot guarantee that $u'(c_{t+1})/u'(c_t) \to 1$ in equation (2d). Likewise, if the allocation converges but μ_t does not, then $v_t(\mu_{t+1} - \mu_t)$ may not vanish in equation (2d). Thus, the two situations that prevent the theorem's application are: (i) non-convergence to an interior steady state; or (ii) non-convergence of multipliers. In general, one expects that (i) implies (ii). The literature has provided an example of (ii) where the allocation does converge to an interior steady state.

Theorem 2. (Lansing, 1999; Reinhorn, 2002 and 2013) *Suppose the allocation converges to an interior steady state, so that c_t, C_t, k_{t+1} converge to positive values. Then, if and only if $\sigma = 1$, multipliers do not converge and*

$$\frac{R_t^*}{R_t} \to \frac{1}{\beta}(1 - \gamma(1 - \beta)v^*),$$

where $v^* = \lim v_t$. This implies a positive long-run tax on capital if redistribution towards workers is desirable, $1 - \gamma v > 0$.

The result follows easily by combining (2c) and (2d) for the case with $\sigma = 1$ and comparing it to the capitalist's Euler equation, which requires $R_t = \frac{1}{\beta}$ at a steady state. Lansing (1999) first presented the logarithmic case as a counterexample to Judd (1985). Reinhorn (2002 and 2013) correctly clarified that in the logarithmic case the Lagrange multipliers explode, explaining the difference in results.[8]

Lansing (1999) depicts the result for $\sigma = 1$ as a knife-edged case: "the standard approach to solving the dynamic optimal tax problem yields the wrong answer in this (knife-edge) case [...]" (from the Abstract, page 423) and "The counterexample turns out to be a knife-edge result. Any small change in the capitalists' intertemporal elasticity of substitution away from one (the log case) will create anticipation effects [...] As capitalists' intertemporal elasticity of substitution in consumption crosses one, the trajectory of the optimal capital tax in this model undergoes an abrupt change." (page 427) This suggests that for $\sigma \neq 1$ the long-run tax on capital is zero. We shall show that this is not the case.

[8] Instead, Lansing (1999) suggests there may be a technical problem with the argument in Judd (1985) specific to $\sigma = 1$. We see no technical difficulty in applying optimal control to the logarithmic case. We believe the issue is exactly what Reinhorn (2002 and 2013) pointed out.

2.1 The Logarithmic Case: IES equal to one

Before studying $\sigma > 1$, it is useful to review the logarithmic case in greater detail. For convenience, we will assume $\gamma = 0$. This guarantees that desired redistribution runs from capitalists to workers.

With logarithmic utility, $\sigma = 1$, capitalists save at a constant rate

$$C_t = (1-s)R_t k_t,$$
$$k_{t+1} = sR_t k_t.$$

Although $s = \beta$ with log preferences, it is worth noting that nothing depends on this fact, and we could start the analysis postulating a constant savings rate s divorced from β.[9]

The planning problem becomes

$$\max_{\{c_t, k_{t+1}\}} \sum_{t=0}^{\infty} \beta^t u(c_t),$$

subject to

$$c_t + \frac{1}{s} k_{t+1} + g = f(k_{t+1}) + (1-\delta)k_t,$$

with k_0 given. This is equivalent to an optimal growth problem with the price of capital set to $\frac{1}{s} > 1$, instead of its true unitary cost. The difference is due to capitalists keeping a fraction $1-s$ for consumption. The government and workers can only save indirectly through capitalists, by giving them more resources today and extracting more tomorrow. From their point of view, technology is less productive because capitalists must be "fed" a fraction of the investment. Note that higher savings rates s imply higher efficiency.[10]

The Euler equation is

$$u'(c_t) = s\beta u'(c_{t+1}) R^*_{t+1}.$$

Since the problem is equivalent to a standard optimal growth problem there is a unique interior steady state and it is globally stable. At this steady state the Euler equation becomes

$$R^* = \frac{1}{\beta s}.$$

[9] This could capture different discount factors between capitalists and workers or an ad hoc behavioral assumption of constant savings, as in the standard Solow growth model.

[10] This kind of wedge in rates of return is similar to that found in countless models where there are financial frictions between "experts" able to produce capital investments and "savers". Often, these models are set up with a moral hazard problem, whereby some fraction of the investment *returns* must be kept by experts, as "skin in the game" to ensure good behavior.

A steady state also requires $k = sRk$, or $R = 1/s$, implying

$$\frac{R^*}{R} = \frac{1}{\beta} > 1.$$

Proposition 1. *Suppose logarithmic utility for capitalists, $U(c) = \log c$. The solution to the planning problem converges monotonically to a unique steady state with a positive tax on capital, $1 - \frac{R}{R^*} = 1 - \beta$.*

This proposition echoes the result in Lansing (1999) specialized to $\gamma = 0$, which we summarized in Theorem 2, with the added conclusion regarding the global stability of the steady state. Capitalists face a positive tax at the steady state. Interestingly, the tax rate depends only on β, not on s or other parameters. In terms of magnitudes, the tax rate on wealth equals $1 - \beta$, since $R = \beta R^*$.

We shall argue that positive long run taxes are not special to logarithmic utility, that these results are not knife edged. One way to proceed would be to exploit continuity of the planning problem with respect to σ to establish that for any fixed time t the solution is continuous in σ, so that $\tau_t(\sigma)$ converges as $\sigma \to 1$ to the positive tax rate in the logarithmic case. While arguing by continuity in this way may be enough to dispel the notion that the logarithmic utility case is irrelevant for $\sigma \neq 1$, it has its limitations. As we shall see, the convergence is not uniform and one cannot invert the order of limits: $\lim_{t\to\infty} \lim_{\sigma\to 1} \tau_t(\sigma)$ will not equal $\lim_{\sigma\to 1} \lim_{t\to\infty} \tau_t(\sigma)$. Therefore, this approach cannot address how the limit tax rate $\lim_{t\to\infty} \tau_t(\sigma)$ behaves as a function of σ. We shall proceed more globally by tackling the problem with $\sigma \neq 1$ directly, which allows us to study the limit tax rate $\lim_{t\to\infty} \tau_t(\sigma)$.

2.2 Positive Long-Run Taxation: IES below one

We now consider the case with $\sigma > 1$ so that the intertemporal elasticity of substitution $\frac{1}{\sigma}$ is below unity. We continue to focus on the situation where no weight is placed on capitalists, $\gamma = 0$, ensuring that redistribution runs from capitalists to workers.

Suppose the allocation converges to an interior steady state $k_t \to k$, $C_t \to C$, $c_t \to c$ with $k, C, c > 0$. This implies that κ_t and υ_t also converge to positive values, κ and υ. In the limit, the first-order conditions imply

$$f'(k) + 1 - \delta = \frac{1}{\beta} + \upsilon(\mu_t - \mu_{t-1}) = \frac{1}{\beta} + \mu_t \frac{\sigma - 1}{\sigma\kappa} \upsilon + \frac{1}{\beta\sigma\kappa}.$$

Since $\sigma > 1$, this implies that μ_t must converge to

$$\mu = -\frac{1}{(\sigma - 1)\beta v} < 0, \tag{3}$$

and so the long-run tax on capital is zero, $f'(k) + 1 - \delta = \frac{1}{\beta}$.

Now consider whether $\mu_t \to \mu < 0$ is possible. From the first-order condition (2a) we have $\mu_0 = 0$. Also, from equation (2c) it follow that whenever $\mu_t \geq 0$ then $\mu_{t+1} \geq 0$. Thus, $\mu_t \geq 0$ for all $t = 0, 1, \ldots$ implying that $\mu_t \to \mu < 0$ is impossible. This shows that the solution cannot converge to the zero-tax steady state. Indeed, it actually proves the solution cannot converge to any interior steady state, since, we argued, the only possible interior steady state is the zero tax steady state. We have shown the following.

Proposition 2. *If $\sigma > 1$ and $\gamma = 0$ then for any initial k_0 the solution to the planning problem does not converge to the zero-tax steady state, or any interior steady state.*

It follows that *if the optimum converges*, then either $k_t \to 0$, $C_t \to 0$ or $c_t \to 0$. With positive spending $g > 0$, then $k_t \to 0$ is not feasible; this then rules out $C_t \to 0$, since capitalists cannot be starved while owning positive wealth.

We have shown that, provided the solution converges, $c_t \to 0$. This in turn implies that either $k_t \to k_g$ or $k_t \to k^g$ where $k_g < k^g$ are the two solutions to $\frac{1}{\beta}k + g = f(k) + (1 - \delta)k$, which uses the fact that $C = \frac{1-\beta}{\beta}k$ at any steady state. We next show that the solution does indeed converge towards k_g and that the long-run tax on capital is strictly positive. The proof uses the fact that $\mu_t \to \infty$ as argued above, but requires many other steps detailed in the appendix.

Proposition 3. *If $\sigma > 1$ and $\gamma = 0$ then for any initial k_0 the solution to the planning problem converges to $c_t \to 0$, $k_t \to k_g$, $C_t \to \frac{1-\beta}{\beta}k_g$, with a positive limit tax on wealth: $1 - \frac{R_t}{R_t^*} \to \tau_g > 0$. The limit tax τ_t is decreasing in spending g, with $\tau_g \to 1$ as $g \to 0$.*

The tax on wealth does not fall towards zero, indeed, the opposite is true. As we illustrate below the tax can be quite sizable and rise over time. The solution does not converge to an interior allocation and multipliers do not converge, invalidating the application Judd (1985), as we summarized here in Theorem 1.

Perhaps counterintuitively, the long-run tax on capital is inversely related to the level of government spending. The reason for this is that the solution drives capital to its lowest feasible value k_g, which is increasing in spending g. Thus, a lower level of spending allows the planner to take long run capital to a lower level. This emphasizes that positive long-run taxation of wealth is not being driven by a budgetary necessity.

A Bellman Equation. The two constraints in the planning problem feature the variables $C_{t-1}, k_t, C_t, k_{t+1}$ and c_t. This suggests a recursive formulation with (k_t, C_{t-1}) as the state and c_t as a control. The associated Bellman equation is then

$$V(k, C_-) = \max_{c \geq 0, (k', C) \in A} \{u(c) + \gamma U(C) + \beta V(k', C)\} \tag{4}$$

$$c + C + k' + g = f(k) + (1 - \delta)k$$

$$\beta U'(C)(C + k') = U'(C_-)k$$

$$c, C, k' \geq 0.$$

for all $(k, C_-) \in A$, where A is the largest set of points (k_0, C_{-1}) such that there exists a sequence $\{k_{t+1}, C_t\}$ satisfying all the constraints in (1) including the transversality condition. Associated with this Bellman equation are policy functions $c = h^c(k, C_-)$, $k' = h^k(k, C_-)$ and $C = h^C(k, C_-)$.

At $t = 0$ capital k_0 is given, but there is no need to impose $\beta U'(C_0)(C_0 + k_1) = U'(C_{-1})k_0$. Thus, the planner maximizes $V(k_0, C_{-1})$ with respect to C_{-1} and the first order condition is

$$V_C(k_0, C_{-1}) = 0.$$

Since $\mu_t = V_C(k_t, C_{t-1})U''(C_{t-1})k_t$ this is akin to the condition $\mu_0 = 0$ in equation (2a).[11]

This recursive approach is useful for a number of reasons. First, it allows us to solve the model numerically. Second, the policy functions allow us to consider the dynamics in the space (k_t, C_{t-1}), which we discuss towards the end of this subsection.

Solution near $\sigma = 1$. Figure 1 displays the time path for the capital stock and the tax rate on wealth, $1 - R_t/R_t^*$, for a range of σ near 1 that straddle the logarithmic case. We set $\beta = 0.95$, $\delta = 0.1$, $f(k) = k^\alpha$ with $\alpha = 0.3$ and $u(c) = U(c)$. Spending g is chosen so that $\frac{g}{f(k)} = 20\%$ at the zero-tax steady state. The initial value of capital, k_0, is set at the

[11] Alternatively, we may can impose that R_0 is taken as given, with $R_0 = R_0^*$ for example, to exclude an initial capital tax. In that case we solve

$$\max_{k_1, c_0, C_0} \{u(c_0) + \gamma U(C_0) + \beta V(k_1, C_0)\}$$

subject to

$$C_0 + k_1 = R_0 k_0$$

$$c_0 + C_0 + k_1 = f(k_0) + (1 - \delta)k_0$$

$$c_0, C_0, k_1 \geq 0$$

This alternative gives rise to similar results.

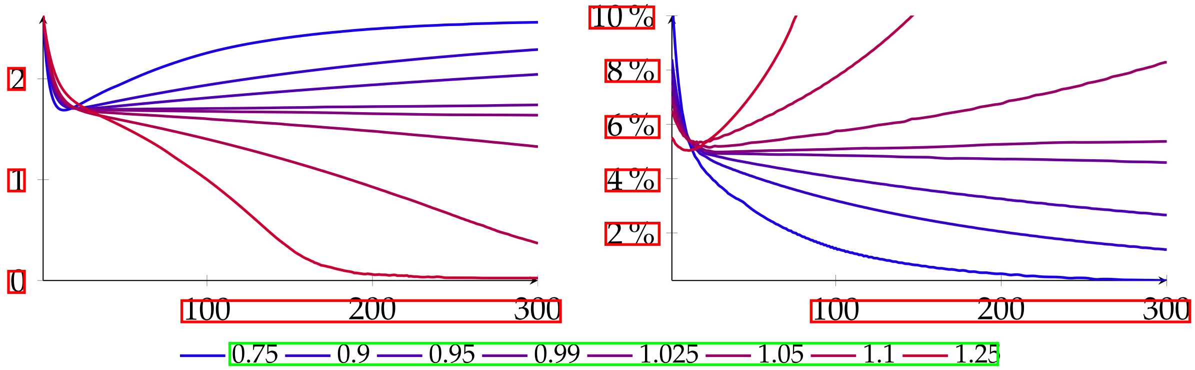

Figure 1: Optimal time paths over 300 years for capital stock (left panel) and wealth taxes (right panel) for various value of σ. Note: tax rates apply to gross returns not net returns, i.e. they represent an annual wealth tax.

zero-tax steady state. Our numerical method is based on the Bellman equation (4) and is described in the appendix.

To clarify the magnitudes of the tax on wealth, consider an example: if $R^* = 1.04$ s so that the before-tax net return is 4%, then a tax on wealth of 1% represents a 25% tax on the net return, a tax of 4% represents a tax rate of 100% on net returns, etcetera.

A few things stand out in Figure 1. First, the results confirm what we showed theoretically in Proposition 3, that for $\sigma > 1$ capital converges to $k_g = 0.0126$. In the figure this convergence is monotone[12] and relatively steady, taking around 200 years for $\sigma = 1.25$. The asymptotic tax rate is very high, approximately $1 - R/R^* = 85\%$, and outside the figure's range. Of course, this implies that the before-tax return $R^* = f'(k_g) + 1 - \delta$ at k_g is exorbitant, because the after-tax return is still $R = 1/\beta$.

Second, for $\sigma < 1$, the path for capital is not monotonic[13] and eventually converges to the zero-tax steady state and the tax rate converges to zero. However, the convergence is relatively slow, especially for values of σ near 1. This makes sense, since, by continuity, for any period t, the solution should converge to that of the logarithmic utility case as $\sigma \to 1$.[14] By implication, for $\sigma < 1$ the rate of convergence to the zero-tax steady state must be zero as $\sigma \uparrow 1$. To further punctuate this point, Figure 2 shows the number of years it takes for the tax on wealth to drop below 1% as a function of $\sigma \in (\frac{1}{2}, 1)$. As σ rises it takes longer and longer and as $\sigma \uparrow 1$ it takes an eternity.

The logarithmic case leaves other imprints on the solutions for $\sigma \neq 1$. Returning to

[12] This depends on the level of initial capital. For lower levels of capital the path first rises then falls.

[13] This is possible because the state variable has two dimensions, (k_t, C_{t-1}). At the optimum, for the same capital k, consumption C is initially higher on the way down than it is on the way up.

[14] Recall that, by Proposition 1, the logarithmic solution converges to positive taxation as $t \to \infty$.

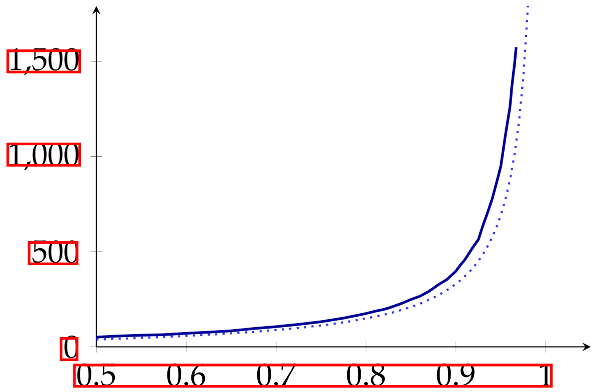

Figure 2: Time elapsed (in years) until tax on wealth falls below 1% for $\sigma \in (\frac{1}{2}, 1)$. The solid line uses the solution of the nonlinear model, the dashed line uses an approximation from the linearized model below.

Figure 1, for both $\sigma < 1$ and $\sigma > 1$ we see that over the first 20-30 years, the path approaches the steady state of the logarithmic utility case, associated with a tax rate around $1 - \frac{R}{R^*} = 1 - \beta = 5\%$. The speed at which this takes place is relatively quick, which is explained by the fact that for $\sigma = 1$ it is driven by the standard rate of convergence in the neoclassical growth model. The solution path then transitions much more slowly either upwards or downwards, depending on whether $\sigma < 1$ or $\sigma > 1$.

An Intuition based on the Intertemporal Manipulation of Saving Incentives. Why does the tax rise for $\sigma > 1$ and fall for $\sigma < 1$? Why are these dynamics relatively slow for σ near 1?

To address these questions about normative results, it helps to back up and review differences in the following positive exercise. Start from a constant tax on wealth and imagine an unexpected announcement for higher future taxation. How do capitalists react today? There are substitution and income effects pulling in opposite directions. When $\sigma > 1$ the substitution effect is weaker and capitalists increase present savings, to partially offset the drop in future consumption.[15] When $\sigma < 1$ the substitution effect is stronger and capitalists decrease present savings, substituting towards current consumption. In the logarithmic case, $\sigma = 1$, the two effects cancel out, so that present consumption and savings are unaffected.

Returning to the normative questions, increasing savings is desirable when capital is currently being taxed, so as to augment the tax base. When $\sigma < 1$, this can be accom-

[15] This does not imply that the supply for savings "bends backward". For instance, if the interest rate were lowered permanently then wealth would rises over time, even with $\sigma > 1$. Higher values of σ are simply associated with a less elastic savings response. Although there is no consensus, the case with $\sigma > 1$ is usually considered the empirically plausible one.

plished by promising lower tax rates in the future; the optimum leans in this direction explaining the declining path for taxes. In contrast, when $\sigma > 1$, the same is accomplished by promising higher future tax rates; this leads to an increasing path for taxes. These incentives are absent in the logarithmic, $\sigma = 1$, case, explaining why the tax rate converges to a constant.

When $\sigma < 1$, the rate of convergence to the zero-tax steady state is driven by these considerations to manipulate savings intertemporally. With σ near 1, the potency and benefit of these manipulations is small, explaining why the rate of convergence is low.

Positive weight on capitalists $\gamma > 0$. At least with $g = 0$, a desire for redistribution is crucial to make the question of wealth taxing nontrivial. Proposition 3 assumes no weight on capitalists, $\gamma = 0$, ensuring that redistribution is desired from capitalists to workers.

When the desire for redistribution is present but weak, so that γ is positive but low, it is natural to expect the same qualitatively results as Proposition 3. When γ is very high, however, desired redistribution may flip and run from workers to capitalists. It is then natural to expect subsidies on wealth, rather than taxes. Figure 3 verifies all of these points, fixing $\sigma = 1.25$ and varying the weight γ.[16]

Nonlinear and Linearized Dynamics. The policy functions associated with the Bellman equation (4) map (k_t, C_{t-1}) into (k_{t+1}, C_t). It is useful to consider the resulting dynamics.

Numerically, when $\sigma < 1$ we find that the zero-tax steady state is globally stable: (k_t, C_{t-1}) converges from any initial condition, even those not satisfying the initial condition $V_C = 0$. When $\sigma > 1$ there are three relevant steady states: two corners, at k_g and k^g, and the zero-tax interior steady state. We find that the latter is saddle path stable. We find that the state (k_t, C_{t-1}) converges to the corners k_g or k^g except on a measure zero of initial conditions given by the stable saddle arms of the zero-tax steady state. The loci $V_C(k, C_-) = 0$ and the stable saddle arm intersect at a single value of k. This point is comparable to the tipping point value of γ in Figure 3, separating solutions with convergence towards k_g from those with convergence towards k^g.

Theoretically, we can verify the local version of these properties. The constraints in (1) and first-order conditions in (2) define a dynamical system for $(k_t, C_{t-1}, \mu_t, \lambda_t)$. We linearize this system around the zero-tax steady state $(k^*, C^*, \mu^*, \lambda^*)$. Transversality con-

[16] Once again, initial capital is set to its zero-tax level. In the legend, rather than displaying γ, we perform a transformation that makes it more easily interpretable: we report the proportional change in consumption for capitalists that would be desired at the steady state, e.g. -0.4 represents that the planner's ideal allocation of the zero-tax output would feature a 40% reduction in the consumption of capitalists, relative to the steady state value $C = \frac{1-\beta}{\beta}k$.

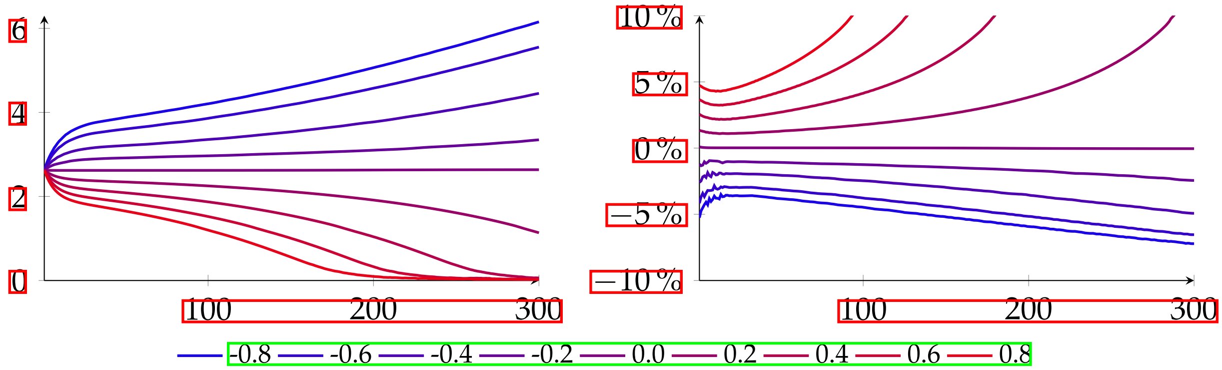

Figure 3: Optimal time paths over 300 years for capital stock (left panel) and wealth taxes (right panel) for various redistribution preferences (zero represents no desire for redistribution; see footnote 16).

ditions then pin down μ_t and λ_t uniquely as a function of (k_t, C_{t-1}). This implies a linearized dynamical system for (k_t, C_{t-1}) of the form

$$\begin{pmatrix} k_{t+1} \\ C_t \end{pmatrix} - \begin{pmatrix} k_t \\ C_{t-1} \end{pmatrix} = \hat{J} \begin{pmatrix} k_t - k^* \\ C_{t-1} - C^* \end{pmatrix}$$

for some $\hat{J}$. We take the continuous-time limit to make our results comparable to those in Kemp et al. (1993) and study

$$\begin{pmatrix} \dot{k} \\ \dot{C} \end{pmatrix} = J \begin{pmatrix} k - k^* \\ C - C^* \end{pmatrix} \tag{5}$$

for some J. The details are found in the appendix. The following properties can be shown.

Proposition 4. *Consider the linearized system (5).*

(a) If $\sigma > 1$, the zero-tax steady state is saddle-path stable.

(b) If $\sigma < 1$ and $\gamma \leq \gamma^*$, the zero-tax steady state is stable.

(c) If $\sigma < 1$ and $\gamma > \gamma^*$, the zero-tax steady state may be stable or unstable and the dynamics may feature cycles.

Here, $\gamma^* = u'(c^*)/U'(C^*)$ is the weight on capitalists which makes the planner indifferent between redistributing towards workers or capitalists at the zero-tax steady state.

The first two points confirm our theoretical and numerical observations for the nonlinear dynamical system. For $\sigma < 1$ the zero tax steady state is locally stable, while for $\sigma > 1$ it is locally saddle-path stable, which explains why the solution does not generally converge to this steady state, as discussed above.[17]

The third point shows that for high γ the system may become unstable or feature cyclical dynamics. This is consistent with Kemp et al. (1993), which studied the linear dynamics around the zero-tax steady state and reported the potential for instability and cycles. Our proposition clarifies that a necessary condition for their results is an assumed desire to redistribute away from workers towards capitalists, $\gamma > \gamma^*$. Instead, when desired redistribution runs from capitalists to workers, so that $\gamma < \gamma^*$ (or when $\sigma > 1$ for any γ) the linearized dynamics cannot feature cycles. Our paper focuses on low values of γ, to ensure redistribution from capitalists to workers. Thus, our results are completely unrelated to those in Kemp et al. (1993).

2.3 Ad Hoc Saving Rules for Capitalists

Up to this point, in keeping with Judd (1985), capitalists have been modeled as infinitely-lived optimizing savers. In addition, we specialized to additively separable utility functions. We now relax both assumptions in one fell swoop.

Assume capitalists save according to an "ad-hoc" savings rule,

$$k_{t+1} = S(R_t k_t; R_{t+1}, R_{t+2}, \dots),$$

where $S(I; R_1, R_2, \dots)$ is some given savings function, taking as arguments current wealth $I = R_t k_t$ and all future interest rates $\{R_{t+1}, R_{t+2}, \dots\}$. This encompasses the case where capitalists maximize an additively separable utility function, but is more general. For example, the savings function may be derived from the maximization of a recursive utility function, or even represent behavior that cannot be captured by optimization, such as hyperbolic discounting or self-control and temptation. This generality will help us discern the key economic mechanism involved in our non-convergence result.

The planning problem is

$$\max_{\{c_t, C_t, k_{t+1}\}} \sum_{t=0}^{\infty} \beta^t \left(u(c_t) + \gamma U(C_t) \right),$$

[17] One can employ the linear system to compute the number of years before the tax rate falls below 1%, a calculation comparable to that performed earlier for the nonlinear model. The dotted line in Figure 2 displays the result, which comes out very close to the nonlinear model. By this measure, for $\sigma < 1$, the dynamics appear to be well approximated by the linearized system.

subject to

$$c_t + R_t k_t + g = f(k_t) + (1-\delta)k_t,$$

$$C_t + S_t = R_t k_t,$$

$$k_{t+1} = S(R_t k_t; R_{t+1}, R_{t+2}, \ldots),$$

with k_0 given.

Non Convergence. Our next result provides a generalization of the negative conclusion in Proposition 2 to our current ad hoc savings setting.

Proposition 5. *Suppose $\gamma = 0$ and assume the savings function is increasing in income, so that $S_I(I, R_1, R_2, \ldots) \geq 0$, and decreasing in future rates, so that $S_{R_\tau}(I, R_1, R_2, \ldots) \leq 0$ for all $\tau = 1, 2, \ldots$ Then the optimum does not converge to the zero-tax steady state.*

The requirement that savings increase with wealth, $S_I \geq 0$, amounts to a standard normality condition. When in addition higher future interest rates lower current savings, $S_{R_\tau} \leq 0$, the economy cannot converge to the zero-tax steady state. The case with separable isoelastic utility with $\sigma > 1$ is a special case satisfying these properties.

As before, the intuition is that the anticipatory effects create the potential to manipulate savings. The planner makes use of these incentives, creating a destabilizing force away from zero taxation.

Elasticities and Taxation. Propositions 2 and 5 show that an interior steady state is not something one can take for granted. Casting these important warnings aside, let us now proceed assuming we do reach an interior steady state for quantities and multipliers.

At a steady state the interest rate R satisfies $k = S(kR; R, R, \ldots)$ and let $v \equiv U'(C)/u'(c)$. Combining the first-order conditions for the above planning problem one can show that

$$\frac{R^*}{R} - 1 = \left(\frac{1}{\beta} - RS_I\right) \frac{1 - \gamma v}{RS_I + \sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau}} + \left(\frac{1}{\beta R} - 1\right) \gamma v, \tag{6}$$

where $\epsilon_{S,\tau} \equiv \frac{R_\tau}{S} \frac{\partial S}{\partial R_\tau}(R_0 k_0; R_1, R_2, \ldots)$ evaluated at a steady state $(kR; R, R, \ldots)$. Note that with additively separable utility the last term in equation (6) is zero, since $\beta R = 1$ at any steady state. This equation is nearly identical to a condition derived by Piketty and Saez (2013, see their Section 3.3, equation 16), although they worked a slightly different model with unlimited government debt, closer to Chamley (1986) in that sense, and also assumed additively separable utility functions.[18] Our derivation shows that these particulars are not essential and allows for an ad hoc savings function.

For simplicity, assume that either $\gamma = 0$ or that preferences are additively separable, so that we can ignore the last term in equation (6). The key implication is then that, assuming $\frac{1}{\beta} - RS_I \neq 0$, an infinite total elasticity $\sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau}$ is required for zero taxation.

As our derivation emphasizes, one virtue of equation (6) is that it does not require optimizing behavior on the part of capitalists. However, when capitalists maximize a recursive utility function—including the special case with additive utility—we can show directly that the total elasticity in the denominator of equation (6) diverges.

Proposition 6. *Suppose capitalists have recursive preferences represented by (7) (see Section 3.1 below) then at any zero tax steady state*

$$\sum_{\tau=1}^{T} \beta^{-\tau} \epsilon_{S,\tau}$$

diverges to $+\infty$ or $-\infty$ as $T \to \infty$.

With additive separable utility, the long run elasticity of savings is often characterized as being infinite, in the sense that at any steady state $R = \frac{1}{\beta}$ and β is a fixed parameter. Away from the additive case, with recursive utility, the elasticity of long-run savings may be finite, in that various values for steady state R are possible. It may then seem natural to expect $\sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau}$ to also be finite, yet our proposition shows that this is not possible.

Thus, with recursive utility, if the allocation converges to the interior steady state this proposition combined with equation (6), anticipates a zero tax. Both Judd (1985) and Chamley (1986) made efforts to state results with recursive non-additive utility, so as to avoid assuming an infinite elasticity for long-run savings. However, Proposition 6 verifies that the relevant long-run elasticity is infinite even when preferences are not additively separable. In this sense, despite their efforts, Judd (1985) and Chamley (1986) did not completely avoid an infinite elasticity.

It is absolutely crucial to understand that any reliance on equation (6) and the total elasticity $\sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau}$ to discuss long run taxation must be qualified: an infinite elasticity is only necessary, not sufficient for zero long run taxation. For example, in the additively separable iso-elastic utility case, when $\sigma < 1$ one can show that $\sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau} = \infty$; we also found numerically that the solution converges to the zero-tax steady state. This is

[18]For example, because of additive separability, in their condition $RS_I = 1$. There are other minor difference, for example, they express the formula in terms of changes in the path for capital, whereas (6) is expressed in terms of the derivatives of our ad-hoc savings function.

consistent with equation (6). In contrast, when $\sigma > 1$ one has $\sum_{t=1}^{\infty} \beta^{-t} \epsilon_{S,\tau} = -\infty$, and if one were to consult equation (6) it would still indicates a zero tax. However, we have shown that the solution does not converge to an interior steady state and that taxation is nonzero. More generally, Proposition 5 shows that even if $\sum_{t=1}^{\infty} \beta^{-t} \epsilon_{S,\tau}$ does not diverge, whenever $\epsilon_{S,\tau} < 0$, we cannot converge to an interior steady state. It follows that equation (6) is not helpful and, indeed, often misleading. The pitfall is that an interior steady state, for quantities and multipliers, required for its derivation, cannot be assumed.[19]

3 Representative Agent Ramsey Model: Chamley (1986) and Judd (1999) Revisited

Up to this point we worked within the two-class model introduced by Judd (1985), without government debt. Chamley (1986) worked with a representative agent Ramsey model with unconstrained government debt. Judd (1999) adopts the same assumptions. This section presents results using a representative agent model, directly comparable to these two papers.

We first discuss the solution when the bounds on capital taxation do not bind in the long run. We then discuss the possibility of bounds binding indefinitely.

3.1 First Best or Zero Taxation of Zero Wealth?

Just as Judd (1985), Chamley (1986) considered non-additively separable utility to avoid an "infinite long-run elasticity of savings", as the additively separable utility case is commonly characterized.[20] Unlike additively separable utility, recursive preferences allows the rate of impatience to vary with consumption levels. We will show that, if the optimum converges to a steady state where bounds on taxation are not binding, then the tax on capital is indeed zero. However, the model yields other, hitherto unnoticed, implications as long as the rate of impatience is not constant.

To distinguish this subsection from the next, we focus here on situations where the bounds on capital taxation are not binding in the long run. The next subsection addresses the important possibility of indefinitely binding bounds.

[19] The derivation also assumes there are no bounds on capital taxation. As we shall see, this is crucial in the context of the Chamley (1986) model.

[20] At any steady state with additive utility one must have $R = 1/\beta$ for a fixed parameter $\beta \in (0,1)$. This is true regardless of the wealth or consumption level. In this sense, the supply of savings is infinitely elastic at this rate of interest.

Preferences. Utility satisfies a Koopmans (1960) recursion

$$V_t = W(U_t, V_{t+1}) \qquad \text{where} \qquad U_t = U(c_t, n_t). \tag{7}$$

Here $U(c, n)$ is the period utility function and $W(U, V')$ is an aggregator function; both functions are taken to be twice continuously differentiable. We assume consumption and leisure are normal goods,

$$\frac{U_{cc}}{U_c} - \frac{U_{nc}}{U_n} \leq 0 \qquad \text{and} \qquad \frac{U_{cn}}{U_c} - \frac{U_{nn}}{U_n} \leq 0,$$

with at least one strict inequality.

Regarding the aggregator function, the additively separable utility case amounts to the particular linear choice $W(U, V') = U + \beta V'$ with $\beta \in (0, 1)$. Nonlinear aggregators allow local discounting to vary with U and V', as in Koopmans (1960), Uzawa (1968) and Lucas and Stokey (1984). Of particular interest is how the discount factor varies across potential steady states. Define $\bar{U}(V)$ as the solution to $V = W(\bar{U}(V), V)$ and let $\bar{\beta}(V) \equiv W_{V'}(\bar{U}(V), V)$ denote the steady state discount factor.

Planning problem. The *implementability condition* for this economy is[21]

$$\sum_{t=0}^{\infty} \beta_t W_{Ut}(U_{ct}c_t + U_{nt}n_t) = W_{U0}U_{c0}R_0a_0, \tag{8}$$

where $\beta_t = \prod_{s=0}^{t-1} W_{V's}$. The derivation is standard. In the additive separable utility case $\beta_t W_{t,U} = \beta^t$ and expression (8) reduces to the standard implementability condition popularized by Lucas and Stokey (1983) and Chari et al. (1994). Given R_0, any allocation satisfying the implementability condition and the resource constraint

$$c_t + k_{t+1} + g_t = F(k_t, n_t) + (1 - \delta)k_t \qquad t = 0, 1, \ldots \tag{9}$$

can be sustained as a competitive equilibrium for some sequence of prices and taxes.[22]

[21] We use the shorthand notation W_{Ut} to represent $W_U(U_t, V_{t+1})$, etc.

[22] The argument is identical to that in Lucas and Stokey (1983) and Chari et al. (1994).

To enforce upper bounds on the taxation of capital in period $t = 1, 2, \ldots$ we impose

$$W_{Ut} U_{ct} = R_{t+1} W_{V't} W_{Ut+1} U_{ct+1}, \tag{10a}$$

$$R_{t+1} = (1 - \tau_t)(F_{kt+1} - \delta) + 1, \tag{10b}$$

$$\tau_t \leq \bar{\tau}. \tag{10c}$$

The planning problem maximizes V_0 subject to (7), (8), (9) and (10). The bounds $\tau_t \leq \bar{\tau}$ may or may not bind forever. In this subsection we are interested in situations where the bounds do not bind asymptotically, i.e. they are slack after some date $T < \infty$. In the next subsection we discuss the possibility of the bounds binding forever.

Chamley (1986) provided the following result, slightly adjusted here to make explicit the need for the steady state to be interior, for multipliers to converge and for the bounds on taxation to be asymptotically slack.

Theorem 3 (Chamley, 1986)**.** *Let $\hat{\Lambda}_t = \beta_t W_{Ut} U_{ct} \Lambda_t$ denote the multiplier on the resource constraint in period t. Suppose the optimum converges to an interior steady state where the constraints on capital taxation are asymptotically slack. Suppose further that the multiplier Λ_t converges to an interior point $\Lambda_t \to \bar{\Lambda} > 0$. Then the tax on capital converges to zero $\frac{R_t}{R_t^*} \to 1$.*

Chamley (1986) actually worked with the particular bound $\bar{\tau} = 1$, implying a constraint $r_t \geq 0$, but stated "the net rate of return [...] is constrained, and assumed to be greater than some arbitrary value, M. Without loss of generality, M can be chosen to be equal to zero".[23]

The main result of this subsection is stated in the next proposition. Relative to Theorem 3, we make no assumptions on multipliers. More importantly, we derive new steady state conclusions.

Proposition 7. *Suppose the planning problem converges to an interior steady state and assume that asymptotically the constraints on capital taxation are slack. Then the tax on capital is zero. In addition, if the discount factor is locally sensitive, so that $\bar{\beta}'(V) \neq 0$, then either*

(a) private wealth converges to zero: $a_t \to 0$; or

(b) the allocation converges to the first-best, with zero taxation of labor.

[23] For $\bar{\tau} = 1$ it is enough to assume that the multiplier converges to state this result and there is no need to qualify that the bounds on capital taxation are not binding. Thus, this is the form that Theorem 1 in Chamley (1986) takes. Although, the assumption of converging multipliers is not stated, but imposed within the proof.

At any interior steady state, if the bounds on capital taxation are not binding, the tax on capital is zero; this much echoes Chamley (1986), or our rendering of the result in Theorem 3. However, as long as the rate of impatience is not locally constant, $\bar{\beta}'(V) \neq 0$, the proposition shows that a zero tax on capital must be accompanied by a zero tax on labor or with zero private wealth. In other words, if taxation of capital is sufficiently unconstrained so that the solution approaches a steady state where the bounds do not bind, there are two possibilities. In the first, the tax base for capital taxation must have been driven to zero, perhaps due to its heavy taxation in the transition to the steady state. The second possibility is that the government accumulate enough wealth against the private sector, perhaps aided by heavy taxation of capital, to finance itself without the distortionary taxation, achieving the first best.

Both scenarios require the government to accumulate a significant positive assets position, i.e. negative debt. In the first case, since $a_t = k_t + b_t \to 0$, the government must own the entire capital stock. In the second, the government must accumulate enough assets to finance government spending, forgoing labor and capital taxation. Quantitatively, the second scenario is likely to require greater asset accumulation than the first.

Both these scenarios are markedly different from the usual interpretation of the zero long-run tax result, which implies that we should forego obtaining revenue from capital taxation but not from labor taxation—relatively unfriendly to capital taxation. Instead, here both labor and capital are either treated symmetrically or nothing remains to be taxed in the case of capital—a scenario that is downright friendly to capital taxation, more symmetrical with labor.

To be sure, in the special, but commonly assumed, case with additive separable utility positive private wealth, a zero tax on wealth and a positive tax on labor coexist at a steady state. However, this is not possible whenever the rate of impatience is locally not constant. In this sense, the usual interpretation—where labor bears the entire burden of taxation and private wealth is spared—is a knife edge case.

3.2 Long Run Capital Taxes with Constraints

As is well understood, without constraints on capital taxation the problem of capital taxation becomes trivial: the solution involves extraordinarily high initial capital taxation, typically complete expropriation, unless the first best is achieved first. Taxing the given initial capital stock mimics the missing lump-sum tax, which has no distortionary effects.

This motivated Chamley (1986) and the subsequent literature to impose upper bounds on taxation, $\tau_t \leq \bar{\tau}$.[24] We now show that these bounds on capital taxes may bind forever, contradicting a claim by Chamley (1986). This claim has been echoed throughout the literature, e.g. Judd (1999), Atkeson et al. (1999) and others.

For our present purposes, and following Chamley (1986), as well as Judd (1999), it is convenient to work with a continuous-time version of the model and restrict attention to additively separable preferences,[25]

$$\int_0^\infty e^{-\rho t} u(c_t, n_t) dt. \tag{11a}$$

$$u(c,n) = U(c) - v(n) \qquad \text{with} \qquad U(c) = \frac{c^{1-\sigma}}{1-\sigma}, \quad v(n) = \frac{n^{1+\zeta}}{1+\zeta}, \tag{11b}$$

where $\sigma, \zeta > 0$. Following Chamley (1986), we adopt an iso-elastic utility function over consumption; this is important because it ensures the bang-bang nature of the solution. For convenience, we also assume iso-elastic disutility from labor; this assumption is not crucial. The resource constraint is

$$c_t + \dot{k}_t + g = f(k_t, n_t) - \delta k_t. \tag{12}$$

where f is concave, homogeneous of degree one and differentiable. The implementability condition is

$$\int_0^\infty e^{-\rho t} \left(u_c(c_t, n_t) c_t + u_n(c_t, n_t) n_t \right) = u_c(c_0, n_0) a_0, \tag{13}$$

where $a_0 = k_0 + b_0$ denotes initial private wealth, consisting of capital k_0 and bonds b_0. To enforce bounds on capital taxation we further impose

$$\theta_t = u_c(c_t, n_t), \tag{14a}$$

$$\dot{\theta}_t = \theta_t(\rho - r_t), \tag{14b}$$

$$r_t = (1 - \tau_t) r_t^*, \tag{14c}$$

$$r_t^* = f_k(k_t, n_t) - \delta, \tag{14d}$$

$$\tau_t \leq \bar{\tau}, \tag{14e}$$

[24] A typical story for the bounds is tax compliance constraints—capital owners would hide capital or mask its returns if taxation were too onerous. Another motivation, although outside the present scope of the representative agent Chamley (1986) model, are political economy constraints on redistribution from capital owners, a point made by Saez (2013). Finally, another possibility is that bounds on capital taxation reflect self-imposed institutional constraints introduced to mitigate the time inconsistency problem.

[25] Continuous time allowed Chamley (1986) to exploit the bang-bang nature of the optimal solution. Since we focus on cases where this is not the case it is less crucial for our results. However, we prefer to keep the analyses comparable.

for some bound $\bar{\tau} > 0$. The planning problem maximizes (11a) subject to (12), (13) and (14).

Chamley (1986, Theorem 2, pg. 615) formulated the following claim regarding the dynamic path of capital taxation.[26]

Claim. *Suppose $\bar{\tau} = 1$ and that preferences are given by* (11)*. Then there exists a time T with the following three properties:*

(a) for $t < T$, the constraint $\tau_t \leq \bar{\tau}$ is binding;

(b) for $t > T$ capital income is untaxed: $r_t = r_t^*$ and $\tau_t = 0$;

(c) $T < \infty$.

At a crucial juncture in the proof of this claim, Chamley (1986) states in support of part (c) that "The constraint $r_t \geq 0$ cannot be binding forever (the marginal utility of private consumption [...] would grow to infinity [...] which is absurd)."[27] Our next proposition establishes that there is nothing absurd about this within the logic of the model and that, quite to the contrary, part (c) of the above claim is incorrect: indefinite taxation, $T = \infty$, may be optimal.

Proposition 8. *Suppose $\bar{\tau} = 1$ and that preferences are given by* (11) *with $\sigma > 1$. Fix any initial capital stock k_0. There exists $\underline{b} < \bar{b}$ such that for $b_0 \in [\underline{b}, \bar{b}]$ the optimum has $\tau_t = \bar{\tau}$ for all $t \geq 0$, i.e. the bound on capital taxation binds forever, and there is no equilibrium with $b > \bar{b}$.*

Here $\bar{b}$ represents the peak of a "Laffer curve", above which there is no equilibrium. The proposition states that for intermediate debt levels it is optimal to tax capital indefinitely. Since these points are below the peak of the Laffer curve, indefinite taxation is not driven by budgetary need—there are feasible plans with $T < \infty$, however, the plan with $T = \infty$ is simply better. This is illustrated in figure 4 with a qualitative plot of the set of states (k_0, b_0) for which indefinite capital taxation is optimal. Although this proposition only considers $\bar{\tau} = 1$, as in Chamley (1986), it is natural to conjecture that lower values of $\bar{\tau}$ make the optimality of $T = \infty$ more likely.

Our next result assumes $g = 0$ and constructs the solution for a set of initial conditions that allow us to guess and verify its form.

[26] Similar claims are made in Atkeson et al. (1999), Judd (1999) and many other papers.

[27] It is worth pointing out, however that although Chamley (1986) claims $T < \infty$ it never states that T is small. Indeed, it cautions to the possibility that it is quite large saying "the length of the period with capital income taxation at the 100 per cent rate can be significant."

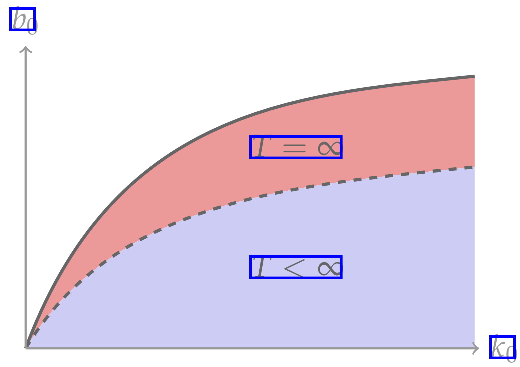

Figure 4: Graphical representation of Proposition 8.

Proposition 9. *Suppose that $\bar{\tau} = 1$, that preferences are given by (11) with $\sigma > 1$, and that $g = 0$. There exists $\underline{k} < \bar{k}$ and $b_0(k_0)$ such that: for any $k_0 \in (\underline{k}, \bar{k}]$ and initial debt $b_0(k_0)$ the optimum satisfies $\tau_t = \bar{\tau}$ for all $t \geq 0$ and $c_t, k_t, n_t \to 0$ exponentially with constant n_t/k_t and c_t/k_t.*

This proposition shows that for well chosen initial conditions and no government spending, the solution path converges to zero in an homogenous, constant growth rate fashion. This feature is certainly special, yet helpful in constructing the appropriate guess explicitly.

This explicit example illustrates that convergence takes place, but not to an interior steady state. Indeed, with additive separable utility no interior steady state exists when $T = \infty$. With $\bar{\tau} = 1$ the after tax interest rate is zero when the bound is binding, but the agent discounts the future positively, preventing a steady state. Instead, with $\bar{\tau} < 1$ the interest rate may be positive and the after tax interest could equal the discount rate ρ. With additive separable utility, this ensures constant consumption, but does not ensure constant labor. If the tax on labor is not constant then labor is generally not constant. Indeed, this appears to be a feature of the optimum, with the planner neutralizing the intertemporal substitution of labor, induced by the capital tax, with an ever increasing labor tax. We conjecture that a steady state with $T = \infty$ may be possible if the tax on\ labor is also constrained by an upper bound.

3.3 Revisiting Judd (1999)

Up to this point we have focused on the Chamley-Judd zero-tax results. A followup literature has offered both extensions and interpretations. One notable case doing both is Judd (1999). This paper follows Chamley (1986) closely, setting up a representative agent economy with perfect financial markets and unrestricted government bonds. More specifically, the paper does two things. First, it provides a variant of the result in Chamley (1986) which does not require the allocation nor multipliers to converge to a steady state. Instead, the paper postulates that an endogenous multiplier lies within an interval and shows that this implies that capital taxation must be zero on average. Second, the paper offers a connection between capital taxation and rising consumption taxes to provides intuition for these results. Let us consider each point in turn.

Bounded Multipliers and Zero Average Capital Taxes. Using our setup from Section 3.2 the main result can be restated as follows. Assuming $\bar{\tau} = 1$, then the planning problem maximizes (11a) subject to (12), (13), (14a), (14b), and $r_t \geq 0$. Let $\hat{\Lambda}_t = e^{-\rho t}\theta_t \Lambda_t$ denote the co-state for capital, i.e. the multiplier on equation (12), satisfying $\dot{\hat{\Lambda}}_t = \rho \hat{\Lambda}_t - r_t^* \hat{\Lambda}_t$. Using that $\frac{\dot{\hat{\Lambda}}_t}{\hat{\Lambda}_t} = -\rho + \frac{\dot{\theta}_t}{\theta_t} + \frac{\dot{\Lambda}_t}{\Lambda_t}$ and $\frac{\dot{\theta}_t}{\theta_t} = \rho - r_t$ this is equivalent to

$$\frac{\dot{\Lambda}_t}{\Lambda_t} = r_t - r_t^*.$$

By the very construction of Λ_t, if one were to assume that Λ_t converges, it would immediately follow that $r_t = r_t^*$ in the long run. This is one way to restate the Chamley (1986) steady state result,[28] which, due to the power of the assumption on Λ_t, even makes any assumptions on the long-run behavior of the allocation unnecessary. Judd (1999, pg. 13, Theorem 6) goes down this route but instead of assuming convergence of Λ_t, he hypothesizes that the endogenous multiplier Λ_t never leaves a certain interval.

Theorem 4 (Judd, 1999)**.** *Let $e^{-\rho t}\theta_t \Lambda_t$ denote the capital co-state on equation* (12) *and assume*

$$\Lambda_t \in [\underline{\Lambda}, \bar{\Lambda}],$$

for $0 < \underline{\Lambda} \leq \bar{\Lambda} < \infty$. Then the cumulative distortion up to t is bounded,

$$\log\left(\frac{\Lambda_0}{\bar{\Lambda}}\right) \leq \int_0^t (r_s - r_s^*)ds \leq \log\left(\frac{\Lambda_0}{\underline{\Lambda}}\right),$$

and the average distortion converges to zero,

$$\frac{1}{t}\int_0^t (r_s - r_s^*)ds \to 0.$$

[28] See our Theorem 3.

In particular, under the conditions of this theorem, the optimum cannot converge to a steady state with a positive tax on capital. More generally, the condition requires departures of r_t from r_t^* to average zero. As stated, the result is somewhat sensitive to the assumption that $\bar{\tau} = 1$; when $\bar{\tau} \neq 1$ and technology is nonlinear, the co-state equation acquires other terms, associated with the bounds on capital taxation.

It is important to note that the proof above proceeded without using any other optimality condition other than that for capital k_t.[29] In particular, it does not invoke the first-order condition for the interest rate r_t or tax rate τ_t. A result on optimal capital taxation is derived without the optimality condition for the tax rate on capital. Naturally, this poses two questions. How is this possible unless the bounds on Λ_t essentially assume the result? And are these bounds on Λ_t consistent with an optimum?

Regarding the first question, the multiplier $\hat{\Lambda}_t$ represents the social marginal value to the planner of resources at time t; thus, its growth rate, $\frac{\dot{\hat{\Lambda}}_t}{\hat{\Lambda}_t}$, should represents the social marginal rate of intertemporal substitution (MRS). When Λ_t is constant, social and private MRSs coincide: $\frac{\dot{\hat{\Lambda}}_t}{\hat{\Lambda}_t} = \frac{\dot{\theta}_t}{\theta_t} - \rho$. Without a conflict between social and private MRS, there is no need to introduce an intertemporal wedge. Therefore, imposing bounds on Λ_t is tantamount to assuming the result.

We already have an answer to the second question, in the form of Proposition 8, showing that indefinite taxation may be optimal.

Corollary. *At the optimum described in Proposition 8 we have that $\Lambda_t \to 0$ as $t \to \infty$. Thus, the hypothesis in Judd (1999) is violated.*

In other words, at the optimum there is no guarantee that the endogenous object Λ_t remains bounded away from zero, as postulated by Judd (1999). Theorem 4 is inapplicable.

Exploding Consumption Taxes. Judd (1999) also offers an intuitive interpretation for the Chamley-Judd result pointing out that a tax on capital is equivalent to an increasing tax on consumption. This casts indefinite taxation of capital as a villain, since increasing and unbounded taxes on consumption do not seem intuitively reasonable and seemingly contradict standard commodity tax principles, as enunciated by Diamond and Mirrlees (1971), Atkinson and Stiglitz (1972) and others.

We believe the equivalence between capital taxation and a rising path for consumption taxes is accurate and useful. It does explain why prolonging capital taxation comes

[29] In this continuous time optimal control formulation, the costate equation for capital is the counterpart to the first-order condition with respect to capital in a discrete time formulation. Indeed, the same result can be easily formulated in a discrete time setting.

at an efficiency cost by distorting the consumption path. If the marginal cost from this distortion were increasing in T and approached infinity at $T = \infty$ this would indeed imply insurmountable costs and provide a strong argument against indefinite taxation of capital. An instinctive aversion to consumption taxes that rise without bound is perhaps based on an intuition for this possibility.

We next formalize the efficiency costs of capital taxation generated by the intertemporal distortions to the consumption path. However, despite the connection with ever increasing consumption taxes, we show that the marginal cost of these distortions remains bounded, even as $T \to \infty$, so that there are no insurmountable costs. These bounded marginal costs must be traded off against the marginal benefits, explaining why the corner solution $T = \infty$ is potentially optimal. The intuitive argument against capital taxation based on a connection with ever increasing consumption taxes does not deliver.

We proceed with a constructive argument and assume, for simplicity, that technology is linear, so that $f(k, n) - \delta k = r^* k + w^* n$ for fixed parameters r^* and n^*.

Proposition 10. *Suppose utility is given by (11), with $\sigma > 1$. Suppose technology is linear. Then the solution to the planning problem can be obtained by solving to the following static problem:*

$$\max_{T,c,n} \quad u(c) - v(n), \tag{15}$$

$$s.t. \quad (1 + \psi(T))\, c + G = k_0 + \omega n,$$

$$c - \frac{v'(n)}{u'(c)} n = (1 - \tau(T)) a_0,$$

where c and n are measures of lifetime consumption and labor supply, respectively, G is the present value of government consumption and ωn is the present value of labor income, for some fixed $\omega > 0$. The functions ψ and τ are increasing with $\psi(0) = \tau(0) = 0$ and

$$\frac{d\psi}{d\tau} \equiv \frac{\psi'(T)}{\tau'(T)}$$

is bounded away from infinity.

Given c, n and T we can compute the paths for consumption c_t and labor n_t. Behind the scenes, the static problem solves the dynamic problem. In particular, it optimizes over the path for labor taxes. In this static representation, $1 + \psi(T)$ is akin to a production cost of consumption and $\tau(T)$ to a non-distortionary capital levy. On the one hand, higher T increases the efficiency cost from the consumption path. On the other hand, it increases revenue in proportion to the level of initial capital. Prolonging capital taxation requires trading off these costs and benefits.

Importantly, despite the connection between capital taxation and an ever increasing unbounded tax on consumption, the proposition shows that the tradeoff between costs and benefits is bounded, $\frac{d\psi}{dT} < \infty$, even as $T \to \infty$. In other words, indefinite taxation does not come at an infinite marginal cost and helps explain why this may be optimal.

Should we be surprised that these results contradict commodity tax principles, as enunciated by Diamond and Mirrlees (1971), Atkinson and Stiglitz (1972) and others? No, not at all. As general as these frameworks may be they do not consider upper bounds on taxation, the crucial ingredient in Chamley (1986) and Judd (1999). Thus, their guiding principles are ill adapted to these settings. In particular, formulas based on local elasticities do not apply, without further modification. When bounds are introduced, this constrains taxation to be lower in the short run, but spreads it over a longer, possibly infinite, horizon.

4 A Hybrid: Redistribution and Debt

Throughout this paper we have strived to stay on target and remain faithful to the original models supporting the Chamley-Judd result. This is important so that our own results are easily comparable to those in Judd (1985) and Chamley (1986). However, many contributions since then offer modifications and extensions of the original Chamley-Judd models and results. In this section we depart briefly from our main focus to show that our results transcend their original boundaries and are relevant to this broader literature.

To make this point with a relevant example, we consider a hybrid model, with redistribution between capitalists and workers as in Judd (1985), but sharing the essential feature in Chamley (1986) of unrestricted government debt. It is very simple to modify the model in Section 2 in this way. We add bonds to the wealth of capitalists $a_t = k_t + b_t$, modifying equation (1c) to

$$\beta U'(C_t)(C_t + k_{t+1} + b_{t+1}) = U'(C_{t-1})(k_t + b_t)$$

and the transversality condition to $\beta^t U'(C_t)(k_{t+1} + b_{t+1}) \to 0$. Equivalently, we have the present value *implementability condition*,

$$\sum_{t=0}^{\infty} \beta^t U'(C_t)C_t = U'(C_0)R_0(k_0 + b_0),$$

With $U(C) = C^{1-\sigma}/(1-\sigma)$ this is

$$(1-\sigma)\sum_{t=0}^{\infty}\beta^t U(C_t) = U'(C_0)R_0(k_0+b_0). \tag{16}$$

Anticipated Confiscatory Taxation. For $\sigma > 1$ the left hand side in equation (16) is decreasing in C_t and the right hand side is decreasing in C_0. It follows that one can take a limit with the property that $C_t \to 0$ for all $t = 0, 1, \ldots$, which is optimal for $\gamma = 0$. Along this limit $R_1 \to 0$, so the tax on capital is exploding to infinity. This same logic applies if the tax is temporarily restricted for periods $t \leq T-1$ for some given T, but is unrestricted in period T.

Proposition 11. *Consider the two-class model from Section 2 but with unrestricted government bonds. Suppose $\sigma > 1$ and $\gamma = 0$. If capital taxation is unrestricted in at least one period, then the optimum features an infinite tax in some period and $C_t \to 0$ for all $t = 0, 1, \ldots$*

This result exemplifies how extreme the tax on capital may be without bounds. In addition to this result and even when $\sigma < 1$, if no constraints are imposed on taxation, except at $t = 0$, then in the continuous time limit as the length of time periods shrinks to zero, taxation tends to infinity. This point was also raised in Chamley (1986) for the representative agent Ramsey model, and served as a motivation for imposing stationary upper bounds, $\tau_t \leq \bar{\tau}$.

Long Run Taxation with Constraints. We now impose upper bounds on capital taxation and show that these constraints may bind forever, just as in Section 3.2. As we did there, it is convenient to switch to a continuous-time version of the model.

Proposition 12. *Consider the two-class model from Section 2 but with unrestricted government bonds. Suppose $\sigma > 1$ and $\gamma = 0$. If capital taxation is restricted by $\tau_t \leq \bar{\tau}$ for some $\bar{\tau} > 0$, then at the optimum $\tau_t = \bar{\tau}$, i.e. capital should be taxed indefinitely.*

These results hold for any value of $\bar{\tau} > 0$, not just $\bar{\tau} = 1$. Intuitively, $\sigma > 1$ is enough to ensure indefinite taxation of capital because $\gamma = 0$ makes it optimal to tax capitalists as much as possible. Similar results hold for positive but low enough levels of γ, so that redistribution from capitalists to workers is desired.

This proposition assumes that transfers are perfectly targeted to workers. However, indefinite taxation, $T = \infty$, remains optimal when this assumption is relaxed, so that transfers are also received by capitalists.

We have also maintained the assumption from Judd (1985) that workers do not save. In a political economy context, Bassetto and Benhabib (2006) study a situation where all agents save (in our context, both workers and capitalists) and are taxed linearly at the same rate. Indeed, they report the possibility that indefinite taxation is optimal for the median voter.

Overall, these results suggest that indefinite taxation is optimal in a range of models that are descendants of Chamley-Judd, with a wide range of assumptions regarding the environment, heterogeneity, social objectives and policy instruments.

5 Conclusions

This study revisited two closely related models and results, Judd (1985) and Chamley (1986). Our own results contradict the standard results and their interpretation, revealing that the long run tax on capital is not necessarily zero in these models.

Why were the proper conclusions missed by Judd (1985), Chamley (1986) and others? These papers assume that endogenous multipliers associated with the planning problem converge. We have shown that this is not necessarily the case. In fact, as we argued, postulating convergence for endogenous multipliers is equivalent to postulating that the planner's marginal rate of substitution equals the planner's, so that no intertemporal distortion is required. In this sense, assuming this convergence amounts to little more than assuming zero long-run taxes. The original proofs are magically so simple that they almost feel too good to be true. And indeed, they are.

In quantitative evaluations it may well be the case that one finds a zero long-run tax on capital, e.g. for the model in Judd (1985) one may set $\sigma < 1$, and in Chamley (1986) the bounds may not bind forever, depending on parameters.[30] In this paper we stay away from making any such claim, one way or another. We confined attention to the original theoretical results, widely perceived as delivering zero long-run taxation as an ironclad conclusion, independent of parameter values. Based on our analysis, we find little basis for such an interpretation.

References

Abel, Andrew B., "Optimal Capital Income Taxation," Working Paper 13354, National

[30] Any quantitative exercise may also evaluate the welfare gains, not just the optimum. Even when $T < \infty$ is optimal, the value of T may be high and it is possible that indefinite taxation provides a good approximation to the optimum.

Bureau of Economic Research 2007.

Atkeson, Andrew, Varadarajan V Chari, and Patrick J Kehoe, "Taxing capital income: a bad idea," *Federal Reserve Bank of Minneapolis Quarterly Review*, 1999, *23*, 3–18.

Atkinson, Anthony B and Joseph E Stiglitz, "The structure of indirect taxation and economic efficiency," *Journal of Public Economics*, 1972, *1* (1), 97–119.

Banks, James and Peter A Diamond, "The base for direct taxation," in "Dimensions of Tax Design: The Mirrlees Review," Oxford University Press for The Institute for Fiscal Studies, 2010, chapter 6, pp. 548–648.

Bassetto, Marco and Jess Benhabib, "Redistribution, taxes, and the median voter," *Review of Economic Dynamics*, 2006, *9* (2), 211–223.

Brunnermeier, Markus K, Thomas M Eisenbach, and Yuliy Sannikov, "Macroeconomics with financial frictions: A survey," Technical Report, National Bureau of Economic Research 2012.

Chamley, Christophe, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 1986, *54* (3), pp. 607–622.

Chari, V.V., Lawrence J. Christiano, and Patrick J. Kehoe, "Optimal Fiscal Policy in a Business Cycle Model," *Journal of Political Economy*, 1994, *102* (4), 617–52.

Coleman II, Wilbur John, "Welfare and optimum dynamic taxation of consumption and income," *Journal of Public Economics*, 2000, *76* (1), 1–39.

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

Gertler, Mark and Nobuhiro Kiyotaki, "Chapter 11 - Financial Intermediation and Credit Policy in Business Cycle Analysis," in Benjamin M. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3, Elsevier, 2010, pp. 547 – 599.

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

—, "Optimal taxation and spending in general competitive growth models," *Journal of Public Economics*, 1999, *71* (1), 1–26.

—, "Capital-Income Taxation with Imperfect Competition," *American Economic Review Papers and Proceedings*, 2002, *92* (2), pp. 417–421.

Kemp, Murray C, Ngo Van Long, and Koji Shimomura, "Cyclical and Noncyclical Redistributive Taxation," *International Economic Review*, May 1993, *34* (2), 415–29.

Koopmans, Tjalling C, "Stationary ordinal utility and impatience," *Econometrica: Journal of the Econometric Society*, 1960, pp. 287–309.

Lansing, Kevin J., "Optimal redistributive capital taxation in a neoclassical growth model," *Journal of Public Economics*, 1999, *73* (3), 423 – 453.

Lucas, Robert E., Jr. and Nancy L. Stokey, "Optimal Fiscal and Monetary Policy in an Economy without Capital," *Journal of Monetary Economics*, 1983, *12*, 55–93.

— **and** — , "Optimal growth with many consumers," *Journal of Economic Theory*, 1984, *32* (1), 139 – 171.

Phelan, Christopher and Ennio Stacchetti, "Sequential Equilibria in a Ramsey Tax Model," *Econometrica*, 2001, *69* (6), 1491–1518.

Piketty, Thomas and Emmanuel Saez, "A Theory of Optimal Inheritance Taxation," *Econometrica*, 2013, *81* (5), 1851–1886.

Reinhorn, Leslie J, "On optimal redistributive capital taxation," *mimeo*, 2002 and 2013.

Saez, Emmanuel, "Optimal progressive capital income taxes in the infinite horizon model," *Journal of Public Economics*, 2013, *97*, 61–74.

Uzawa, Hirofumi, "Time preference, the consumption function, and optimum asset holdings," *Value, Capital, and Growth: Papers in Honour of Sir John Hicks (University of Edinburgh Press, Edinburgh)*, 1968, pp. 485–504.

Appendix

A Proof of Proposition 3

The proof of Proposition 3 consists of three parts. In the first part, we provide a few definitions that are necessary for the proof. In particular, we define the *feasible* set of states. In the second part, we characterize the feasible set of states geometrically. The proofs for the results are somewhat cumbersome and lengthy, so they are relegated to the end of this section to ensure greater readability. Finally, in the third part, we use these results to prove Proposition 3. Readers interested only in the main steps of the proof are advised to jump straight to the third part.

A.1 Definitions

For the proof of Proposition 3 we make a number of definitions, designed to simplify the exposition. A state (k, C_-) as in the recursive statement (4) of problem (1a) will sometimes be abbreviated by z, and a set of states by Z. The total state space is denoted by Z_{all}. It will prove useful at times to express the set of constraints in (4) as

$$k' = x - C_- \left(\frac{\beta x}{k}\right)^{1/\sigma} \tag{17a}$$

$$C = C_- \left(\frac{\beta x}{k}\right)^{1/\sigma} \tag{17b}$$

$$C_-^{\sigma/(\sigma-1)} \left(\frac{\beta}{k}\right)^{1/(\sigma-1)} \leq x \leq f(k) + (1-\delta)k - g, \tag{17c}$$

where $x = k' + C$ replaces c as control. In the last equation, the first inequality ensures non-negativity of k' while the second inequality is merely the resource constraint. Substituting out x, we can also write the law of motion for capital as $k' = \frac{1}{\beta}\frac{k}{C_-^{\sigma}} C^{\sigma} - C$, which we will be using below.

The whole set of future states z' which can follow a given state $z = (k, C_-)$ is denoted by $\Gamma(z)$, which can be the empty set. We will call a path $\{z_t\}$ *feasible* if (a) $z_{t+1} \in \Gamma(z_t)$, which precludes $\Gamma(z_t)$ from being empty; and (b) if the transversality condition holds along the path, $C_t^{-\sigma} k_{t+1} \to 0$. Similarly, a state z will be called *feasible*, if there exists a feasible (infinite) path $\{z_t\}$ starting at $z_0 = z$. In this case, z is *generated by* $\{z_t\}$. Because $z_1 \in \Gamma(z)$, we also say z is *generated by* z_1. A *steady state* $z = (k, C_-) \in Z$ is defined to be a state with $C_- = (1-\beta)/\beta k$. Not all steady states turn out to be feasible, hence a *feasible steady state* z is a steady state that is *self-generating*, $z \in \Gamma(z)$. Denote by Z^* the set of all feasible states. An integral part of the proof will be to characterize Z^*.

It will be important to specify between which capital stocks the economy is moving. For this purpose, define k_g and $k^g > k_g$ to be the two roots to the equation

$$k = \underbrace{f(k) + (1-\delta)k - g}_{\equiv F(k)} - \frac{1-\beta}{\beta} k. \tag{18}$$

Demanding that $k^g > k_g$ is tantamount to specifying $F'(k^g) < 1/\beta < F'(k_g)$. Equation (18) was derived from the resource constraint, demanding that capitalists' consumption is at the steady state level of $C = \frac{1-\beta}{\beta} k$ and workers' consumption is equal to zero. Equation (18) need not have two solutions, not even a single one, in which case government consumption is unsustainably high for *any* capital stock. Such values for g are uninteresting and therefore ruled out. Corresponding to k_g and k^g, we define $C_g \equiv (1-\beta)/\beta\, k_g$ and $C^g \equiv (1-\beta)/\beta\, k^g$ as the respective steady state consumption of capitalists. The steady states (k_g, C_g) and (k^g, C^g) represent the lowest and highest feasible steady states, respectively. The reason for this is that the steady state resource constraint (18) is violated for any $k \notin [k_g, k^g]$.

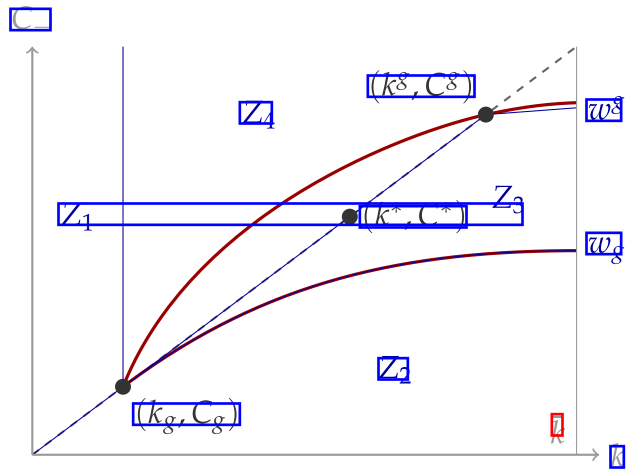

Figure 5: The state space Z_{all}, including the feasible set Z^* (between the two red curves), and all sets Z_i (separated by the blue curves). The point (k^*, C^*) is the zero-tax steady state. Showing that this is the qualitative shape of the feasible set Z^* is an integral part of the proof.

As in the Neoclassical Growth Model, the set of feasible states of this model is easily seen to allow for arbitrarily large capital stocks. This is why we cap the state space for high values of capital, and we take the total state space to be $Z_{\text{all}} = [0, \bar{k}] \times \mathbb{R}_+$ for states (k, C_-), where $\bar{k} \equiv \max\{k_{\max}, k_0\}$ and $k = k_{\max}$ solves $k = f(k) + (1-\delta)k - g$. This way, the set of capital stocks that are resource feasible given an initial capital stock of k_0 must necessarily lie in the interval $[0, \bar{k}]$, so the restriction for $\bar{k}$ is without loss of generality for any given initial capital stock k_0. Note that with this state space, the set of feasible states Z^* is also capped at $\bar{k}$ in its k-component.

The regularity conditions we impose for this proof are Inada conditions on f and u,[31] $f'(0) = \infty, f'(\infty) = 0$ and $u'(0) = \infty, u'(\infty) = 0$.

The outline of this proof is as follows. In section A.2 we characterize the geometry of the set of feasible states Z^*. The results derived there are essential for the actual proof of Proposition 3 in section A.3.

A.2 Geometry of Z^*

For better guidance through this section, we refer the reader to figure 5, which shows the typical shape of Z^*. The main results in this section are characterizations of the bottom and top boundaries of Z^*. We proceed by splitting up the state space, $Z_{\text{all}} = [0, \bar{k}] \times \mathbb{R}_+$, into four pieces and characterizing the feasible states in each of the four pieces.

Define

$$w_g(k) \equiv \begin{cases} \frac{1-\beta}{\beta} k & \text{for } 0 \le k \le k_g \\ C_g \left(\frac{k}{k_g}\right)^{1/\sigma} & \text{for } k_g \le k \le \bar{k} \end{cases}$$

[31] In addition we make standard assumptions throughout, such as $f(0) = 0$, and f, u strictly increasing and concave.

$$w^g(k) \equiv \begin{cases} \frac{1-\beta}{\beta}k & \text{for } 0 \le k \le k^g \\ C^g \left(\frac{k}{k^g}\right)^{1/\sigma} & \text{for } k^g \le k \le \bar{k}, \end{cases}$$

and split up the state space as follows (see figure 5)

$$Z_{\text{all}} \equiv \underbrace{\left\{k < k_g, C_- \ge \frac{1-\beta}{\beta}k\right\}}_{Z_1} \cup \underbrace{\{C_- < w_g(k)\}}_{Z_2} \\ \cup \underbrace{\{k \ge k_g, w_g(k) \le C_- \le w^g(k)\}}_{Z_3} \cup \underbrace{\{k \ge k_g, C_- \ge w^g(k)\}}_{Z_4}.$$

Lemma 1 characterizes the feasible states in sets Z_1 and Z_2.

Lemma 1. *$Z^* \cap Z_1 = Z^* \cap Z_2 = \emptyset$. All states with $k < k_g$ or $C_- < w_g(k)$ are infeasible.*

Proof. See subsection A.4.1. □

In particular, Lemma 1 showed that all states with $C_- < w_g(k)$ are infeasible. Lemma 2 below complements this result stating that all states with $w_g(k) \le C_- \le w^g(k)$ (and $k \ge k_g$) in fact are feasible, that is, lie in Z^*. This means, $\{C_- = w_g(k), k \ge k_g\}$ constitutes the lower boundary of the feasible set Z^*.

Lemma 2. *$Z_3 \subseteq Z^*$, or equivalently, all states with $w_g(k) \le C_- \le w^g(k)$ and $k \ge k_g$ are feasible. Moreover, states on the boundary $\{C_- = w_g(k), k \ge k_g\}$ can only be generated by a single feasible state, (k_g, C_g). Thus, there is only a single "feasible" control for those states, $c \ge 0$.*

Proof. See subsection A.4.2. □

Lemma 2 finishes the characterization of all feasible states with $C_- \le w^g(k)$. What remains is a characterization of feasible states with $C_- > w^g(k)$, or in terms of the $k - C_-$ diagram of figure 5, the characterization of the red top boundary. This boundary is inherently more difficult than the bottom boundary because it involves states that are not merely one step away from a steady state. Rather, paths might not reach a steady state at all in finite time. The goal of the next set of lemmas is an iterative construction to show that the boundary takes the form of an increasing function $w(k)$ such that states with $C_- > w^g(k)$ are feasible if and only if $C_- \le \bar{w}(k)$.

For this purpose, we need to make a number of new definitions: Let $\psi(k, C_-) \equiv (k + C_-)/C^\sigma$. Applying the ψ function to the successor (k', C) of a state (k, C_-) and using the IC constraint (1c) gives $\psi(k', C) = \beta^{-1}k/C_-^\sigma$, a number that is *independent of the control* x. Hence, for every state (k, C_-) there exists an iso-ψ curve containing all its potential successor states.

In some situations it will be convenient to abbreviate the laws of motion for capitalists' consumption and capital, equations (17a) and (17b), as $k'(x, k, C_-)$ and $C(x, k, C_-)$.

Finally, define an operator T on the space of continuous, increasing functions $v : [k_g, \bar{k}] \to \mathbb{R}_+$, as,

$$Tv(k) = \sup\{C_- \mid \exists x \in (0, F(k)] : v(k'(x, k, C_-)) \ge C(x, k, C_-)\}. \tag{19}$$

The operator is designed to extend a candidate top boundary of the set of feasible states by one iteration. To make this formal, let $Z^{(i)}$ be the set of states with $C_- \geq w^g(k)$ which are i steps away from reaching $C_- = w^g(k)$. For example, $Z^{(0)} = \{C_- = w^g(k)\}$. Lemma 3 proves some basic properties of the operator T.

Lemma 3. *T maps the space of continuous, strictly increasing functions $v : [k_g, \bar{k}] \to \mathbb{R}_+$ with $\psi(k, v(k))$ strictly decreasing in k and $v(k_g) = C_g$, $v(k^g) = C^g$, into itself.*

Proof. See subsection A.4.3. □

Lemma 4 uses the operator T to describe the sets $Z^{(i)}$.

Lemma 4. $Z^{(i)} = \{w^g(k) \leq C_- \leq T^i w^g(k)\}$. *In particular $T^i w^g(k) \geq T^j w^g(k) \geq w^g(k)$ for $i \geq j$.*

Proof. See subsection A.4.4. □

The next lemma characterizes the limit function $\bar{w}(k)$ whose graph will describe the top boundary of the set of feasible states.

Lemma 5. *There exists a continuous limit function $\bar{w}(k) \equiv \lim_{i\to\infty} T^i w^g(k)$, with $\bar{w}(k_g) = C_g$ and $\bar{w}(k^g) = C^g$. All states with $C_- = \bar{w}(k)$ are feasible, but only with policy $c = 0$.*

Proof. See subsection A.4.5. □

Lemma 6. *No state with $C_- > \bar{w}(k)$ (and $k_g \leq k \leq \bar{k}$) is feasible.*

Proof. See subsection A.4.6. □

Finally, Lemma 7 shows an auxiliary result which is both used in the proof of Lemma 6 and in Lemma 9 below.

Lemma 7. *Let (k_{t+1}, C_t) be a path starting at (k_0, C_{-1}) with controls $c_t = 0$. Let $k_g < k_0 \leq \bar{k}$. Then:*

(a) If $C_{-1} = \bar{w}(k_0)$, $(k_{t+1}, C_t) \to (k^g, C^g)$.

(b) If $C_{-1} > \bar{w}(k_0)$, $(k_{t+1}, C_t) \not\to (k^g, C^g)$.

Proof. See subsection A.4.7. □

A.3 Proof of Proposition 3

Armed with the results from section A.2 we now prove Proposition 3 in a series of intermediate results. For all statements in this section, we consider an economy with an initial capital stock of $k_0 \in [k_g, \bar{k}]$. We call a path (k_{t+1}, C_t) *globally optimal path*, if initial C_{-1} was optimized over given the initial capital stock k_0. Analogously, we call a path (k_{t+1}, C_t) *locally optimal path*, if initial C_{-1} was not optimized over but rather taken as given at a certain level, respecting the constraint that (k_0, C_{-1}) be feasible.

The first lemma proves that the multiplier on capitalists' IC constraint explodes along an optimal path, and at the same time, workers' consumption drops to zero.

Lemma 8. *Along an optimal path, $\mu_t \geq 0$, $\mu_t \to \infty$ and $c_t \to 0$, where μ_t, c_t are as in problem (1a).*

Proof. Consider the law of motion for μ_t,

$$\mu_{t+1} = \mu_t \left(\frac{\sigma - 1}{\sigma \kappa_{t+1}} + 1 \right) + \frac{1}{\beta \sigma \kappa_{t+1} v_t}.$$

From section A.2 we know that $\kappa_{t+1} = k_{t+1}/C_t$ is bounded away from ∞. Since $\mu_0 = 0$ and $\sigma > 1$, it follows that $\mu_t \geq 0$ and $\mu_t \to \infty$.

Suppose $c_t \not\to 0$. In this case, there exists $\underline{c} > 0$ and an infinite sequence of indices (t_s) such that $c_{t_s+1} \geq \underline{c}$ and $c_{t_s+1} \geq c_{t_s}$ for all s. Along these indices, the FOC for capital implies

$$\underbrace{u'_{t_s+1}}_{< u'(\underline{c})} (f'_{t_s+1} + (1-\delta)) = \frac{1}{\beta} u'_{t_s} + \underbrace{U'_{t_s}}_{\text{bounded}} \cdot \underbrace{(\mu_{t+1} - \mu_t)}_{\geq \text{const} \cdot \mu_t \to \infty},$$

and so $k_{t_s+1} \to 0$ for $s \to \infty$, which is impossible within a s-g set because it violates $k \geq k_g$. $\square$

Lemma 8 is mainly important because it shows that workers' consumption drops to zero. Together with the following lemma, this gives us a crucial geometric restriction of where an optimal path goes in the long run.

Lemma 9. *In the interior of Z^*, the optimal control policy is always $c > 0$. It follows that a globally optimal path approaches either (k_g, C_g) or (k^g, C^g).*

Proof. Note that any point in the interior of Z^* is element of some $Z^{(i)}$, $i < \infty$, and can thus reach the set $\{C_- \leq w^g(k)\} \setminus \{(k_g, C_g), (k^g, C^g)\}$ in finite time. From there, at most two steps are necessary to reach a interior steady state (k_{ss}, C_{ss}) with $k_g < k_{ss} < k^g$ and hence positive consumption $c_{ss} > 0$. By continuity, such an interior steady state can be reached without leaving the interior of the feasible set.

Now take an interior state (k, C_-) and suppose the optimal control was $c = 0$. Then, by the Euler equation[32], it would have to stay at zero for the whole (locally) optimal path, and so the value of this path would be $u(0)/(1-\beta)$. Clearly, this is less than the value of a path converging to an interior steady state with positive workers' consumption along the whole path.

We conclude that the set $\{c = 0\}$ lies in the boundary of Z^*. By Lemmas 2 and 5 this means that $\{c = 0\}$ is exactly equal to the top boundary $\{C_- = w(k)\}$. A globally optimal path which approaches $\{c = 0\}$ must then share the same limiting behavior as states in the set $\{c = 0\}$. By virtue of Lemma 7, it must either converge to (k_g, C_g) or (k^g, C^g). $\square$

Lemma 9 gives a sharp prediction for the behavior of an optimal path: It converges to one of two $c = 0$ steady states: One with little capital or one with abundant capital. Which one that is, will be clear from the next lemma.

[32] Note that the $c \geq 0$ restriction need not be imposed due to Inada conditions for u.

Lemma 10. *If a locally optimal path (k_{t+1}, C_t) converges to (k^g, C^g), then the value function V is locally decreasing in C at point (k_{t+1}, C_t) for all $t > T$ with T large enough.*

Proof. Let $x_t \equiv F(k_t) - c_t$ and consider the following variation: Suppose that at a point T, (k_{T+1}, C_T) is not at the lower boundary (in which case it cannot converge to (k^g, C^g) anyway) and that $c_t < F(k_t) - F'(k_t)k_t$ for all $t \geq T$.[33] For simplicity, call this $T = -1$. Do the perturbation $\hat{C}_{-1} \equiv C_{-1} - \epsilon$, $\hat{k}_0 = k_0$, but keep the controls c_t at their optimal level for (k_0, C_{-1}), that is $\hat{c}_t = c_t$. Denote the perturbed capital stock and capitalists' consumption by $\hat{k}_{t+1} = k_{t+1} + dk_{t+1}$ and $\hat{C}_t = C_t + dC_t$. Then the control x changes by $dx_t = F'_t dk_t$ to first order. We want to show that $dk_{t+1} > 0$ and $dC_t < 0$ for all $t \geq 0$, knowing that $dC_{-1} = -\epsilon$ and $dk_0 = 0$.

From the constraints we find,

$$dk_{t+1} = \underbrace{F'(k_t)dk_t}_{\geq 0} \underbrace{- \frac{C_t}{C_{t-1}} dC_{t-1}}_{\geq 0} + \underbrace{\frac{1}{\sigma} \frac{C_t}{x_t} \frac{F(k_t) - F'(k_t)k_t - c_t}{k_t} dk_t}_{\geq 0} > 0$$

$$dC_t = \underbrace{\frac{C_t}{C_{t-1}} dC_{t-1}}_{\leq 0} \underbrace{- \frac{1}{\sigma} \frac{C_t}{x_t} \frac{F(k_t) - F'(k_t)k_t - c_t}{k_t} dk_t}_{\leq 0} < 0.$$

Using matrix notation, this local law of motion can be written as

$$\begin{pmatrix} dk_{t+1} \\ dC_t \end{pmatrix} = \begin{pmatrix} a_t + b_t & -d_t \\ -b_t & d_t \end{pmatrix} \begin{pmatrix} dk_t \\ dC_{t-1} \end{pmatrix},$$

with $a_t = F'(k_t)$, $d_t = C_t / C_{t-1}$, $b_t = \frac{1}{\sigma} \frac{C_t}{x_t} \frac{F(k_t) - F'(k_t)k_t - c_t}{k_t}$. Close to (k^g, C^g), this matrix has $d \approx 1$. Suppose for one moment that a was zero; the fact that $a > 0$ only works in favor of the following argument. With $a = 0$, the matrix has a single nontrivial eigenvalue of $b + d$, which exceeds 1 strictly in the limit, and the associated eigenspace is spanned by $(1, -1)$. The trivial eigenvalue's eigenspace is spanned by (d, b). Notice that the latter eigenvector is not collinear with the initial perturbation $(0, -1)$, implying that $dk_\infty > 0$ and $dC_\infty < 0$. Hence, $\hat{k}_\infty > k_\infty = k^g$ and $\hat{C}_\infty < C_\infty = C^g$.

But notice that to the bottom right of (k^g, C^g), the new point is interior, which implies a continuation value of $u(0)/(1-\beta)$. More formally, this means there must exist a time $T' > 0$ for which the continuation value of $(k_{T'+1}, C_{T'})$ is strictly dominated by the one for $(\hat{k}_{T'+1}, \hat{C}_{T'})$, that is, $V(k_{T'+1}, C_{T'}) < V(\hat{k}_{T'+1}, \hat{C}_{T'})$. Because all controls were equal up until time T', this implies that $V(k_{T+1}, C_T) < V(k_{T+1}, C_T - \epsilon)$ for ϵ small (Recall that we had set $T = -1$ during the proof). Thus, the value function must increase if C_T is lowered, for a path starting at (k_{T+1}, C_T), for large enough T. This proves that the value function is locally decreasing in C at that point. $\square$

And finally, Lemma 11 proves Proposition 3.

[33] Such a finite $T > 0$ exists for two reasons: (a) because $c_t \to 0$; and (b) because $F(k) - F'(k)k$ which is positive in a neighborhood around $k = k^g$ since k^g was defined by $F(k^g) = k^g/\beta$ and $F'(k^g) < 1/\beta$.

Lemma 11. *A (globally) optimal path converges to (k_g, C_g).*

Proof. By Lemma 9 it is sufficient to prove that a globally optimal path does not converge to (k^g, C^g). Suppose the contrary held and there was a globally optimal path converging to (k^g, C^g). By Lemma 10, this means that the value function is locally decreasing around the optimal path (k_{t+1}, C_t) for $t \geq T$, with $T > 0$ sufficiently large. Consider the following feasible variation for $t = -1, 0, \ldots, T$, $\hat{C}_t = C_t(1 - d\epsilon_t)$, $\hat{k}_{t+1} = k_{t+1}$, $\hat{x}_t = x_t - C_t d\epsilon_t$ where

$$d\epsilon_t \equiv \left(1 - \frac{1}{\sigma}\frac{C_t}{x_t}\right)^{-1} d\epsilon_{t-1}. \tag{20}$$

Observe that (20) is precisely the relation which ensures that the variation satisfies all the constraints of the system (in particular (17b) of which (20) is the linearized version). Workers' consumption increases with this variation by$dc_t = C_t d\epsilon_t > 0$. Therefore, the value of this path changes by

$$dV = \underbrace{\sum_{t=0}^{T} \beta^t u'(c_t) dc_t}_{>0} + \beta^{t+1} \underbrace{\left(V(k_{t+1}, C_t - C_t d\epsilon_t) - V(k_{t+1}, C_t)\right)}_{>0, \text{ by Lemma 10}} > 0,$$

which is contradicting the optimality of (k_{t+1}, C_t). Ergo, a globally optimal path converges to (k_g, C_g). □

A.4 Proofs of Auxiliary Lemmas

A.4.1 Proof of Lemma 1

Proof. Focus on Z_1 first and consider a state (k, C_-) with $k < k_g$ and $C_- \geq \frac{1-\beta}{\beta}k$. Then, $x \equiv k' + C \leq f(k) + (1-\delta)k - g = \frac{1}{\beta}k(1 - \epsilon(k))$, where $\epsilon(k_g) = 0$ and $\epsilon(k) > 0$ for all $k < k_g$, by definition of k_g and Inada conditions for f. Also, by $g > 0$ there is a capital stock $\tilde{k} \in (0, k_g)$ where $\epsilon(\tilde{k}) = 1$ (namely when $f(\tilde{k}) + (1-\delta)\tilde{k} - g = 0$). The highest k' which can generate a state $(k, C_-) \in Z_1$ is then bounded by

$$k' = x - C_-\left(\frac{\beta x}{k}\right)^{-1/\sigma} \leq \frac{1}{\beta}k(1 - \epsilon(k)) - \frac{1-\beta}{\beta} k \underbrace{(1 - \epsilon(k))^{1/\sigma}}_{\geq 1 - \epsilon(k)} \leq k(1 - \epsilon(k)),$$

where in the first inequality we used the fact that k' is increasing in x in the relevant range for x, as specified in (17c). This implies that k always strictly falls in that range, and after a finite number of periods crosses $\tilde{k}$. For $k < \tilde{k}$, the constraint set is empty because then $f(k) + (1-\delta)k - g < 0$. Therefore, no state in $\left\{k < k_g, C_- \geq \frac{1-\beta}{\beta}k\right\}$ can be generated by an infinite path and so $Z^* \cap Z_1 = \emptyset$.

Now consider a state (k, C_-) with $C_- < w_g(k)$, thus, in particular $(C_-/C_g)^\sigma < k/k_g$.[34]

[34] This inequality even holds if $k < k_g$ because there, $C_g(k/k_g)^{1/\sigma} > (1-\beta)/\beta k$. To see this recall that

Define $h(k, C_-) \equiv k/C_-^\sigma$. Suppose next period's state satisfies $C < \frac{1-\beta}{\beta}k'$, or else $C \geq \frac{1-\beta}{\beta}k'$ which we already know leads to an empty constraint set in finite time.[35] Then,

$$h(k', C) = \frac{k'}{C^\sigma} = \frac{k'}{C^\sigma \beta x/k} = \frac{k}{C^\sigma} \underbrace{\frac{k'}{\beta(k'+C)}}_{>1} > h(k, C_-). \tag{21}$$

This implies that, along any feasible path, h is strictly increasing. Suppose $h \not\to \infty$. Then, by the monotone sequence convergence theorem, there exists an $H > 0$ such that $h \to H$ along the path. Using (21) this implies that $k_{t+1}/\left(\beta(k_{t+1} + C_t)\right) \to 1$, or equivalently that $k_{t+1}/C_t \to \beta/(1-\beta)$. If $k_{t+1} \not\to 0$ (in the case $k_{t+1} \to 0$ we are done because for any $k < \tilde{k}$ the constraint set is empty, as before), then this means the state (k_t, C_{t-1}) converges to a feasible steady state.[36] However, the lowest feasible steady state is (k_g, C_g) and since $(C_-/C_g)^\sigma < k/k_g$,

$$h > h(k_g, C_g) = \sup_{k_g < k < k^g} h(k, (1-\beta)/\beta k),$$

which follows because $k/((1-\beta)/\beta k)^\sigma$ is decreasing in k. This is a contradiction to (k_t, C_{t-1}) converging to a feasible steady state. Therefore, $h \to \infty$, and thus $C_t \to 0$ because k is bounded from above by the resource constraint. Again, if k_{t+1} eventually drops below $\tilde{k}$, the constraint set is empty. Assume $k_{t+1} \geq \underline{k}$ for some $\underline{k} > 0$. Then, $U'(C_t)k_{t+1} \to \infty$, contradicting the transversality condition. We conclude that no state (k, C_-) with $(C_-/C_g)^\sigma < k/k_g$ can be generated by an infinite path satisfying the necessary constraints. Hence, $Z_2^* = \emptyset$. □

A.4.2 Proof of Lemma 2

Proof. Consider a state (k, C_-) with $w_g(k) \leq C_- \leq w^g(k)$ and $k \geq k_g$. In particular, $C_- \leq (1-\beta)/\beta k$, $(C_-/C_g)^\sigma \geq k/k_g$ and $(C_-/C^g)^\sigma \leq k/k^g$.[37] The idea of the proof is to show that in fact such a state can be generated by a steady state (k_{ss}, C_{ss}) (with $C_{ss} = (1-\beta)/\beta k_{ss}$ and $k_g \leq k_{ss} \leq k^g$). By definition of k_g and k^g, such a steady state is always self-generating.

Guess that the right steady state has $k_{ss} = (\beta C_-/(1-\beta))^{\sigma/(\sigma-1)} k^{-1/(\sigma-1)}$ and $C_{ss} = (1-\beta)/\beta k_{ss}$. It is straightforward to check that this steady state can be attained with control $x = (C_{ss}/C_-)^\sigma k/\beta$. This steady state is self-generating because $k_g \leq k_{ss} \leq k^g$, which follows from $(C_-/C_g)^\sigma \geq k/k_g$ and $(C_-/C^g)^\sigma \leq k/k^g$. Finally, the control x is

---

$C_g = (1-\beta)/\beta k_g$ and so $C_g(k/k_g)^{1/\sigma}/((1-\beta)/\beta k) = (k/k_g)^{1/\sigma - 1} > 1$, where we used $\sigma > 1$.

[35] Note that if $C \geq (1-\beta)/\beta k'$, then $k' < k_g$. The reason is as follows: The constraints (17a) and (17b) can be rewritten as $k' = (C/C_-)^\sigma k/\beta - C$. Because $(C_-/C_g)^\sigma < k/k_g$, this implies that $k' > \left(C/C_g\right)^\sigma k_g/\beta - C$. Note that the right hand side of this inequality is increasing in C as long as it is positive (which is the only interesting case here). Substituting in $C > (1-\beta)/\beta k'$, this gives $k' > \left(k'/k_g\right)^\sigma k_g/\beta - (1-\beta)/\beta k'$. Rearranging, $k'/k_g > \left(k'/k_g\right)^\sigma$, a condition which can only be satisfied if $k'/k_g < 1$ (recall that $\sigma > 1$).

[36] Notice that, if $k_{t+1}/C_t^\sigma \to H > 0$ and $k_{t+1}/C_t \to \beta/(1-\beta)$ then convergence of k_{t+1} and C_{t+1} themselves immediately follow.

[37] These inequalities hold for all $k \geq k_g$. The proofs are analogous to the proofs in footnotes 34 and 38.

resource-feasible because $C_- \leq (1-\beta)/\beta k$ and thus,

$$x = \frac{1}{\beta}\left[\frac{\left(\frac{\beta}{1-\beta}C_-\right)^{\sigma}}{k}\right]^{1/(\sigma-1)} \leq \frac{k}{\beta} \leq f(k) + (1-\delta)k - g,$$

where the latter inequality follows from the fact that $k_g \leq k \leq k^g$ and the definition of k_g and k^g. This concludes the proof that all states with $w_g(k) \leq C_- \leq w^g(k)$ and $k \geq k_g$ are feasible.

Now regard a state on the boundary $\{C_- = w_g(k), k > k_g\}$, so we also have that $C_- < (1-\beta)/\beta k$.[38] For such a state, $k_{ss} = k_g$ and $C_{ss} = C_g$, and so such a state is generated by (k_g, C_g). Moreover, the unique control which moves (k, C_-) to (k_g, C_g) is $x < k/\beta \leq f(k) + (1-\delta)k - g$, or in terms of c, $c > 0$.

To show that (k_g, C_g) is in fact the only feasible state generating (k, C_-), let (k', C) be a state generating (k, C_-). If $k' < k_g$, then (k', C) is not feasible by Lemma 1, and $k' = k_g$ is exactly the case where (k_g, C_g) generates (k, C_-). Suppose $k' > k_g$. Then, $C < (1-\beta)/\beta k'$,[39] and so we can recycle equation (21) to see $h(k', C) > h(k, C_-)$. Because $h(k, C_-) = h(k_g, C_g)$ however, this implies that $h(k', C) > h(k_g, C_g)$, or put differently, $C < w_g(k')$. Again by Lemma 1 such a (k', C) is not feasible. Therefore, the only state that can generate a state on the boundary $\{C_- = w_g(k), k > k_g\}$ is (k_g, C_g), and the associated unique control involves positive c. □

A.4.3 Proof of Lemma 3

Proof. First note that T can be rewritten as

$$Tv(k) = \max\{C_- \mid v(k'(F(k), k, C_-)) = C(F(k), k, C_-)\}. \tag{22}$$

There are two ways in which (22) differs from (19):

- Suppose that the supremum in (19) is attained with $x_0 < F(k)$. Because $\psi(k, v(k))$ is strictly decreasing in k and $\psi(k'(\ldots), C(\ldots))$ is constant in x, there can at most be a single crossing between the graph of v and $\{(k', C) \mid x \in (0, F(k)]\}$. Further, notice that the function

$$\Phi : x \mapsto \underbrace{\psi\left(k'(x, k, C_-), C(x, k, C_-)\right)}_{\text{constant in } x} - \underbrace{\psi\left(k'(x, k, C_-), v(k'(x, k, C_-))\right)}_{\text{decreasing in } x} \tag{23}$$

  is strictly increasing in x with $\Phi(x_0) \geq 0$. Therefore, $\Phi(F(k)) > 0$, or, in other words, $v(k'(F(k), k, C_-)) > C(F(k), k, C_-)$. Since v is continuous, this means that

[38] This holds because $C_- = w_g(k) = C_g(k/k_g)^{1/\sigma}$ and thus $C_-/((1-\beta)/\beta k) = (k/k_g)^{1/\sigma - 1} < 1$.

[39] This holds because by the IC constraint (??), $\beta(k' + C)/C^{\sigma} = k_g/C_g^{\sigma}$ or equivalently $(k' + C)/C = 1/(1-\beta)\,(C/C_g)^{\sigma}$. Thus, letting $\kappa = k'/C$, $(\kappa+1)\kappa^{\sigma} = (1-\beta)^{-1} \cdot (\beta/(1-\beta))^{\sigma} \cdot (k'/k_g)^{\sigma}$. Since the right hand side is increasing in κ, the fact that $k' > k_g$ tells us that $\kappa > \beta/(1-\beta)$, which is what we set out to show.

C_- can be increased without violating $v(k') \geq C$ — a contradiction to C_- attaining the supremum.

- Suppose that the supremum in (19) is attained with $x = F(k)$ but $v(k'(F(k), k, C_-)) > C(F(k), k, C_-)$. Again, this means increasing C_- does not violate the condition that $v(k'(x, k, C_-)) \geq C(x, k, C_-)$.

These two arguments prove that (22) is a valid way to write $Tv(k)$. Now pick a continuous, increasing function $v : [k_g, \bar{k}] \to \mathbb{R}_+$ with $\psi(k, v(k))$ strictly decreasing in k and $v(k_g) = C_g$, $v(k^g) = C^g$ and check the claimed properties in turn:

- $Tv(k_g) = C_g$ because $k'(F(k_g), k_g, C_g) = k_g$ and $C(F(k_g), k_g, C_g) = C_g$. However, $k'(F(k_g), k_g, C_-)$ is strictly decreasing in C_- and so $k'(F(k_g), k_g, C_-) < k_g$ for $C_- > C_g$ (for $k < k_g$, $v(k)$ is not even defined).

- Note that $v(k'(F(k), k, C_-)) = C(F(k), k, C_-)$ has exactly one solution $C^*(k)$ for C_- since $v(k')$ is increasing in k' but k' is strictly decreasing in C_- and C strictly increasing in C_-. Also, it is easy to see that for $C_- < C_g(k/(\beta F(k)))^{1/\sigma}$, $C(F(k), k, C_-) < C_g$ and so $v(k'(F(k), k, C_-)) > C(F(k), k, C_-)$. Similarly, for C_- sufficiently high, $k' = k_g$ but $C > C_g$.[40]

- To show that $Tv(k)$ is increasing note that $\psi(k'(\ldots), C(\ldots)) = \beta^{-1}k/C_-^\sigma$ is strictly increasing in k and strictly decreasing in C_-. Further, recall that

$$k'(F(k), k, C_-) = F(k)\left(1 - C_-\left(\frac{\beta}{kF(k)^{\sigma-1}}\right)^{1/\sigma}\right)$$

is strictly increasing in k and strictly decreasing in C_-, and that v was such that $\psi(k, v(k))$ is strictly decreasing in k. Then, the function

$$\Psi : (k, -C_-) \mapsto \underbrace{\psi\left(k'(F(k), k, C_-), C(F(k), k, C_-)\right)}_{\nearrow \text{ in } k \text{ and } (-C_-)} - \underbrace{\psi\left(k'(F(k), k, C_-), v(k'(F(k), k, C_-))\right)}_{\searrow \text{ in } k \text{ and } (-C_-)}$$

is strictly increasing in k and $-C_-$. Because the $\{\Psi = 0\}$ locus is exactly the graph of $Tv(k)$, it follows that $Tv(k)$ is strictly increasing.

- Then, it also easily follows that $Tv(k)$ is continuous because Ψ is strictly increasing and continuous, and has exactly one zero for each value of $k \in [k_g, \bar{k}]$.[41]

[40] $C > C_g$ must hold if $k' = k_g$, and $k > k_g$ because: From $k > k_g$ it follows that $k'(F(k), k, (1-\beta)/\beta k) > k_g$ and $C(F(k), k, (1-\beta)/\beta k) > C_g$. Since $k'(\ldots)$ is decreasing in C_-, it follows that $C_- > (1-\beta)/\beta k$ is necessary to achieve $k' = k_g$. Because $C(\ldots)$ is increasing in C_- it follows that $C > C_g$.

[41] This is a fact that holds more generally: If $f(x, y)$ is a strictly increasing two-dimensional function and for each x there exists a unique $y^*(x)$ s.t. $f(x, y^*(x)) = 0$, then $y^*(x)$ must be continuous in x.

- For $\psi(k, Tv(k))$ decreasing in k, pick $k_1 < k_2$. Suppose $\psi(k_1, Tv(k_1)) \leq \psi(k_2, Tv(k_2))$. Since $Tv(k)$ is strictly increasing, it follows that

$$\frac{k_1}{Tv(k_1)^{\sigma}} - \frac{k_2}{Tv(k_2)^{\sigma}} < \underbrace{\frac{k_1}{Tv(k_1)^{\sigma}} + Tv(k_1)^{1-\sigma}}_{\psi(k_1, Tv(k_1))} \underbrace{- \frac{k_2}{Tv(k_2)^{\sigma}} - Tv(k_2)^{1-\sigma}}_{-\psi(k_2, Tv(k_2))} \leq 0,$$

  and so

$$\psi(k_1', C_1) = \beta^{-1} \frac{k_1}{Tv(k_1)^{\sigma}} < \beta^{-1} \frac{k_2}{Tv(k_2)^{\sigma}} = \psi(k_2', C_2). \tag{24}$$

  This, however, implies that $Tv(k_2)$ cannot have been optimal: Pick an alternative consumption level $C_{2,-}$ as $C_{2,-} = Tv(k_1)(k_2/k_1)^{1/\sigma}$, which exceeds $Tv(k_2)$ by (24). Moreover, denoting by x_1 the policy to take state $(k_1, Tv(k_1))$ to state (k_1', C_1), pick x_1 as alternative policy for $(k_2, C_{2,-})$. Note that x_1 is feasible in state $(k_2, C_{2,-})$ because $x_1 \leq F(k_1) \leq F(k_2)$. Since $k_1/Tv(k_1)^{\sigma} = k_2/C_{2,-}^{\sigma}$ by construction, it follows that the state succeeding $(k_2, C_{2,-})$ is $(k'(x_1, k_2, C_{2,-}), C(x_1, k_2, C_{2,-})) = (k_1', C_1)$, which lies on the graph of v. Hence $Tv(k_2)$ cannot have been optimal and so $\psi(k, Tv(k))$ is decreasing in k.

- Finally, $Tv(k^g) = C^g$. The reason for this is that one the one hand, $k'(F(k^g), k^g, C^g) = k^g$ and $C(F(k^g), k^g, C^g) = C^g$. On the other hand, because $k'(\ldots)$ is decreasing and $C(\ldots)$ is increasing in C_-, it follows that $k'(F(k^g), k^g, C_-) < k^g$ but $C(F(k^g), k^g, C_-) > C^g$ for $C_- > C^g$. Such a state can never lie on the graph of v given $v(k^g) = C^g$ and its monotonicity.

$\square$

A.4.4 Proof of Lemma 4

Proof. Note that any state (k, C_-) reaches the space $\{C_- \leq v(k)\}$ in one step if and only if $C_- \leq Tv(k)$ (provided that v satisfies the regularity properties in Lemma 3). Thus, by iteration, $Z^{(i)} = \{w^g(k) \leq C_- \leq T^i w^g(k)\}$. Because $Z^{(i)} \supseteq Z^{(j)}$ for $i \geq j$, it holds that $T^i w^g(k) \geq T^j w^g(k)$.[42] $\square$

A.4.5 Proof of Lemma 5

Proof. The existence of the limit $\lim_{i \to \infty} T^i w^g(k)$ is straightforward for every k (monotone sequence, bounded above because for large values of C_-, $k'(F(k), k, C_-) < k_g$ for any k). By Lemma 2, w must be weakly increasing, $w(k_g) = C_g$, $w(k^g) = C^g$, and $\psi(k, w(k))$ must be weakly decreasing. Suppose that $\bar{w}$ was not continuous. Then, there need to be two arbitrarily close values of k, $k_1 < k_2$ with a gap between $w(k_1)$ and $w(k_2)$. Because $k'(F(k_1), k_1, C_-)$ is decreasing in C_- and $C(F(k_1), k_1, C_-)$ is increasing in C_-, the

[42] A subtlety here is that $Z^{(i)} \supseteq Z^{(j)}$ only holds because states in the set $\{C_- = w^g(k)\}$ is "self-generating", that is, if a path hits the set $\{C_- = w^g(k)\}$ after j steps, it can stay in that set forever. In particular, it can hit the set after $i \geq j$ steps as well. This explains why $Z^{(i)} \supseteq Z^{(j)}$.

fixed point property $T\bar{w} = \bar{w}$ can only hold if $\bar{w}$ were locally decreasing around state $(k'(F(k_1), k_1, C^*(k_1)), C(F(k_1), k_1, C^*(k_1))$, a contradiction. Therefore, $\bar{w}$ is continuous.

Note that by the fixed point property, $T\bar{w}(k) = \bar{w}(k)$, from which it follows that $c = 0$, or in other words $x = F(k)$, is the only feasible policy for states with $C_- = \bar{w}(k)$ (consider the representation of T in equation (22) – this implies that $v(k') < C$ for any $c > 0$, by a similar logic as in (23)). $\square$

A.4.6 Proof of Lemma 6

Proof. Define h as before, $h(k', C) \equiv k'/C^\sigma$. Fix a state (k, C_-) with $C_- > \bar{w}(k)$. First, consider the case $C_- \geq (1-\beta)/\beta k$. Note that such a path must have $C_t > (1-\beta)/\beta k_{t+1}$ along the whole path unless $k_{t+1} < k_g$. This follows directly from $C_t > \bar{w}(k_{t+1})$ (which must hold by construction of w) for $k_{t+1} \leq k^g$. If $k_{t+1} > k^g$ it must be the case that $k_{s+1} > k^g$ for all $s < t$ as well.[43] But then, using $x_t \leq F(k_t) < k_t/\beta$,

$$\frac{k_{t+1}}{C_t} = \frac{x_t}{C_t} - 1 < \frac{k_t/\beta}{C_{t-1}} - 1 = \frac{\beta}{1-\beta}.$$

We established that $C_t > (1-\beta)/\beta k_{t+1}$ along the whole path.

Thus,

$$h(k_{t+1}, C_t) = \frac{k_{t+1}}{C_t^\sigma} = \frac{k_t}{C_{t-1}^\sigma} \underbrace{\frac{k_{t+1}}{\beta(k_{t+1} + C_t)}}_{<1} < h(k_t, C_{t-1}).$$

If $h(k_{t+1}, C_t)$ converges to zero, then either $k_{t+1} \to 0$ or $C_t \to \infty$ (in which case $k_{t+1} \to 0$ by the law of motion for capital and the fact that $k_t \leq \bar{k}$). Such a path is not feasible because k_{t+1} drops below $\tilde{k}$ in finite time (see proof of Lemma 1 for $\tilde{k}$). Hence, suppose $h(k_{t+1}, C_t) \to \underline{h} > 0$. Then, $k_{t+1}/(\beta(k_{t+1} + C_t)) \to 1$, so the path must approximate the steady state line described by $C_- = (1-\beta)/\beta k$. Because $C_t > \bar{w}(k_{t+1})$ along the path, (k_{t+1}, C_t) must be converging to (k^g, C^g).

Next we show that along this convergence, c_t can be zero. Suppose there were times with $c_t > 0$. Then, define a new path $(\hat{k}_{t+1}, \hat{C}_t)$, starting at the same initial state (k, C_-)

[43] The reason for this is that for any state (k, C_-) with $k \leq k^g$ and $C_- > \bar{w}(k)$ we have that $k' \leq k^g$:

- if $\psi(k', C) \geq \psi(k^g, C^g)$, then the curve $\{(k'(x, k, C_-), C(x, k, C_-)), x > 0\}$ (without resource constraint restriction) and the graph of $\bar{w}$ intersect at a state with capital less than k^g. In particular this implies that the intersection of $\{(k'(x, k, C_-), C(x, k, C_-)), x > 0\}$ and the steady state line $\{C = (1-\beta)/\beta k\}$ lies in the interior of $\{C < \bar{w}(k)\}$. Therefore, if C were smaller than $(1-\beta)/\beta k'$ this would mean that $C < \bar{w}(k')$ – a contradiction to $C_- > \bar{w}(k)$ given the construction of $\bar{w}$
- if $\psi(k', C) = k/C_-^\sigma < \psi(k^g, C^g) = k^g/(C^g)^\sigma$, then $k' \leq F(k) - C_-\left(\frac{\beta F(k)}{k}\right)^{1/\sigma} < F(k^g) - C^g\left(\frac{\beta F(k^g)}{k^g}\right)^{1/\sigma} = k^g$.

but with controls $c_t = 0$. Observe that

$$h(\hat{k}_{t+1}, \hat{C}_t) = \psi(\hat{k}_{t+1}, \hat{C}_t) - \hat{C}_t^{1-\sigma} = \beta^{-1} h(\hat{k}_t, \hat{C}_{t-1}) - h(\hat{k}_t, \hat{C}_{t-1})^{(\sigma-1)/\sigma} (\beta F(\hat{k}_t))^{-(\sigma-1)/\sigma}$$

$$\hat{k}_{t+1} = F(\hat{k}_t) - \left(\frac{\beta F(\hat{k}_t)}{h(\hat{k}_t, \hat{C}_{t-1})} \right)^{-1/\sigma},$$

where the first equation is increasing in $h(\hat{k}_t, \hat{C}_{t-1})$ for the relevant parameters for which $h(\hat{k}_{t+1}, \hat{C}_t) \geq 0$, and similarly the second equation is increasing in $F(\hat{k}_t)$ if $\hat{k}_{t+1} \geq 0$. By induction over t, if $h(\hat{k}_t, \hat{C}_{t-1}) \geq h(k_t, C_{t-1})$ and $\hat{k}_t \geq k_t$ (induction hypothesis), then, because $F(\hat{k}_t) \geq x_t$,

$$h(\hat{k}_{t+1}, \hat{C}_t) \geq \beta^{-1} h(k_t, C_{t-1}) - h(k_t, C_{t-1})^{(\sigma-1)/\sigma} (\beta x_t)^{-(\sigma-1)/\sigma} = h(k_{t+1}, C_t)$$

$$\hat{k}_{t+1} \geq F(k_t) - \left(\frac{\beta F(k_t)}{h(k_t, C_{t-1})} \right)^{-1/\sigma},$$

confirming that $\hat{k}_t \geq k_t$ and $h(\hat{k}_t, \hat{C}_{t-1}) \geq h(k_t, C_{t-1})$ for all t. Given that $h(k_{t+1}, C_t) \to \underline{h} > 0$, either $\left(\hat{k}_{t+1}, \hat{C}_t\right) \to (k^g, C^g)$ as well or $\left(\hat{k}_{t+1}, \hat{C}_t\right)$ converges to some steady state between k_g and k^g. The latter cannot be because of $C_t > \bar{w}(k_{t+1})$ along the path. But the former is precluded by Lemma 7 below.

Now, consider the case $k > k^g$ and $C_- < (1-\beta)/\beta k$. If the succeeding state is above the steady state line, $C \geq (1-\beta)/\beta k'$, the case above applies. Hence, suppose the path stayed below the steady state line forever, i.e. $C_t < (1-\beta)/\beta k_{t+1}$ for all t. In that case,

$$h(k_{t+1}, C_t) = \frac{k_{t+1}}{C_t^\sigma} = \frac{k_t}{C_{t-1}^\sigma} \underbrace{\frac{k_{t+1}}{\beta(k_{t+1} + C_t)}}_{>1} > h(k_t, C_{t-1}).$$

Note that $h(k_{t+1}, C_t)$ is bounded from above, for example by $h(k_g, C_g)$ (because all states below the steady state line with h equal to $h(k_g, C_g)$ are below the graph of w^g and thus below $\bar{w}$ as well). So, $h(k_{t+1}, C_t)$ converges and $k_{t+1}/(\beta(k_{t+1} + C_t)) \to 1$. The state approximates the steady state line. Because the only feasible steady state with below the steady state line but above the graph of $\bar{w}$ is (k^g, C^g) it follows that $(k_{t+1}, C_t) \to (k^g, C^g)$.

Following the same steps as before, it can be shown that without loss of generality controls c_t can be taken to be zero along the path. By Lemma 7 below this is a contradiction. □

A.4.7 Proof of Lemma 7

Proof. We prove each of the results in turn.

(a) Notice that $c = 0$ takes any state on the graph of $\bar{w}$ to another state on the graph of $\bar{w}$ (because $T\bar{w} = \bar{w}$). Suppose $k_0 < k^g$ (the case $k_0 > k^g$ is analogous). Then, no future capital stock k_{t+1} can exceed k^g. Because if it did, there would have to be a capital stock $k \in (k_g, k^g)$ with $k'(F(k), k, C_-^*(k)) = k^g$, by continuity of $k \mapsto$

$k'(F(k), k, C^*_-(k))$. But this is impossible by definition of k^g.[44] Thus, along the path, $C_t > (1-\beta)/\beta k_{t+1}$ and so $h(k_{t+1}, C_t)$ is decreasing. As $h(k_g, C_g) > h(k, \bar{w}(k))$ for all $k > k_g$,[45] this means $(k_{t+1}, C_t) \to (k^g, C^g)$.

(b) For simplicity, focus on the case $k_0 < k^g$. Again, the case $k_0 > k^g$ is completely analogous. Suppose (k_{t+1}, C_t) was converging to (k^g, C^g). Note that at k^g, $F(k)/k$ is decreasing[46]. Thus, there exists a time $T > 0$ for which the capital stock k_T is sufficiently close to k^g that $F(k)/k$ is decreasing for all k in a neighborhood of k^g which includes $\{k_t\}_{t\geq T}$. Let $\left(\hat{k}_{t+1}, \hat{C}_t\right)$ denote the path with $c_t = 0$, starting from $(k_T, w(k_T))$. Observe that both (k_{t+1}, C_t) and $(\hat{k}_{t+1}, \hat{C}_t)$ have controls $c_t = 0$ here, unlike in the proof of Lemma 6. Denote the zero-control laws of motion for capital and capitalists' consumption by $L_k(k, C_-) \equiv k'(F(k), k, C_-)$ and $L_C(k, C_-) \equiv C(F(k), k, C_-)$. Since $F(k)/k$ is locally decreasing, it follows that $dL_k/dk > 0$, $dL_k/dC_- < 0$ and $dL_C/dk < 0$, $dL_C/dC_- > 0$. This implies that because $C_{T-1} > \bar{w}(k_T)$ (which must hold or else $C_- \leq w(k)$ by construction of w), $C_t > \hat{C}_t$ and $k_{t+1} > \hat{k}_{t+1}$ for all $t \geq T$. Moreover, borrowing from equation (21), we know that

$$h(k_{t+1}, C_t) = h(k_t, C_{t-1}) \left(\frac{1}{\beta} - \left(\frac{1}{h(k_t, C_{t-1})} \right)^{1/\sigma} \frac{1}{(\beta F(k_t))^{1-1/\sigma}} \right),$$

which implies that by induction $h(k_{t+1}, C_t) \leq h(\hat{k}_{t+1}, \hat{C}_t)$, that is,

$$\begin{aligned}
&\log h(k_{t+T}, C_{t+T-1}) \\
&\quad = \log h(k_T, C_{T-1}) + \sum_{s=0}^{t-1} \log \left(\frac{1}{\beta} - \left(\frac{1}{h(k_{T+s}, C_{T+s-1})} \right)^{1/\sigma} \frac{1}{(\beta F(k_{T+s}))^{1-1/\sigma}} \right) \\
&\quad \leq \log h(k_T, C_{T-1}) + \sum_{s=0}^{t-1} \log \left(\frac{1}{\beta} - \left(\frac{1}{h(\hat{k}_{T+s}, \hat{C}_{T+s-1})} \right)^{1/\sigma} \frac{1}{(\beta F(\hat{k}_{T+s}))^{1-1/\sigma}} \right) \\
&\quad = \log h(\hat{k}_{t+T}, \hat{C}_{t+T-1}) + \log h(k_T, C_{T-1}) - \log h(\hat{k}_T, \hat{C}_{T-1}).
\end{aligned}$$

As $t \to \infty$, this equation yields

$$\log h(k^g, C^g) \leq \log h(k^g, C^g) + \underbrace{\log h(k_T, C_{T-1}) - \log h(\hat{k}_T, \hat{C}_{T-1})}_{=-k_T(\hat{C}_{T-1}^{-\sigma} - C_{T-1}^{-\sigma})<0},$$

which is a contradiction. Therefore, $(k_{t+1}, C_t) \not\to (k^g, C^g)$.

$\square$

[44] By definition of k^g, $F(k^g) = k^g + C^g$, and so, $F(k) < k^g + C^g$ for $k < k^g$.

[45] Note that $\bar{w}(k) > w_g(k)$ and $h(k, w_g(k)) = const$, see Lemmas 1 and 2 above.

[46] This holds because $F'(k^g) < 1/\beta$ and $F(k^g) = 1/\beta k^g$.

B Numerical Method

To solve the Bellman equation we must first compute the set A. As is standard, in practice, especially when $\sigma > 1$, we must restrict the range of capital to a closed interval $[\underline{k}, \bar{k}]$ with $\underline{k} > 0$ to avoid $k = 0$. This leads us to seek a subset $A^{\underline{k}} \subset A$. We compute this set numerically as follows.

Start with the set A_0 defined by $C_- = \frac{1-\beta}{\beta} k$ and $k \in [\underline{k}, \bar{k}]$. This set is self generating and thus $A_0 \subset A^{\underline{k}}$. We define an operator that finds all points (k, C_-) for which one can find c, K', C satisfying the constraints of the Bellman equation and $(k', C) \in A_0$. This gives a set A_1 with $A_0 \subset A_1$. Iterating on this procedure we obtain $A_0, A_1, A_2 \dots$ and we stop when the sets do not grow much.

We then solve the Bellman equation by value function iteration. We start with a guess for V_0 that uses a feasible policy to evaluate utility. This ensures that our guess is below the true value function. Iterating on the Bellman equation then leads to a monotone sequence $V_0, V_1, \dots$ and we stop when iteration n yields a V_n that is sufficiently close to V_{n-1}. Our procedure uses a grid that is defined on a transformation of (k, C_-) that maps A into a rectangle. We linearly interpolate between grid points.

The code was programmed in Matlab and executed with parallel 'parfor' commands, to improve speed and allow denser grids, on a cluster of 64-128 workers. Grid density was adjusted until no noticeable difference in the optimal paths were observed.

C Proof of Proposition 4

The problem in continuous time is

$$\max \int_0^\infty e^{-\rho t} (u(c_t) + \gamma U(C_t))\, dt$$

$$\text{s.t. } c_t + C_t + g + \dot{k}_t = f(k_t) - \delta k_t$$

$$\dot{C}_t = \frac{C_t}{\sigma} \left(\frac{f(k_t)}{k_t} - \delta - \frac{c_t}{k_t} - \rho \right).$$

Let p_t and q_t denote the costates corresponding respectively to the states k_t and C_t. The FOCs are,

$$u'_t = p_t c_t + q_t \frac{1}{\sigma} \frac{C_t}{k_t}$$

$$\dot{p}_t = \rho p_t - p_t (f'(k_t) - \delta) + q_t \frac{C_t}{k_t} - q_t \frac{C_t}{k_t} (f'(k_t) - \delta)$$

$$\dot{q}_t = \rho q_t - \gamma U'(C_t) - q_t \frac{1}{\sigma} \left(\frac{f(k_t)}{k_t} - \delta - \frac{c_t}{k_t} - \rho \right).$$

In addition to the FOCs, we require the two transversality conditions to hold,

$$\lim_{t\to\infty} e^{-\rho t} q_t C_t = 0 \quad \text{and} \quad \lim_{t\to\infty} e^{-\rho t} p_t k_t = 0. \tag{25}$$

Denote the 4-dimensional state of this dynamic system by x_t and its unique positive steady state (the zero-tax steady state) by $x^* = (k^*, C^*, p^*, q^*)$. The linearized system is,

$$\dot{x}_t = J(x_t - x^*), \tag{26}$$

where the 4×4 matrix J can be written as

$$J \equiv \begin{pmatrix} A & B \\ C & \rho I - A' \end{pmatrix},$$

with 2×2 matrices

$$A \equiv \begin{pmatrix} \rho + z & -1 - z/\rho \\ \rho z/\sigma & -z/\sigma \end{pmatrix},$$

$$B \equiv \begin{pmatrix} -1/u'' & -\rho/(\sigma u'') \\ -\rho/(\sigma u'') & -\rho^2/(\sigma^2 u'') \end{pmatrix} = B'$$

$$C \equiv \begin{pmatrix} z^2 u'' + \rho q^*(1 - 1/\sigma) f'' - \gamma U' f'' & -zq/(\sigma k^*) \\ -zq/(\sigma k^*) & z^2 u''/\rho^2 \end{pmatrix} = C',$$

where $z \equiv \rho q^*/(\sigma k^* u'')$. Despite J's somewhat cumbersome form, its determinant simplifies to

$$\det J = (1-\sigma) \underbrace{\frac{f'' u'}{u''}}_{>0} \frac{\rho^2}{\sigma^2}, \tag{27}$$

its characteristic polynomial is, $\det(J - \lambda I) = \lambda^4 - c_1\lambda^3 + c_2\lambda^2 - c_3\lambda + c_4$, with $c_1 = \text{trace}(J) = 2\rho$, $c_2 = \rho^2 + \rho z(1-\sigma)/\sigma - f''u'/u''$, $c_3 = \rho(c_2 - \rho^2) = \rho^2 z(1 - \sigma/\sigma - \rho f''u'/u''$, $c_4 = \det J$, and that its eigenvalues can be written as,

$$\lambda_{1-4} = \frac{\rho}{2} \pm \left[\left(\frac{\rho}{2}\right)^2 - \frac{\delta}{2} \pm \frac{1}{2}\left(\delta^2 - 4 \det J\right)^{1/2} \right]^{1/2}, \tag{28}$$

with $\delta = c_2 - \rho^2 = \rho z(1-\sigma)/\sigma - f''u'/u''$. Substituting in the formulas of z and q^*, δ can also be written as,

$$\delta \equiv \frac{\rho}{\sigma} \frac{u' - \gamma U'}{u'' k^*} - \frac{f'' u'}{u''}. \tag{29}$$

In the remainder, let eigenvalues be numbered as follows: λ_1 has $++$, λ_2 has $+-$, λ_3 has $-+$, and λ_4 has $--$. For convenience, define γ^* by $\gamma^* = u'/U'$.

Note that in general, a solution x_t to the linearized FOCs (26) can load on all four eigenvalues. However, taking the two transversality conditions into account, restricts the system to only load on eigenvalues with $\text{Re}(\lambda_i) \leq \rho/2$. In Lemma 12 below, we show that this means the solution loads on eigenvalues λ_3 and λ_4. Let Q be an invertible matrix such that $QJQ^{-1} = \text{diag}(\lambda_1, \ldots, \lambda_4)$. Write

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}.$$

Thus, the initial values for the two multipliers, p_0 and q_0, need to satisfy

$$Q_{11} \begin{pmatrix} k_0 \\ C_0 \end{pmatrix} + Q_{12} \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} = 0.$$

This completely specifies the trajectory of state x_t in the linearized system.

The following lemma proves properties about J's eigenvalues $\{\lambda_i\}$, in particular about λ_3 and λ_4, which are the relevant eigenvalues for the local dynamics of the state.

Lemma 12. *The eigenvalues in (28) can be shown to satisfy the following properties.*

(a) It is always the case that

$$Re\lambda_1 \geq Re\lambda_2 \geq \rho/2 \geq Re\lambda_4 \geq Re\lambda_3.$$

(b) If $\sigma > 1$, then $\det J < 0$, implying that

$$Re\lambda_1 = \lambda_1 > \rho > Re\lambda_2 \geq \rho/2 \geq Re\lambda_4 > 0 > \lambda_3 = Re\lambda_3. \tag{30}$$

In particular, there is a exactly one negative eigenvalue. The system is saddle-path stable.

(c) If $\sigma < 1$ and $\gamma \leq \gamma^*$, then $\det J > 0$ and $\delta < 0$, implying that

$$Re\lambda_1, Re\lambda_2 > \rho > 0 > Re\lambda_4, Re\lambda_3. \tag{31}$$

In particular, there exist exactly two eigenvalues with negative real part. The system is locally stable.

(d) If $\sigma < 1$ and $\gamma > \gamma^*$, the system may either be locally stable, or locally unstable (all eigenvalues having positive real parts).

Proof. We follow the convention that the square root of a complex number a is defined as the *unique* number b that satisfies $b^2 = a$ *and has nonnegative real part* (if $\text{Re}(b) = 0$ we also require $Im(b) \geq 0$). Hence, the set of all square roots of a is given by $\{\pm b\}$. We prove the results in turn.

(a) First, observe the following fact: Given a real number x and a complex number b with nonnegative real part, it holds that $\text{Re}\left(\sqrt{x+b}\right) \geq \text{Re}\left(\sqrt{x-b}\right)$.[47] From

[47] To prove this, let $\bar{b}$ denote the complex conjugate of b and note that $\text{Re}\left(\sqrt{x+b}\right)$ is monotonic in the real number x. Then, $\text{Re}\left(\sqrt{x+b}\right) = \text{Re}\left(\sqrt{x+\bar{b}}\right) = \text{Re}\left(\sqrt{x-b+(\bar{b}+b)}\right) \geq \text{Re}\left(\sqrt{x-b}\right)$ where $\bar{b}+b = 2\text{Re}(b) \geq 0$ and monotonicity are used.

there, it is straightforward to see that $\mathrm{Re}\lambda_1 \geq \mathrm{Re}\lambda_2$ and $\mathrm{Re}\lambda_4 \geq \mathrm{Re}\lambda_3$. Finally $\mathrm{Re}\lambda_2 \geq \rho/2 \geq \mathrm{Re}\lambda_4$ holds according to our convention of square roots having nonnegative real parts.

(b) The negativity of $\det J$ follows immediately from (27). This implies

$$-\frac{\delta}{2} + \frac{1}{2}\left(\delta^2 - 4\det J\right)^{1/2} > 0 > -\frac{\delta}{2} - \frac{1}{2}\left(\delta^2 - 4\det J\right)^{1/2},$$

and so (30) holds, using monotonicity of $\mathrm{Re}\sqrt{x}$ for real numbers x.

(c) The signs of $\det J$ and δ follow immediately from (27) and (29). In this case, $-\delta/2 \pm 1/2\mathrm{Re}\left(\delta^2 - 4\det J\right)^{1/2} > 0$ proving (31).

(d) This is a simple consequence of the fact that if $\det J > 0$, then either $-\delta/2 \pm 1/2\mathrm{Re}\left(\delta^2 - 4\det J\right)^{1/2} > 0$, or $-\delta/2 \pm 1/2\mathrm{Re}\left(\delta^2 - 4\det J\right)^{1/2} < 0$, where under the latter condition the system is locally unstable. □

D Proof of Proposition 5

First, we define the following object,

$$\omega_\tau = \frac{dW_\tau}{dk_{\tau+1}} = \sum_{\tau' \geq \tau+1} \beta^{\tau'-\tau} u'(c_{\tau'})(F'(k_{\tau'}) - R_{\tau'}) \left(\prod_{s=\tau+1}^{\tau'-1} S_{I,s} R_s\right),$$

which corresponds to the welfare response, measured in units of period τ utility, of a change in savings by an infinitesimal unit between periods τ and $\tau+1$. Now consider the effect of a one-time change in the capital tax, effectively changing R_t to $R_t + dR$ in period t. This has three types of effects on total welfare: It changes savings behavior in all periods $\tau < t$ through the effect of R_t on S_τ. It changes capitalists' income in period t through the effect of R_t on $R_t k_t$. And finally it changes workers' income in period t directly through the effect of R_t on $F(k_t) - R_t k_t$. Summing up these three effects, one obtains a total effect of

$$\begin{aligned} dW = \sum_{\tau=0}^{t-1} \beta^{\tau-t} \omega_\tau \underbrace{S_{\tau,R_t} dR}_{\text{change in savings in period } \tau<t} \\ + \omega_t \underbrace{S_{I,t} k_t dR}_{\text{change in savings in period } t} - u'(c_t) \underbrace{k_t dR}_{\text{change in workers' income in period } t}. \end{aligned}$$

The total effect needs to net out to zero along the optimal path, that is,

$$\omega_t S_{I,t} - u'(c_t) = -\frac{1}{k_t} \sum_{\tau=0}^{t-1} \beta^{\tau-t} \omega_\tau S_{\tau,R_t}. \tag{32}$$

Note that by optimization over the initial interest rate R_0, we find the condition

$$\omega_0 S_{I,0} - u'(c_0)k_0 = 0. \tag{33}$$

Due to the recursive nature of (32), if $\omega_\tau > 0$ for $\tau < t$, then

$$\omega_t S_{I,t} - u'(c_t) = -\frac{1}{k_t}\sum_{\tau=0}^{t-1} \beta^{\tau-t} \underbrace{\omega_\tau}_{>0} \underbrace{S_{\tau,R_t}}_{<0} \geq 0.$$

In particular, using the initial condition (33), this proves by induction that

$$\omega_t S_{I,t} - u'(c_t) \geq 0 \quad \text{for all } t \geq 0. \tag{34}$$

Now suppose the economy were converging to the zero tax steady state, that is, $\Delta_t \equiv F'(k_t) - R_t \to 0$, $c_t \to c > 0$, and $S_{I,t}R_t \to S_I R > 0$. In that case, for large τ, we can approximate ω_τ by

$$\tilde{\omega}_\tau = \beta u'(c) \sum_{\tau' \geq \tau+1} \Delta_{\tau'} (\beta S_I R)^{\tau' - (\tau+1)}.$$

Note that we did not approximate $\Delta_{\tau'}$ to zero, but only quantities with a nonzero limit. Lemma 13 below proves that because $\tilde{\omega}_\tau$ exists for all τ, and $\Delta_{\tau'} \to 0$ it must be that $\tilde{\omega}_\tau \to 0$ and so $\omega_\tau \to 0$. Observe, however, that $\omega_{\tau'} \geq u'(c)$ for sufficiently large τ' by (34) and the assumed convergence to the zero tax steady state. This is a contradiction. Thus the economy cannot converge to the zero tax steady state.

Lemma 13. *Let $\Delta_t \to 0$ be a real-valued sequence. Define $x_T \equiv \sum_{t \geq T} \Delta_t b^{t-T}$ for $b > 0$ and assume x_T exists for all T. Then, $x_T \to 0$ as $T \to \infty$.*

Proof. We distinguish three cases.

- Case 1: $b < 1$. In that case, for any $\epsilon > 0$ we can find $T > 0$ such that $|\Delta_t| < \epsilon(1-b)$ for all $t \geq T$. But this implies that $|x_{T'}| < \epsilon$ for all $T' \geq T$, establishing the result.

- Case 2: $b > 1$. By the recursive nature of x_T we can express it as

$$x_{T+k} = x_T b^{-k} - \sum_{\ell=0}^{k} \Delta_{T+k-\ell} b^{-\ell}.$$

Here, $x_T b^{-k} \to 0$ as $k \to \infty$, so the first term vanishes. Write the absolute value of the second term as

$$\underbrace{\left| \sum_{\ell=0}^{\lfloor k/2 \rfloor} \Delta_{T+k-\ell} b^{-\ell} \right|}_{\leq b/(b-1) \max_{s \geq T+k-\lfloor k/2 \rfloor} |\Delta_s|} + \underbrace{\left| \sum_{\ell=\lfloor k/2 \rfloor+1}^{k} \Delta_{T+k-\ell} b^{-\ell} \right|}_{\leq (\max_{s \geq T} |\Delta_s|) b^{-\lfloor k/2 \rfloor}/(b-1)}$$

which therefore converges to zero as well.

- Case 3: $b = 1$. This case follows because $\sum_t \Delta_t$ is a convergent series, so it must necessarily be the case that the sequence of truncated series $\sum_{t \geq T} \Delta_t$ converges to zero. □

E Proof of Proposition 6

The conditions for optimality are

$$V_t = W\left(R_t a_t - a_{t+1}, V_{t+1}\right)$$
$$W_c\left(R_t a_t - a_{t+1}, V_{t+1}\right) = R_{t+1} W_v\left(R_t a_t - a_{t+1}, V_{t+1}\right) W_c\left(R_{t+1} a_{t+1} - a_{t+2}, V_{t+2}\right).$$

The first equation is the recursion for utility V_t and the second equation is the Euler equation. Linearizing these equations, around the steady state (denoted without time subscripts)

$$W_v\, dV_{t+1} = -W_c R\, da_t + W_c\, da_{t+1} + dV_t - W_c a\, dR_t \tag{35}$$

and

$$\begin{aligned}(RW_cW_{vc} - RW_{cc} - W_{cc})\, da_{t+1} + W_{cc}\, da_{t+2} - (W_vW_c + W_{cc}a)\, dR_{t+1} \\ + (W_{cv} - RW_cW_{vv})\, dV_{t+1} - W_{cv}\, dV_{t+2} \\ = \left(R^2 W_c W_{vc} - W_{cc} R\right) da_t + \left(RW_cW_{vc}a - W_{cc}a\right) dR_t.\end{aligned} \tag{36}$$

Where we have used that $RW_v = 1$ at a steady state. All derivatives are evaluated at the steady state $((R-1)\,a, V)$. We solve (35) and (36) by the method of undetermined coefficients, guessing

$$da_{t+1} = \bar{\lambda}\, da_t + \sum_{s=0}^{\infty} \theta_s\, dR_{t+s} \tag{37a}$$

$$dV_t = W_c R\, da_t + (W_c a) \sum_{s=0}^{\infty} W_v^s\, dR_{t+s}. \tag{37b}$$

The form of equation (37a) is what is required by the Envelope condition. We are left to find $\bar{\lambda}$ and the sequence $\{\theta_s\}$.

Replacing the guesses (37a) and (37b) into (36) we obtain an expression featuring da_t, da_{t+1}, da_{t+2} and dR_{t+s} for $s = 0, 1, \ldots$ Setting the coefficient on da_t to zero gives a quadratic

$$\bar{\lambda}^2 + \left(\frac{2RW_cW_{vc} - W_{cc}(1+R) - R^2W_c^2W_{vv}}{W_{cc} - W_{vc}W_cR}\right)\bar{\lambda} + R = 0 \tag{38}$$

for $\bar{\lambda}$. Note that in the additive separable case (when $W(c, V)$ linear) $\bar{\lambda} = 1$ is a solution.

Setting the coefficient on dR_t to zero gives

$$\theta_0 = \frac{\left(RW_cW_{vc} - W_{cc}\right)a}{2RW_cW_{vc} - W_{cc}\left(1+R\right) + \left(W_{cc} - RW_{cv}W_c\right)\bar{\lambda} - R^2W_c^2W_{vv}}$$
$$= \frac{\bar{\lambda}\left(RW_cW_{vc} - W_{cc}\right)a}{-R\left(W_{cc} - W_{vc}W_cR\right)} = \bar{\lambda}\frac{a}{R}$$

Similarly for dR_{t+1} (after various simplifications)

$$\theta_1 = W_v\bar{\lambda}\theta_0 + \bar{\lambda}W_v\frac{W_v^2 + \left(\frac{W_{cc}}{W_c} + R^*W_cW_{vv} - W_{cv}\right)W_va}{W_{vc} - \frac{W_v}{W_c}W_{cc}}$$

for dR_{t+s} (after many simplifications)

$$\theta_s = W_v\bar{\lambda}\theta_{s-1} + \bar{\lambda}\left(W_v\right)^s\left(1 - W_v\right)\frac{W_{vc} + \frac{W_c}{1-W_v}W_{vv}}{W_{vc} - \frac{W_v}{W_c}W_{cc}}a,$$

for $s = 2, 3, \ldots$

The result follows immediately from this expression. If $W_{vc} + \frac{W_c}{1-W_v}W_{vv} = 0$ then $\bar{\lambda} = 1$ and $\theta_s = W_v\theta_{s-1}$. Otherwise, the second term is nonzero and is geometric in W_v^s.

F Derivation of Implementability Condition

The agent is subject to the budget constraint

$$c_t + a_{t+1} \leq w_tn_t + R_ta_t$$

and the No Ponzi condition $\frac{a_{t+1}}{R_1R_2\cdots R_t} \to 0$. The intratemporal optimality condition is

$$U_n(c_t, n_t) = w_tU_c(c_t, n_t)$$

and the intertemporal Euler equation is

$$W_U(U_t, V_{t+1})U_c(c_t, n_t) = R_{t+1}W_{V'}(U_t, V_{t+1})W_U(U_{t+1}, V_{t+1})U_c(c_{t+1}, n_{t+1}).$$

Substituting these into the budget constraint and using that $\frac{a_{t+1}}{R_1R_2\cdots R_t} \to 0$ we obtain the implementability condition.

G Proof of Proposition 7

Defining $A_t = \frac{\partial}{\partial V_{t+1}} \sum_{t=0}^{\infty} \beta_t W_{t,U}(U_{t,c}c_t + U_{t,n}n_t)$ and $B_t = \sum_{t=0}^{\infty} \frac{\partial(\beta_t W_{t,U})}{\partial U_t}(U_{t,c}c_t + U_{t,n}n_t)$, and using the current value multipliers ν_t and λ_t, the first order conditions are

$$1 + \nu_0 = 0$$

$$-\nu_t + \nu_{t+1} + \mu \frac{A_t}{\beta_{t+1}} = 0$$

$$-\nu_t W_{t,U} U_{t,c} + \mu W_{t,U}\left(U_{t,c} + U_{t,cc}c_t + U_{t,nc}n_t\right) + \mu \frac{B_t}{\beta_t} U_{t,c} = \lambda_t$$

$$\nu_t W_{t,U} U_{t,n} - \mu W_{t,U}\left(U_{t,n} + U_{t,cn}c_t + U_{t,nn}n_t\right) - \mu \frac{B_t}{\beta_t} U_{t,n} = \lambda_t f_{n,t}$$

$$-\lambda_t + \lambda_{t+1} W_{t,V} f_{k,t+1} = 0.$$

If the allocation converges to a steady state, then $A_t/\beta_{t+1} \to A$ and $B_t/\beta_t \to B$ so

$$-\nu_t + \nu_{t+1} + \mu A = 0$$

$$-\nu_t + \mu\left(1 + \frac{U_{cc}c}{U_c} + \frac{U_{nc}}{U_c}n\right) + \mu \frac{B}{W_U} = \lambda_t \frac{1}{W_U U_{t,c}}$$

$$-\nu_t + \mu\left(1 + \frac{U_{cn}c_t}{U_n} + \frac{U_{nn}n}{U_n}\right) + \mu \frac{B}{W_U} = -\lambda_t \frac{f_n}{W_U U_{t,n}}$$

$$-\lambda_t + \lambda_{t+1}\beta f_k = 0,$$

where $\beta \equiv W_V$. Note that

$$\beta f_k - 1 = \frac{\lambda_t}{\lambda_{t+1}} - 1 = -\frac{W_U U_c}{\lambda_t}\mu A. \tag{39}$$

We now argue that this implies that $\beta f_k = 1$ at *any* steady state. If $A = 0$ or $\mu = 0$ the result is immediate from the last equation. If instead $A \neq 0$ and $\mu \neq 0$ then $-\nu_t + \nu_{t+1} + \mu A = 0$ implies that ν_t and hence λ_t diverges to $+\infty$ or $-\infty$. The result then follows since $\beta f_k - 1 = -\frac{W_U U_c}{\lambda_t}\mu A \to 0$. The case with $\mu = 0$ implies that the entire solution is first best, which is uninteresting. The cases with $A = 0$ and $A \neq 0$ are discussed below.

Combining

$$\lambda_t \frac{f_n}{W_U U_n}\tau^L = \mu\left(\frac{U_{cc}c}{U_c} + \frac{U_{nc}}{U_c}n - \frac{U_{cn}c}{U_n} - \frac{U_{nn}n}{U_n}\right),$$

where τ^L is the steady state tax on labor. By normality of consumption and labor $\frac{U_{cc}c}{U_c} + \frac{U_{nc}}{U_c}n - \frac{U_{cn}c}{U_n} - \frac{U_{nn}n}{U_n} < 0$.

Now distinguish three cases according to the asymptotic behavior of ν_t:

- Case A: $\nu_t \to +\infty$, then, $\lambda_t \to -\infty$ and thus $\tau^L = 0$. This requires $A \leq 0$.

- Case B: $\nu_t \to \nu \in \mathbb{R}$, then: $\lambda_t \to \lambda$. This requires $A \leq 0$. There are two subcases to consider. If $\lambda \neq 0$ then $\tau^L \neq 0$ is possible. If instead $\lambda = 0$, then $\tau^L = 0$. However, the first order conditions also imply that $\mu = 0$. Thus, the economy was first best to start with.

- Case C: $\nu_t \to -\infty$. Then, $\lambda_t \to \infty$ and we converge to a first best with $\tau^L = 0$. This case requires $A \geq 0$.

What are the condition for Case B with $\tau^L > 0$? We require $A = 0$. Starting from the definition, we have

$$
\begin{aligned}
A_t &\equiv \frac{\partial}{\partial V_{t+1}} \sum_{t=0}^{\infty} \beta_t W_{t,U} \left(U_{t,c} c_t + U_{t,n} n_t \right) \\
&= \beta_t W_{t,UV} \left(U_{t,c} c_t + U_{t,n} n_t \right) + \beta_t W_{t,VV} \sum_{s=t+1}^{\infty} \beta_{t+1}^{-1} \beta_s W_{s,U} \left(U_{t,c} c_t + U_{t,n} n_t \right) \\
&= \beta_t W_{t,UV} \left(U_{t,c} c_t + U_{t,n} n_t \right) + \beta_t W_{t,VV} W_{t+1,U} u'_{t+1} R_{t+1} a_{t+1}.
\end{aligned}
$$

and so

$$
\begin{aligned}
\frac{A_t}{\beta_{t+1}} &\equiv \frac{W_{t,UV}}{W_{t,V}} \left(U_{t,c} c_t + U_{t,n} n_t \right) + \frac{W_{t,VV}}{W_{t,V}} W_{t+1,U} U_{t+1,c} R_{t+1} a_{t+1} \\
\rightarrow A &\equiv \frac{\beta_U}{\beta} \left(U_{t,c} c_t + U_{t,n} n_t \right) + \frac{\beta_V}{\beta} W_U U_{t,c} R a,
\end{aligned}
$$

where $\beta_X \equiv W_{VX}$ and $X = U, V$.

Note that $\sum_{s=0}^{\infty} \beta_{t,t+s} W_{t+s,U} \left(U_{t,c} c_t + U_{t,n} n_t \right) = W_{t,U} U_{t,c} R_t a_t$, so at a steady state $U_c c + U_n n = (1 - \beta) U_c R a$. Hence,

$$
\begin{aligned}
A &= \frac{\beta_U}{\beta} \left(U_c c + U_n n \right) + \frac{\beta_V}{\beta} W_U U_c R a \\
&= \left(\beta_U + \beta_V \frac{W_U}{1 - \beta} \right) \frac{U_c R a}{\beta (1 - \beta)} = \bar{\beta}'(V) \frac{U_c R a}{\beta (1 - \beta)}.
\end{aligned}
$$

This implies that either $a = 0$ or $\bar{\beta}'(V) = 0$.

H Proof of Proposition 8

The problem under scrutiny in this proof is

$$
V(k_0, b_0) \equiv \max \int_0^{\infty} e^{-\rho t} \left(u(c_t) - v(n_t) \right), \tag{40a}
$$

$$
\dot{c}_t \geq -\frac{\rho}{\sigma} c_t, \tag{40b}
$$

$$
c_t + g_t + \dot{k}_t = f(k_t, n_t) - \delta k_t, \tag{40c}
$$

$$
\int_0^{\infty} e^{-\rho t} \left(u'(c_t) c_t - v'(n_t) n_t \right) \geq u'(c_0)(k_0 + b_0), \tag{40d}
$$

where we parametrize the isoelastic disutility from labor by $v(n) = \frac{1}{1+\zeta} n^{1+\zeta}$. The value function is decreasing in b_0. Problem (40a) has the following necessary first order conditions

$$\Phi_v^W v'(n_t) = \lambda_t f_n(k_t, n_t), \tag{41a}$$

$$\dot{\eta}_t - \rho \eta_t = \eta_t \frac{\rho}{\sigma} + \lambda_t - \Phi_u^W u'(c_t), \tag{41b}$$

$$\dot{\lambda}_t = (\rho - r_t^*) \lambda_t, \tag{41c}$$

$$\eta_0 = -\mu \sigma c_0^{-\sigma-1}(k_0 + b_0), \tag{41d}$$

where we defined $\Phi_v^W \equiv 1 + \mu(1+\zeta)$ and $\Phi_u^W \equiv 1 + \mu(1-\sigma)$ and denoted by r_t^* the before-tax interest rate $f_k - \delta$. Here, μ is the multiplier on the IC constraint (40d), λ_t is the multiplier of the resource constraint (40c), and η_t denotes the costate of consumption c_t. If $\eta_t < 0$, then constraint (40b) is binding. Further, if at $T < \infty$ we have $\eta_t = 0$ for $t \in [T, T+\varepsilon)$ then we must have

$$\eta_T = 0 \quad \text{and} \quad \lambda_T = \Phi_u^W u'(c_T).$$

Otherwise, the conditions for optimality of $T = \infty$ are satisfied provided

$$\eta_t < 0 \quad \text{for all } t, \tag{42}$$

$$e^{-\rho t} \eta_t c_t \to 0. \tag{43}$$

We first prove a helpful lemma relating the occurrence of $T = \infty$ to the multiplier on the IC constraint, μ. This lemma will become important below.

Lemma 14. *If $\mu > 1/(\sigma - 1)$ then $T = \infty$. Thus, if $T < \infty$, μ is bounded from above by $1/(\sigma-1)$.*

Proof. If $\mu > 1/(\sigma-1)$ then $\Phi_u^W < 0$. Suppose T were finite. Using the laws of motion for λ and η, (41b) and (41c) it is straightforward to show that this implies that

$$\eta_T = 0 \quad \text{and} \quad \dot{\eta}_T = r^* \Phi_u^W u'(c_T) > 0.$$

In particular η can only reach 0 from above, so T would have to be zero. But from (41d) we see that $\eta_0 < 0$. Therefore, η_t cannot be hitting zero at all, and $T = \infty$. □

It is convenient to characterize a restricted problem, where T is required to be infinite. Effectively, this implies that constraint (40b) holds with equality throughout and the path of c_t is entirely characterized by c_0. First define

$$\tilde{v}(c_0, k_0) \equiv \min \int e^{-\rho t} v(n_t)$$
$$\text{s.t. } c_t + g_t + \dot{k}_t \leq f(k_t, n_t) - \delta k_t$$
$$c_t = c_0 e^{-\rho/\sigma t},$$

Notice that $\tilde{v}$ is continuous, strictly increasing in c_0, and $\tilde{v}$ is convex in c_0. As c_0 goes to zero, $\tilde{v}$ remains finite and for large values of c_0, $\tilde{v}$ tends to infinity. Next, define the restricted problem

$$V_\infty(k_0, b_0) \equiv \max_{c_0} u(c_0)\frac{\sigma}{\rho} - \tilde{v}(c_0, k_0) \tag{44a}$$

$$c_0^{1-\sigma}\frac{\sigma}{\rho} - (1+\zeta)\tilde{v}(c_0, k_0) \geq u'(c_0)(k_0 + b_0). \tag{44b}$$

By definition, $V_\infty \leq V$, but whenever $T = \infty$ is optimal, $V_\infty = V$. We now show the following result about the restricted problem.

Lemma 15. *There exists a level of initial debt $\bar{b}$ such that a solution to the restricted planner's problem exists for all $b_0 \leq \bar{b}$. The restricted value function is differentiable in a neighborhood of $\bar{b}$ and becomes infinitely steep at $\bar{b}$, $V'_\infty(b) \to -\infty$ as $b_0 \nearrow \bar{b}$.*

Proof. Denote the multiplier on the IC constraint (44b) by $\tilde{\mu}$. Define

$$\bar{b} \equiv \max_{c_0} c_0\frac{\sigma}{\rho} - (1+\zeta)c_0^{\sigma}\tilde{v}(c_0, k_0) - k_0. \tag{45}$$

Using the convexity and the limit properties of $\tilde{v}$ that were derived above, we see that $c_0^{\sigma}\tilde{v}(c_0, k_0)$ is flat in c_0 around 0 and rises faster than c_0 for large c_0. In particular, this implies that the max in (45) is attained for some interior c_0^*. Also, for any $b_0 < \bar{b}$ one can find a value for c_0 satisfying the IC constraint (44b). In particular, the planner's problem (40a) has a solution for all $b_0 \leq \bar{b}$. Note that around c_0^*, the objective function is strictly increasing because for $c_0 = c_0^*$,

$$u'(c_0)\frac{\sigma}{\rho} - \tilde{v}_{c_0} = \frac{\zeta}{1+\zeta}u'(c_0)\frac{\sigma}{\rho} + \sigma c_0^{-1}\tilde{v} > 0.$$

Therefore, for b_0 in a neighborhood of $\bar{b}$, the larger of the two solutions for c_0 is picked and the optimal choice of c_0, $c_0(b)$, is continuous, strictly decreasing, and differentiable. Therefore, the value function $V_\infty(b)$ is differentiable in a neighborhood $(b_1, \bar{b})$ with $b_1 < \bar{b}$.

Now write the Lagrangian for problem (44a) with $b_0 < \bar{b}$, letting $\tilde{\mu}$ be the multiplier on (44b) in problem (44a).

$$L(c_0, \mu) \equiv u(c_0)\frac{\sigma}{\rho} - \tilde{v}(c_0, k_0) + \mu c_0^{-\sigma}\left(c_0\frac{\sigma}{\rho} - (1+\zeta)c_0^{\sigma}\tilde{v}(c_0, k_0) - (k_0 + b_0)\right).$$

The necessary first order condition for c_0 is then,

$$u'(c_0)\frac{\sigma}{\rho} - \tilde{v}_{c_0} + \tilde{\mu}c_0^{-\sigma}\frac{\mathrm{d}}{\mathrm{d}c_0}\left(\frac{\sigma}{\rho}c_0 - (1+\zeta)c_0^{\sigma}\tilde{v}(c_0, k_0)\right) = 0,$$

which, as $b_0 \to \bar{b}$ and $c_0 \to c_0^*$, implies that $\tilde{\mu} \to \infty$ because $\frac{\mathrm{d}}{\mathrm{d}c_0}\left(\frac{\sigma}{\rho}c_0 - (1+\zeta)c_0^{\sigma}\tilde{v}(c_0, k_0)\right) \to$

0 but $u'(c_0)\frac{\sigma}{\rho} - \tilde{v}_{c_0} \not\to 0$.[48] When changing variables from c_0 to $u_0 \equiv c_0^{1-\sigma}/(1-\sigma)$ in problem (44a), it is evident that the problem is convex. Therefore, by Berge's Maximum Theorem, there is a unique maximizer c_0 and a unique multiplier $\bar{\mu}$, for every $b_0 \leq \bar{b}$. Therefore, the value function is continuously differentiable in b_0 with $V'_\infty(b) = -\bar{\mu}c_0^{-\sigma}$. In particular, as $b \to \bar{b}$, $c_0(b) \to c_0^* > 0$, and so

$$\lim_{b \to \bar{b}} V'_\infty(b) = -\infty. \qquad \square$$

Lemma 15 provides a characterization of the problem conditional on $T = \infty$. As a corollary, for $b \leq \bar{b}$ the constraint set of the original problem (40a) is nonempty as well.

To show that there is an interval $[\underline{b}, \bar{b}]$ with $\underline{b} < \bar{b}$ for which $T = \infty$ is optimal, or equivalently $V = V_\infty$, assume to the contrary that there exists a sequence (b_n) approaching $\bar{b}$ for which $T < \infty$ is optimal. In particular, $V(b) > V_\infty(b)$ for all $b = b_n$ along the sequence. Because V and V_∞ are both continuous functions, the set $\{V \neq V_\infty\} = \{T < \infty\}$ has nonzero measure in any neighborhood $(b_1, \bar{b})$ of $\bar{b}$. For $b = \bar{b}$, we prove that the value functions need to coincide in the following lemma.

Lemma 16. *The maximum $b = \bar{b}$ is only feasible if $T = \infty$. Hence, $V(\bar{b}) = V_\infty(\bar{b})$.*

Proof. Suppose $b = \bar{b}$ was feasible for $T < \infty$. Then, the process for consumption at the optimum is governed by[49]

$$\begin{cases} \dot{c}_t = -\frac{\rho}{\sigma} c_t & \text{for } t < T \\ \dot{c}_t = c_t(r_t^* - \rho)/\sigma & \text{for } t \geq T, \end{cases}$$

with a particular initial consumption value c_0. Denote by $\hat{c}_t$ the path which starts at the same initial consumption $\hat{c}_0 = c_0$ but keeps falling at rate $-\rho/\sigma$ forever. Similarly, define by $\hat{n}_t$ the path for labor which keeps k_t fixed but satisfies the resource constraint with consumption equal to $\hat{c}_t$. Clearly, $\hat{n}_t \leq n_t$ for all t. Because the right hand side of (40d) is strictly decreasing in c_t and for $t > 0$, this strictly relaxes the IC constraint. Hence,

$$\int e^{-\rho t}\hat{c}_t^{1-\sigma} - \int e^{-\rho t} v(\hat{n}_t) > \hat{c}_0^{-\sigma}(k_0 + \bar{b}).$$

Notice, however, that for $T = \infty$, we can do even better by optimizing over labor (not necessarily keeping capital constant), leading to

$$\hat{c}_0^{1-\sigma}\frac{\sigma}{\rho} - (1+\zeta)\tilde{v}(\hat{c}_0, k_0) > \hat{c}_0^{-\sigma}(k_0 + \bar{b}).$$

By definition of $\bar{b}$ this is a contradiction. Therefore, an initial level of public debt of $\bar{b}$ necessarily requires $T = \infty$. $\square$

[48] This holds because if $u'(c_0)\frac{\sigma}{\rho} - \tilde{v}_{c_0} = 0$ at c_0^* then, $\zeta\tilde{v}_{c_0} + (1+\zeta)\sigma c_0^{-1}\tilde{v} = 0$ at c_0^*, contradicting the fact that $\tilde{v}$ is positive and strictly increasing.

[49] Note that $r_t^* > 0$ by standard Inada conditions for v and (41a).

Together with Lemma 15, this implies that $V(b)$ must become infinitely steep as $b \to \bar{b}$. We would like to use the Envelope Theorem now to link the local behavior of V to what we know about the μ multiplier from Lemma 14. In order to be able to do so, notice that the value function $V(b)$ is actually the value of a convex problem. To see this, regard the problem in terms of $u_t \equiv u(c_t)$ and $v_t \equiv v(n_t)$,

$$V(b_0) \equiv \max_{u_t, v_t, k_t} \int_0^\infty e^{\ \rho t} (u_t - v_t), \tag{46}$$

$$u_t \geq (\sigma - 1)\frac{\rho}{\sigma} u_t,$$

$$((1-\sigma)u_t)^{-1/(\sigma-1)} + g_t + \dot{k}_t = f\left(k_t, ((1+\zeta)v_t)^{1/(1+\zeta)}\right) - \delta k_t,$$

$$\int_0^\infty e^{\ \rho t}((1-\sigma)u_t - (1+\zeta)v_t) \geq ((1-\sigma)u_t)^{\sigma/(\sigma-1)}(k_0 + b_0).$$

Therefore, there is a globally unique maximizer and a unique multiplier μ on the IC constraint. By Berge's Maximum Theorem, the multiplier μ and the optimal choice of u_0 (and thus c_0) vary continuously in b_0. In particular, this implies that V is continuously differentiable in b_0 with derivative $V'(b) = -\mu c_0{}^{\sigma}$. By optimality of $T < \infty$ and Lemma 14, the multiplier μ is bounded from above by $1/(\sigma - 1)$ and c_0 can be uniformly bounded from below by $\underline{c}_0$ using the IC constraint, where $\underline{c}_0 > 0$ is the smaller of the two solutions to $c_0\sigma/\rho - (1+\zeta)c_0^{\sigma}\tilde{v}(c_0, k_0) = k_0$.[50] Therefore,

$$V'(b) \geq -\frac{\underline{c}_0^{-\sigma}}{\sigma - 1}, \tag{47}$$

for all $b \in \{V \neq V_\infty\}$. For $b \in \{V = V_\infty\}$ it is the case that $V'(b) = V'_\infty(b)$ while for $b \in \{V \neq V_\infty\}$ close enough to $\bar{b}$, $V'(b) < V'_\infty(b)$ by (47). Because $\{V \neq V_\infty\}$ has nonzero measure in any neighborhood of $\bar{b}$, the fundamental theorem of calculus implies for b close to $\bar{b}$,

$$V(b) = V(\bar{b}) + \int_{\bar{b}}^{b} V'(\tilde{b})d\tilde{b} < V(\bar{b}) + \int_{\bar{b}}^{b} V'_\infty(\tilde{b})d\tilde{b} = V_\infty(b),$$

contradicting the optimality of $T < \infty$. Hence, there exists a neighborhood $[\underline{b}, \bar{b}]$ for which $T = \infty$ is optimal.

I Proof of Proposition 9

We proceed by solving the necessary first order conditions to problem (40a). Noting that the problem is convex, see (46), this implies that we are characterizing the unique solution.

By demanding equal growth rates of c_t, n_t, k_t, we demand that $c_t/k_t = c_0/k_0$ and $n_t/k_t = n_0/k_0$ at all times. Define $\Phi_u^W \equiv 1 + \mu(1-\sigma)$ and $\Phi_v^W = 1 + \mu(1+\zeta)$. Solving

[50] Write the IC constraint as $g(c_0, T) = (k_0 + b_0)$. Notice that in the proof of Lemma 16 we showed that $g(\cdot, T)$ can only increase as we move to $T = \infty$, where $g(c_0, \infty) = c_0\frac{\sigma}{\rho} - (1+\zeta)c_0^{\sigma}\tilde{v}(c_0, k_0)$ is strictly concave and $g(0, \infty) = 0$. Therefore, the smallest c_0 allowed by the IC constraint occurs if $T = \infty$ and $b_0 = 0$.

the necessary FOCs,

$$\bar{\Phi}_v^W v'(n_t) = \lambda_t f_n(k_t, n_t) \tag{48}$$

$$c_t + \dot{k}_t = f(k_t, n_t) - \delta k_t$$

$$\int e^{-\rho t}\left(c_t^{1-\sigma} - n_t^{1+\zeta}\right) = c_0^{-\sigma}(k_0 + b_0),$$

and defining $g \equiv (f_k(1,\cdot))^{-1}$ we find expressions for c_0, n_0, b_0 and the constant interest rate $r^* = f_k(k_0, n_0) - \delta$ and wage $w^* = f_n(k_0, n_0)$,

$$r^* = \zeta\frac{\rho}{\sigma} + \rho$$

$$w^* = f_n\left(1, g\left(\zeta\frac{\rho}{\sigma} + \rho + \delta\right)\right)$$

$$n_0 = k_0 g\left(\zeta\frac{\rho}{\sigma} + \rho + \delta\right)$$

$$c_0 = k_0\left[f\left(1, g\left(\zeta\frac{\rho}{\sigma} + \rho + \delta\right)\right) - \delta + \frac{\rho}{\sigma}\right]$$

$$b_0 = c_0\frac{\sigma}{\rho} - \frac{1}{\rho + (1+\zeta)\rho/\sigma} c_0^{\sigma} n_0^{1+\zeta} - k_0.$$

Here, it is straightforward to show that $c_0 > 0$ by definition of g.

The process for η_t can be inferred from its law of motion, $\dot{\eta}_t - \rho\eta_t = \eta_t\frac{\rho}{\sigma} + \lambda_t - \bar{\Phi}_u^W u'(c_t)$, and the transversality condition, $e^{-\rho t}\eta_t c_t \to 0$,

$$\eta_t = -\frac{\hat{\lambda}_0}{\rho + (1+\zeta)\rho/\sigma} e^{\zeta\rho/\sigma t} + \frac{\sigma}{\rho}\bar{\Phi}_u^W c_0^{-\sigma} e^{\rho t}.$$

This leaves us with three conditions for λ_0, η_0, μ,

$$\eta_0 = -\frac{\hat{\lambda}_0}{\rho + (1+\zeta)\rho/\sigma} + \frac{\sigma}{\rho}\bar{\Phi}_u^W c_0^{-\sigma}$$

$$\eta_0 = -\mu\sigma c_0^{-\sigma-1}(k_0 + b_0)$$

$$\bar{\Phi}_v^W n_0^{\zeta} = \lambda_0 w^*,$$

and the inequality $\bar{\Phi}_u^W \leq 0$ ensuring that $\eta_t < 0$ for all t. Define $1 - \tau_0^{\ell} \equiv n_0^{\zeta} c_0^{\sigma} / w^*$. Then, μ can be determined as

$$\mu = \frac{\tau_0^{\ell} + \sigma + \zeta}{\sigma\left((1-\tau_0^{\ell})w^* n_0/c_0 - 1\right) - \tau_0^{\ell}(1+\zeta)}.$$

Note that τ_0^{ℓ} is a decreasing function of k_0, with $\tau_0^{\ell} \to 1$ as $k_0 \to 0$. In particular μ varies with k_0 according to[51]

$$\mu < 0 \quad \text{for} \quad k_0 < \underline{k}$$
$$\mu \geq 1/(\sigma - 1) \quad \text{for} \quad k_0 \in (\underline{k}, \bar{k}]$$
$$\mu < 1/(\sigma - 1) \quad \text{for} \quad k_0 > \bar{k}.$$

This proves that for $k_0 \in (\underline{k}, \bar{k}]$, there exists a debt level $b_0(k_0)$ for which the quantities c_t, n_t, k_t all fall to zero at equal rate $-\rho/\sigma$ and all the necessary optimality conditions of the problem are satisfied.

J Proof of Proposition 10

First, we show that the planner's problem is equivalent to (15). Then we show that the functions ψ and $\bar{\tau}$ are increasing, have $\psi(0) = \tau(0) = 0$ and bounded derivatives.

The planner's problem in this linear economy can be written using a present value resource constraint, that is,

$$\max \int e^{-\rho t} \left(u(c_t) - v(n_t)\right) \tag{49}$$

$$\text{s.t. } \dot{c} \geq c\frac{1}{\sigma}((1-\tau)r^* - \rho)$$
$$\int e^{-r^* t}(c_t - w^* n_t) + G = k_0$$
$$\int e^{-\rho t}\left[(1-\sigma)u(c_t) - (1+\zeta)v(n_t)\right] \geq u'(c_0)a_0,$$

where $G = \int_0^\infty e^{-r^* t} g_t$ is the present value of government expenses, k_0 is the initial capital stock, a_0 is the representative agent's initial asset position, and per-period utility from consumption and disutility from work are given by $u(c_t) = c_t^{1-\sigma}/(1-\sigma)$ and $v(n_t) = n_t^{1+\zeta}/(1+\zeta)$. Note that we assumed $\sigma > 1$. The FOCs for labor imply that given n_0,

$$n_t = n_0 e^{(r^* - \rho)t/\zeta}. \tag{50}$$

Part (a) and (b) in Claim 3.2 imply the existence of $T \in [0, \infty]$ such that $\tau_t = \tau$ for $t \leq T$ and zero thereafter. In particular, the after-tax (net) interest rate will be $r_t = (1 - \bar{\tau})r^* \equiv \bar{r}$ for $t \leq T$ and $r_t = r^*$ for $t > T$. Then, by the representative agent's Euler equation, the path for consumption is determined by

$$c_t = c_0 e^{\frac{\rho - \bar{r}}{\sigma}t + \frac{r^* - \bar{r}}{\sigma}(t-T)^+}. \tag{51}$$

[51] In particular, μ has a pole at $\tau^\ell_{0,\text{pole}} = \sigma(w^* n_0/c_0 - 1)/(1 + \zeta + \sigma w^* n_0/c_0)$. We define $\underline{k}$ to be the value of k_0 corresponding to $\tau^\ell_{0,\text{pole}}$. Notice that one can show that $\tau^\ell_{0,\text{pole}} > -(\sigma + \zeta)$, implying that the pole is always to the left of $\mu = 0$

Substituting equations (50) and (51) into (49), the planner's problem simplifies to,

$$\max_{T,c_0,\bar{n}} \quad \psi_1(T)\frac{c_0^{1-\sigma}}{1-\sigma} - \psi_3\frac{n_0^{1+\zeta}}{1+\zeta} \tag{52}$$

$$\text{s.t. } \psi_2(T)\frac{1}{\chi^*}c_0 + G = k_0 + \psi_3 w^* n_0$$

$$\psi_1(T)c_0^{1-\sigma} - \psi_3 n_0^{1+\zeta} = \chi^* c_0^{-\sigma} a_0,$$

where $\psi_1(T) = \frac{\chi^*}{\chi}\left(1-e^{-\chi T}\right) + e^{-\chi T}$, $\psi_2(T) = \frac{\chi^*}{\hat{\chi}}\left(1-e^{-\hat{\chi}T}\right) + e^{-\hat{\chi}T}$, $\psi_3 = \chi^*\left(r^* + \frac{\rho}{\zeta}\right)^{-1}$ and $\chi = \frac{\sigma-1}{\sigma}\bar{r} + \frac{\rho}{\sigma}$, $\chi^* = \frac{\sigma-1}{\sigma}r^* + \frac{\rho}{\sigma}$, $\hat{\chi} = r^* + \frac{\rho-\bar{r}}{\sigma}$. Notice that $\hat{\chi} > \chi^* > \chi$.

Now normalize consumption and labor

$$c \equiv \psi_1(T)^{1/(1-\sigma)}c_0/\chi^* \qquad n \equiv \psi_3^{1/(1+\zeta)}n_0/\left(\chi^*\right)^{(1-\sigma)/(1+\zeta)}$$

and define an efficiency cost $\psi(T) \equiv \psi_2(T)\psi_1(T)^{1/(\sigma-1)} - 1$, a capital levy $\tau(T) \equiv 1 - \psi_1(T)^{-\sigma/(\sigma-1)}$, and the present value of wage income $\omega n \equiv w^*\psi_3^{\zeta/(1+\zeta)}n$. Then, we can rewrite problem (52) as

$$\max_{T,c,n} \quad u(c) - v(n)$$

$$\text{s.t. } (1+\psi(T))c + G = k_0 + \omega n$$

$$c - \frac{v'(n)}{u'(c)}n = (1-\tau(T))a_0,$$

which is what we set out to show. Notice that $\psi_1(0) = \psi_2(0) = 1$ and so $\psi(0) = \tau(0) = 0$. Further, given our assumption that $\sigma > 1$, $\psi_1(T)$ and $\tau(T)$ are increasing in T. To show that $\psi'(T) \geq 0$, notice that, after some algebra,

$$\frac{d}{dT}\left(\psi_2\psi_1^{1/(\sigma-1)}\right) \geq 0 \quad \Leftrightarrow \quad \hat{\chi}\left(e^{\chi T}-1\right) \leq \chi\left(e^{\hat{\chi}T}-1\right),$$

which is true for any $T \geq 0$ because $\hat{\chi} > \chi$. Therefore, $\psi'(T) \geq 0$, with strict inequality for $T > 0$, implying that $\psi(T)$ is strictly increasing in T.

Now consider the ratio of derivatives,

$$\frac{\psi'(T)}{\tau'(T)} = \frac{1}{\sigma}\psi_2\psi_1^{(1+\sigma)/(\sigma-1)}\left((\sigma-1)\frac{\psi_2'}{\psi_2}\frac{\psi_1}{\psi_1'} + 1\right).$$

Notice that $\psi_1(T) \in [1, \chi^*/\chi]$ and $\psi_2(T) \in [\chi^*/\hat{\chi}, 1]$, so both are bounded away from infinity and zero. Further, the ratio ψ_2'/ψ_1' is also bounded away from infinity, $\psi_2'/\psi_1' = -\frac{1}{\sigma-1}e^{-(\hat{\chi}-\chi)T} \in [-1/(\sigma-1), 0]$, implying that $\psi'(T)/\tau'(T)$ is bounded away from ∞.

K Proof of Proposition 12

We proceed as in the first part of the proof of Proposition 8. As in Section 2, labor supply is inelastic at $n_t = 1$. The problem is then

$$\max \int_0^\infty e^{-\rho t} u(c_t), \tag{53a}$$

$$\dot{C}_t \geq -\frac{\rho}{\sigma} C_t, \tag{53b}$$

$$c_t + C_t + \dot{k}_t = f(k_t) - \delta k_t, \tag{53c}$$

$$\int_0^\infty e^{-\rho t} u'(C_t) C_t \geq u'(C_0)(k_0 + b_0). \tag{53d}$$

Problem (53a) has the following necessary first order conditions

$$\dot{\eta}_t - \rho \eta_t = \eta_t \frac{\rho}{\sigma} + \lambda_t - \Phi_u^W U'(C_t), \tag{54a}$$

$$\dot{\lambda}_t = (\rho - f'(k_t) + \delta)\lambda_t, \tag{54b}$$

$$\eta_0 = -\mu \sigma C_0^{-\sigma-1}(k_0 + b_0), \tag{54c}$$

$$u'(c_t) = \lambda_t \tag{54d}$$

where we defined $\Phi_v^W \equiv \mu(1+\zeta)$ and $\Phi_u^W \equiv \mu(1-\sigma)$. Here, μ is the multiplier on the IC constraint (53d), λ_t is the multiplier of the resource constraint (53c), and η_t denotes the costate of capitalists' consumption C_t. If $\eta_t < 0$, then constraint (53b) is binding.

Suppose $T < \infty$, in which case we have $\eta_T = 0$. Using the laws of motion for λ and η, (54a) and (54b), it is straightforward to show that

$$\eta_T = 0 \quad \text{and} \quad \dot{\eta}_T = r^* \Phi_u^W U'(C_T) > 0.$$

In particular η can only reach 0 from above. A contradiction.