Positive Long Run Capital Taxation: Chamley-Judd Revisited*

Ludwig Straub
MIT

Iván Werning
MIT

February 2015

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 it. 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 tax rates. The model in Chamley (1986) imposes an upper bound on capital taxes. We provide conditions under which these constraints bind forever, implying positive long run taxes. When this is not the case, the long-run tax may be zero. However, if preferences are recursive and discounting is locally non-constant (e.g., not additively separable over time), a zero long-run capital tax limit must be 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 increasing consumption taxes does not provide a firm rationale against capital taxation.

Keywords: Optimal taxation; Capital taxation

1 Introduction

One of the most startling results in optimal tax theory is the famous finding by Chamley (1986) and Judd (1985). Although working in somewhat different settings, their conclu-

*This paper benefited from detailed comments by Fernando Alvarez, Peter Diamond, Mikhail Golosov and Stefanie Stantcheva and was very fortunate to count on research assistance from Greg Howard, Lucas Manuelli and Andrés Sarto. All remaining errors are our own. This paper supersedes "A Reappraisal of Chamley-Judd Zero Capital Taxation Results" presented by Werning at the Society of Economic Dynamics' 2014 meetings in Toronto.

sions 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, taking convergence to a steady state 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 was not anticipated by economists' intuitions, 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 remains downright puzzling, as witnessed by the fact that economists have continued to take turns putting forth various intuitions to interpret it, none definitive nor universally accepted.

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 Chamley (1986) and Judd (1985), 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, indeed, quite the contrary. We discuss what went wrong with the original results, their interpretations and proofs.

Before summarizing our results in greater detail, it is useful to briefly recall the setups in Chamley (1986) and Judd (1985). Start with the similarities. Both papers assume infinitely-lived agents and take as given an initial stock of capital. Taxes are basically restricted to proportional taxes on capital and labor—lump-sum taxes are either ruled out or severely limited. To prevent expropriatory capital levies, the tax rate on capital is constrained by an upper bound.[2] Turning to differences, Chamley (1986) focused on

[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]Consumption taxes (Coleman II, 2000) and dividend taxes with capital expenditure (investment) deductions (Abel, 2007) can mimic initial wealth expropriation. Both are disallowed.

a representative agent and assumed 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. As emphasized by Judd (1985), it is most remarkable that a zero long-run tax result obtains despite the restriction to budget balance.[3] 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 a host of difficulties governments may face managing their debts or assets.[4]

We begin with the model in Judd (1985) and focus on situations where desired redistribution runs from capitalists to workers. Working with an isoelastic utility over consumption for capitalists, $U(C) = \frac{C^{1-\sigma}}{1-\sigma}$, we establish that when the intertemporal elasticity of substitution (IES) is below one, $\sigma > 1$, taxes rise and converge towards a positive limit tax, instead of declining towards zero. This limit tax is significant, driving capital to its lowest feasible level. Indeed, with zero government spending the lowest feasible capital stock is zero and the limit tax rate on wealth goes to 100%. The long-run tax is not only *not* zero, it is far from that.

The economic intuition we provide for this result is based on the anticipatory savings effects of future tax rates. When the IES is less than one, any anticipated increase in taxes leads to higher savings today, since the substitution effect is relatively small and dominated by the income effect. When the day comes, higher tax rates do eventually lower capital, but if the tax increase is sufficiently far off in the future, then the increased savings generate a higher capital stock over a lengthy transition. This is desirable, since it increases wages and tax revenue. To exploit such anticipatory effects, the optimum involves an increasing path for capital tax rates. This explains why we find positive tax rates that rise over time and converge to a positive value, rather than falling towards zero.

When the IES is above one, $\sigma < 1$, we verify numerically that the solution converges to the zero-tax steady state.[5] This also relies on anticipatory savings effects, working in reverse. However, we show that this convergence may be very slow, potentially taking

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

[4] Another issue may arise on the other end. Without constraints on debt, capitalists may become highly indebted or not own the capital they manage. The 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).

[5] We complement these numerical results by proving a local convergence result around the zero-tax steady state when $\sigma < 1$.

centuries for wealth taxes to drop below 1%. Indeed, the speed of convergence is not bounded away from zero in the neighborhood of a unitary IES, $\sigma = 1$. Thus, even for those cases where the long-run tax on capital is zero, this property provides a misleading summary of the model's tax prescriptions.

We confirm our intuition based on anticipatory effects by generalizing our results for the Judd (1985) economy to a setting with arbitrary savings behavior of capitalists. Within this more general environment we also derive an inverse elasticity formula for the steady state tax rate, closely related to one in Piketty and Saez (2013). However, our derivation stresses that the validity of this formula requires sufficiently fast convergence to an interior steady state, a condition that we show fails in important cases.

We then turn to the representative agent Ramsey model studied by Chamley (1986). As is well appreciated, in this setting upper bounds on the capital tax rate are imposed to prevent expropriatory levels of taxation. We provide two sets of results.

Our first set of results show that in cases where the tax rate does converge to zero, there are other implications of the model, hitherto unnoticed. These implications undermine the usual interpretation against capital taxation. Specifically, if the optimum converges to a steady state where the bounds on tax rates are slack, we show that the tax is indeed zero. However, for recursive nonadditive utility, we also show that this zero-tax steady state is necessarily accompanied by either zero private wealth—in which case the tax base is zero—or a zero tax on labor income—in which case the first best is achieved. This suggests that zero taxes on capital are attained only after taxes have obliterated private wealth or allowed the government to proceed without any distortionary taxation. Needless to say, these are not the scenarios typically envisioned when interpreting zero long-run tax results. Away from additive utility, the model simply does not justify a steady state with a positive tax on labor, a zero tax on capital and positive private wealth.

Our second set of results show that the tax rate may not converge to zero. In particular, we show that the upper bounds imposed on the tax rate may bind forever, implying a positive long-run tax on capital. We prove that this is guaranteed if the IES is below one and 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: the planner chooses to tax capital indefinitely, but is not compelled to do so. Intuitively, higher debt leads to higher labor taxes, making capital taxation attractive to ease the labor tax burden. However, because the tax rate on capital is capped, the only way to expand capital taxation is to prolong the time spent at the bound. At some point, for high enough debt, indefinite taxation becomes optimal.

All of these results run counter to what is certainly by now established wisdom, cemented by a significant followup literature, extending and interpreting long-run zero tax results. In particular, Judd (1999) presents an argument against positive capital taxation without requiring convergence to a steady state, using a model close to the one in Chamley (1986). However, as we explain, these arguments fail because they invoke assumptions on endogenous multipliers that are violated at the optimum. We also explain why the intuition offered in that paper, based on the observation that a positive capital tax is equivalent to a rising tax on consumption, does not provide a valid rationale against indefinite capital taxation.

To conclude, we present a hybrid model that combines heterogeneity and redistribution as in Judd (1985), but allows for government debt as in Chamley (1986). Capital taxation turns out to be especially potent in this setting: whenever the IES is less than one, the optimal policy sets the tax rate at the upper bound forever. This suggests that positive long-run capital taxation should be expected for a wide range of models that are descendants of Chamley (1986) and Judd (1985).

2 Capitalists and Workers

We start with the two-class economy without government debt laid out in Judd (1985). Time is indefinite and discrete, with periods labeled by $t = 0, 1, 2, \ldots$.[6] 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 returns to capital to pay for transfers targeted to workers.

Preferences. Both capitalists and workers 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. Capitalists have utility

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

[6] Judd (1985) formulates the model in continuous time, but this difference is immaterial. As usual, the continuous-time model can be thought of as a limit of the discrete time one as the length of each period shrinks to zero.

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 (IES). Workers have utility

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

where u is increasing, concave, continuously differentiable and $\lim_{c \to 0} u'(c) = \infty$.

Technology. Output is obtained from capital and labor using a neoclassical constant returns production function $F(k_t, n_t)$ satisfying standard conditions.[7] 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

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

There is some given positive level of initial capital, $k_0 > 0$.

Markets and Taxes. Markets are perfectly competitive, with labor being paid wage $w_t^* = F_L(k_t, n_t)$ and the before-tax return on capital being given by

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

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

$$R_t = (1 - \tau_t)(R_t^* - 1) + 1 \qquad \text{or} \qquad R_t = (1 - \mathcal{T}_t) R_t^*,$$

where τ_t is the tax rate on the net return to wealth and $\mathcal{T}_t$ the tax rate on the gross return to wealth, or wealth tax for short. Whether we consider a tax on net returns or on gross returns is irrelevant and a matter of convention. We say that capital is taxed whenever $R_t < R_t^*$ and subsidized whenever $R_t > R_t^*$.

[7] We assume that F is increasing and strictly concave in each argument, continuously differentiable, and satisfying the standard Inada conditions $F_k(k, L) \to \infty$ as $k \to 0$ and $F_k(k, L) \to 0$ as $k \to \infty$. Moreover assume that capital is essential for production, that is, $F(0, n) = 0$ for all n.

Capitalist and Worker Behavior. Capitalists solve

$$\max_{\{C_t,a_{t+1}\}} \sum_{t=0}^{\infty} \beta^t U(C_t) \quad \text{s.t.} \quad C_t + a_{t+1} = R_t a_t \quad \text{and} \quad a_{t+1} \geq 0,$$

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

$$U'(C_t) = \beta R_{t+1} U'(C_{t+1}) \qquad \text{and} \qquad \beta^t U'(C_t) a_{t+1} \to 0,$$

are necessary and sufficient for optimality.

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

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

which uses the fact that $F_n = F - F_k k$. Here $T_t \in \mathbb{R}$ represent government lump-sum transfers (when positive) or taxes (when negative) to workers.[8]

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[9]

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

subject to

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

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

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

[8] Equivalently, one can set up the model without lump-sum transfers/taxes to workers, but allowing for a proportional tax or subsidy on labor income. Such a tax perfectly targets workers without creating any distortions, since labor supply is perfectly inelastic in the model.

[9] 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).

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 characterize their associated policies. We often focus on the case with no weight on capitalists, $\gamma = 0$, to ensure that desired redistribution runs from capitalists towards workers. Equation (1b) is the resource constraint. Equation (1c) combines the capitalists' first-order condition and budget constraint and (1d) imposes the transversality condition; together conditions (1c) and (1d) ensure the optimality of the capitalists' saving decision.

The necessary first-order conditions are

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

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

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

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

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 $\beta^t \lambda_t$ and $\beta^t \mu_t$, respectively.

2.1 Previous Steady State Results

Judd (1985, pg. 72, Theorem 2) provided a zero-tax result, which we adjust in the following statement to stress 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: $\mathcal{T}_t = 1 - R_t / R_t^* \to 0$.*

The proof is immediate: from equation (2d) we obtain $R_t^* \to 1/\beta$, while the capitalists' Euler equation requires that $R_t \to 1/\beta$. 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 elaborate on a similar point in Section 3.3.

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, even if the allocation converges to an interior steady state but μ_t does not converge, 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 $\mu_{t+1} - \mu_t$ to zero. 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) *Assume $\sigma = 1$. Suppose the allocation converges to an interior steady state, so that c_t, C_t and k_{t+1} converge to strictly positive values. Then,*

$$\mathcal{T}_t \to \frac{1-\beta}{1+\gamma v \beta/(1-\gamma v)},$$

where $v = \lim v_t$ and the multiplier μ_t in the system of first-order conditions (2c) does not converge. 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.[10]

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 suggest that whenever $\sigma \neq 1$ the long-run tax on capital is zero. We shall show that this is not the case.

2.2 Main Result: Positive Long-Run Taxation

Logarithmic Utility. Before studying $\sigma > 1$, our main case of interest, it is useful to review the special case with logarithmic utility, $\sigma = 1$. We assume $\gamma = 0$ to guarantee

[10] Lansing (1999) suggests a technical difficulty with the argument in Judd (1985) that is specific to $\sigma = 1$. Indeed, at $\sigma = 1$ one degree of freedom is lost in the planning problem, since C_{t-1} must be proportional to k_t. However, since equations (2) can still satisfied by the optimal allocation for some sequence of multipliers, we believe the issue can be framed exactly as Reinhorn (2002 and 2013) did, emphasizing the non convergence of multipliers.

that desired redistribution runs from capitalists to workers.

When $U(C) = \log C$ capitalists save at a constant rate $s > 0$,

$$C_t = (1-s)R_t k_t \qquad \text{and} \qquad k_{t+1} = sR_t k_t.$$

Although $s = \beta$ with logarithmic preferences, nothing we will derive depends on this fact, so we can interpret s as a free parameter that is potentially divorced from β.[11]

The planning problem becomes

$$\max_{\{c_t, k_{t+1}\}} \sum_{t=0}^{\infty} \beta^t u(c_t) \quad \text{s.t.} \quad c_t + \frac{1}{s}k_{t+1} + g = f(k_t) + (1-\delta)k_t,$$

with k_0 given. This amounts to an optimal neoclassical growth problem, where the price of capital equals $\frac{1}{s} > 1$ instead of the actual unit cost. The difference arises from the fact that capitalists consume a fraction $1-s$. The government and workers must save indirectly through capitalists, entrusting them with resources today by holding back on current taxation, so as to extract more tomorrow. From their perspective, technology appears less productive because capitalists feed off a fraction of the investment. Lower saving rates s increase this inefficiency.[12]

Since the planning problem is equivalent to a standard optimal growth problem, we know that there exists a unique interior steady state and that it is globally stable. The modified golden rule at this steady state is $\beta s R^* = 1$. A steady state also requires $sRk = k$, or simply $sR = 1$. Putting these conditions together gives $R/R^* = \beta < 1$.

Proposition 1. *Suppose $\gamma = 0$ and that capitalists have logarithmic utility, $U(C) = \log C$. Then the solution to the planning problem converges monotonically to a unique steady state with a positive tax on capital given by $\mathcal{T} = 1 - \beta$.*

This proposition echoes the result in Lansing (1999), as summarized by Theorem 2, but also establishes the convergence to the steady state. Interestingly, the long-run tax rate depends only on β, not on the savings rate s or other parameters.

Although Lansing (1999) and the subsequent literature interpreted this result as a knife-edged counterexample, we will argue that this is not the case, that positive long

[11]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. We pursue this line of thought in Section 2.3 below.

[12]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.

run taxes are not special to logarithmic utility. 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 tax rate $\mathcal{T}_t(\sigma)$ converges as $\sigma \to 1$ to the tax rate obtained in the logarithmic case (which we know is positive for large t). While this is 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: $\lim_{t\to\infty} \lim_{\sigma\to 1} \mathcal{T}_t(\sigma)$ does not equal $\lim_{\sigma\to 1} \lim_{t\to\infty} \mathcal{T}_t(\sigma)$. Therefore, arguing by continuity does not help characterize the long run tax rate $\lim_{t\to\infty} \mathcal{T}_t(\sigma)$ as a function of σ. We proceed by tackling the problem with $\sigma \neq 1$ directly.

Positive Long-Run Taxation: IES < 1. 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, $\gamma = 0$. Section 2.4 shows that the same results apply for other value of γ, as long as redistribution from capitalists to workers is desired.

Towards a contradiction, suppose the allocation were to converge 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. Combining equations (2c) and (2d) and taking the limit for the allocation, we obtain

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

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

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

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), whenever $\mu_t \geq 0$ then $\mu_{t+1} \geq 0$. It follows that $\mu_t \geq 0$ for all $t = 0, 1, \ldots$, a contradiction to $\mu_t \to \mu < 0$. This proves that the solution cannot converge to any interior steady state, including the zero-tax steady state.

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 other interior steady state.*

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

Thus, 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 $\frac{1}{\beta}k + g = f(k) + (1-\delta)k$, using the fact that (1c) implies $C = \frac{1-\beta}{\beta}k$ at any steady state.[13] We next show that the solution does indeed converge, and that it does so towards the lowest sustainable value of capital, k_g, so that the long-run tax on capital is strictly positive. The proof uses the fact that $\mu_t \to \infty$ and $c_t \to 0$, 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: $\mathcal{T}_t = 1 - \frac{R_t}{R_t^*} \to \mathcal{T}^g > 0$. The limit tax $\mathcal{T}^g$ is decreasing in spending g, with $\mathcal{T}^g \to 1$ as $g \to 0$.*

The zero-tax conclusion in Judd (1985) is invalidated here because the allocation does not converge to an interior steady state and multipliers do not converge. According to our result, the tax rate not only does not converge to zero, it reaches a sizable level. Perhaps counterintuitively, the long-run tax on capital, $\mathcal{T}^g$, is inversely related to the level of government spending, since k_g is increasing with spending g. This underscores that long-run capital taxation is not driven by budgetary necessity.

As the proposition shows, optimal taxes may reach very high levels. Up to this point, we have placed no limits on tax rates. It may be of interest to consider a situation where the planner is further constrained by an upper bound on the tax rate for net returns (τ) or gross wealth ($\mathcal{T}$), perhaps due to evasion or political economy considerations. If these bounds are sufficiently tight to be binding, it is natural to conjecture that the optimum converges to these bounds, and to an interior steady state allocation with a positive limit for worker consumption, $\lim_{t\to\infty} c_t > 0$.

Solution for IES near 1. Figure 1 displays the time path for the capital stock and the tax rate on wealth, $\mathcal{T}_t = 1 - R_t/R_t^*$, for a range of σ that straddles the logarithmic $\sigma = 1$ 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 zero-tax steady state. Our numerical method is based on a recursive formulation of the problem described in the appendix.

To clarify the magnitudes of the tax on wealth, $\mathcal{T}_t$, consider an example: If $R^* = 1.04$ 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 wealth tax of 4% represents a tax rate of 100% on net returns, and so on.

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

[13] Here we assume that government spending g is feasible, that is, $g < \max_k\{f(k) + (1-\delta)k - \frac{1}{\beta}k\}$.

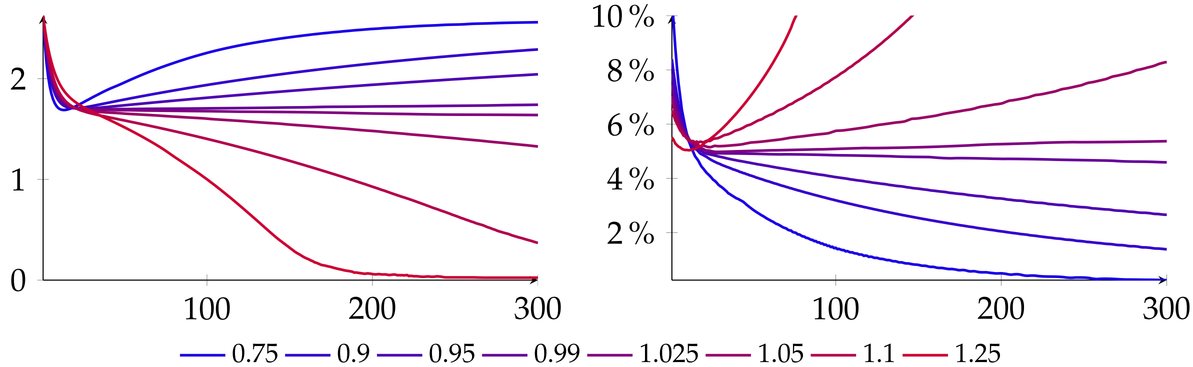

Figure 1: Optimal time paths over 300 years for capital k_t (left panel) and wealth taxes $\mathcal{T}_t$ (right panel) for various values of σ.

convergence is monotone[14], taking around 200 years for $\sigma = 1.25$. The asymptotic tax rate is very high, approximately $\mathcal{T}^g = 1 - R/R^* = 85\%$, lying outside the figure's range since the after-tax return equals $R = 1/\beta$ in the long run, this implies that the before-tax return $R^* = f'(k_g) + 1 - \delta$ is exorbitant.

Second, for $\sigma < 1$, the path for capital is non-monotonic[15] and eventually converges to the zero-tax steady state. 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 $\sigma \to 1$, with positive taxation as described in Proposition 1. 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 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 $\mathcal{T} = 1 - \beta = 5\%$. 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 $\sigma < 1$.

Intuition: Anticipatory Effects of Future Taxes on Current Savings. Why does the optimal tax eventually rise for $\sigma > 1$ and fall for $\sigma < 1$? Why are the dynamics relatively

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

[15]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.

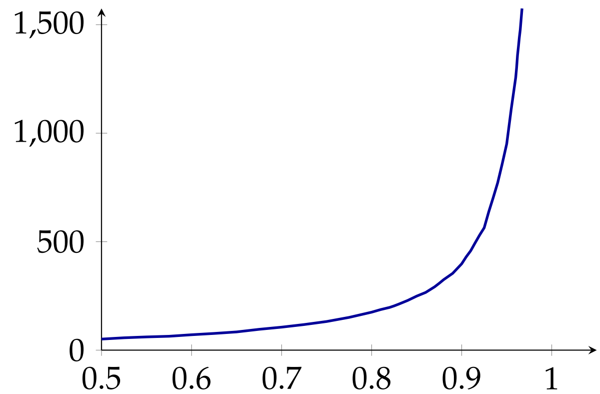

Figure 2: Time elapsed (in years) until wealth tax rate $\mathcal{T}_t$ falls below 1% for $\sigma \in (\frac{1}{2}, 1)$.

slow for σ near 1?

To address these normative questions it helps to back up and review the following positive exercise. Start from a constant tax on wealth and imagine an unexpected announcement of higher future taxes on capital. How do capitalists react today? There are substitution and income effects pulling in opposite directions. When $\sigma > 1$ the substitution effect is muted, compared to the income effect, and capitalists lower their consumption to match the drop in future consumption. As a result, capital rises in the short run, even if it may fall in the far future.[16] When $\sigma < 1$ the substitution effect is stronger and capitalists increase current consumption. In the logarithmic case, $\sigma = 1$, the two effects cancel out, so that current consumption and savings are unaffected.

Returning to the normative questions, lowering capitalists' consumption and increasing capital is desirable for workers. When $\sigma < 1$, this can be accomplished by promising lower tax rates in the future. This explains why a declining path for taxes is optimal. In contrast, when $\sigma > 1$, the same is accomplished by promising higher tax rates in the future; explaining the increasing path for taxes. These incentives are absent in the logarithmic case, when $\sigma = 1$, explaining why the tax rate converges to a constant.

When $\sigma < 1$ the rate of convergence to the zero-tax steady state is also driven by these anticipatory effects. With σ near 1, the potency of these effects is small, explaining why the rate of convergence is low and indeed becomes vanishingly small as $\sigma \uparrow 1$.

In contrast to previous intuitions offered for zero long-run tax results, the intuition we provide for our results—zero and nonzero long-run taxes alike, depending on σ—is not

[16] It is important to note that $\sigma > 1$ does *not* imply that the supply for savings "bends backward". Indeed, as a positive exercise, if taxes are raised permanently within the model, then capital falls over time to a lower steady state for any value of σ, including $\sigma > 1$. Higher values of σ imply a less elastic response over any *finite* time horizon, and thus a slower convergence to the lower capital stock. The case with $\sigma > 1$ is widely considered more plausible empirically.

about the desired *level* for the tax. Instead, we provide a rationale for the desired *slope* in the path for the tax: an upward path when $\sigma > 1$ and a downward path when $\sigma < 1$. The conclusions for the optimal long-run tax then follow from these desired slopes, rather than the other way around.

2.3 General Savings Functions and Inverse Elasticity Formula

The intuition suggests that the essential ingredient for positive long run capital taxation in the model of Judd (1985) is that capitalists' savings decrease in future interest rates. To make this point even more transparently, we now modify the model and assume capitalists behave according to a general "ad-hoc" savings rule,

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

where $S(I_t; R_{t+1}, R_{t+2}, \dots) \in [0, I_t]$ is a continuously differentiable function taking as arguments current wealth $I_t = R_t k_t$ and future interest rates $\{R_{t+1}, R_{t+2}, \dots\}$. We assume that savings increase with income, $S_I > 0$. This savings function encompasses the case where capitalists maximize an additively separable utility function, as in Judd (1985), but is more general. For example, the savings function can 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.

Again, we focus on the case $\gamma = 0$. The planning problem is then

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

subject to

$$\begin{aligned} c_t + R_t k_t + g &= f(k_t) + (1-\delta)k_t, \\ k_{t+1} &= S(R_t k_t; R_{t+1}, R_{t+2}, \dots), \end{aligned}$$

with k_0 given.

We can show that, consistent with the intuition spelled out above, long-run capital taxes are positive whenever savings decrease in future interest rates.

Proposition 4. *Suppose $\gamma = 0$ and assume the savings function is decreasing in future rates, so that $S_{R_t}(I; R_1, R_2, \dots) \leq 0$ for all $t = 1, 2, \dots$ and all arguments $\{I, R_1, R_2, \dots\}$. If the optimum converges to an interior steady state in c, k, and R, and at the steady state $\beta R S_I \neq 1$,* *then the limit tax rate is positive and $\beta RS_I < 1$.*

This generalizes Proposition 2, since the case with iso-elastic utility and IES less than one is a special case satisfying the hypothesis of the proposition. Once again, the intuition here is that the planner exploits anticipatory effects by raising tax rates over time to increase present savings.

The result requires $\beta RS_I < 1$ at the steady state, which is satisfied when savings are linear in income, since then $S_I R = 1$ at a steady state. Note that savings are linear in income in the isoelastic utility case. More generally, $RS_I < 1$ is natural, as it ensures local stability for capital given a fixed steady-state return, i.e. the dynamics implied by the recursion $k_{t+1} = S(Rk_t, R, R, \cdots)$ for fixed R.

Inverse Elasticity Formula. There is a long tradition relating optimal tax rates to elasticities. In the context of our general savings model, spelled out above, we derive the following "inverse elasticity rule"

$$\mathcal{T} = 1 - \frac{R}{R^*} = \frac{1 - \beta RS_I}{1 + \sum_{t=1}^{\infty} \beta^{-t+1} \epsilon_{S,t}}, \tag{4}$$

where $\epsilon_{S,t} \equiv \frac{R_t}{S} \frac{\partial S}{\partial R_t}(R_0 k_0; R_1, R_2, \ldots)$ denotes the elasticity of savings with respect to future interest rates evaluated at the steady state in c, k, and R. Although the right hand side is endogenous, equation (4) is often interpreted as a formula for the tax rate. Our inverse elasticity formula is closely related to a condition derived by Piketty and Saez (2013, see their Section 3.3, equation 16).[17]

We wish to make two points about our formula. First, note that the relevant elasticity in this formula is *not* related to the response of savings to *current*, transitory or permanent, changes in interest rates. Instead, the formula involves a sum of elasticities of savings with respect to *future* changes in interest rates. Thus, it involves the anticipatory effects discussed above. Indeed, the variation behind our formula changes the after-tax interest rate at a single future date T, and then takes the limit as $T \to \infty$. For any finite T, the term $\sum_{t=1}^{T} \beta^{-t+1} \epsilon_{S,t}$ represents the sum of the anticipatory effects on capitalists' savings behavior in periods 0 up to $T - 1$; while $\sum_{t=1}^{\infty} \beta^{-t+1} \epsilon_{S,t}$ captures the limit as $T \to \infty$. It is important to keep in mind that, precisely because it is anticipatory effects that matter,

[17] Their formula is derived under the special assumptions of additively separable utility, an exogenously fixed international interest rate and an exogenous wage. None of this is important, however. The two formulas remain different because of slightly different elasticity definitions; ours is based on partial derivatives of the primitive savings function S with respect to a single interest rate change, while theirs is based on the implicit total derivative of the capital stock sequence with respect to a permanent change in the interest rate.

the relevant elasticities are negative in standard cases, e.g. with additive utility and IES below one.

Second, the derivation we provide in the appendix requires convergence to an interior steady state as well as additional conditions (somewhat cumbersome to state) to allow a change in the order of limits and obtain the simple expression $\sum_{t=1}^{\infty} \beta^{-t+1} \epsilon_{S,t}$. These latter conditions seem especially hard to guarantee ex ante, with assumptions on primitives, since they may involve the endogenous speed of convergence to the presumed interior steady state.[18] As we have shown, in this model one cannot take these properties for granted, neither the convergence to an interior steady state (Proposition 3) nor the additional conditions. Indeed, Proposition 4 already supplies counterexamples to the applicability of the inverse elasticity formula.

Corollary. *Suppose $\gamma = 0$ and assume the savings function is decreasing in future rates, so that $S_{R_t}(I; R_1, R_2, \dots) \leq 0$ for all $t = 1, 2, \dots$ and all arguments $\{I, R_1, R_2, \dots\}$. Then the economy cannot converge to a point where $\beta R S_I \neq 1$, the denominator in the inverse elasticity formula (4) is negative, $1 + \sum_{t=1}^{\infty} \beta^{-t+1} \epsilon_{S,t} < 0$, and formula (4) holds.*

This result provides conditions under which the formula (4) cannot characterize the long run tax rate. Whenever the discounted sum of elasticities with respect to future rates, $\sum_{t=1}^{\infty} \beta^{-t+1} \epsilon_{S,t}$, is negative and less than -1, the formula implies a negative limit tax rate. Yet, under the same conditions as in Proposition 4, this is not possible since this result shows that if convergence takes place, the tax rate is positive.

The case with additive and iso-elastic utility is an extreme example where the sum of elasticities $\sum_{t=1}^{\infty} \beta^{-t+1} \epsilon_{S,t}$ diverges. As it turns out, in this case $\beta^{-t} \epsilon_{S,t} = -\frac{\sigma-1}{\sigma} \frac{1-\beta}{\beta}$ at a steady state and the sum of elasticities diverges. It equals $+\infty$ if the IES is greater than one, or $-\infty$ if the IES is less than one.[19] In both cases, formula (4) suggests a zero steady state tax rate. Piketty and Saez (2013) use this to argue that this explains the Chamley-Judd result of a zero long-run tax. However, as we have shown, when the IES is less than one the limit tax rate is not zero. This counterexample to the applicability of the inverse elasticity formula (4) assumes additive utility and, thus, an infinite sum of elasticities. However, the problem may also arise for non-additive preferences or with ad hoc saving functions. Indeed, the conditions for the corollary may be met in cases where the sum of elasticities is finite, as long as its value is sufficiently negative.

[18] Unfortunately, one cannot ignore transitions by choice of a suitable initial condition. For example, even in the additive utility case with $\sigma < 1$ and even if we start at the zero capital tax steady state, capital does not stay at this level forever. Instead, capital first falls and then rises back up at a potentially slow rate.

[19] Proposition 12 in the appendix shows that a non finite value for the sum of elasticities is a general feature of recursive preferences.

It should be noted that our corollary provides sufficient conditions for the formula to fail, but other counterexamples may exist outside its realm. Suggestive of this is the fact that when the denominator is positive but small the formula may yield tax rates above 100%, which seems nonsensical, requiring $R < 0$. More generally, very large tax rates may be inconsistent with the fact that steady state capital must remain above $k_g > 0$.

To summarize, the inverse elasticity formula (4) fails in important cases, providing misleading answers for the long run tax rate. This highlights the need for caution in the application of steady state inverse elasticity rules.

2.4 Redistribution Towards Capitalists

In the present model, a desire to redistribute towards workers, away from capitalists, is a prerequisite to create a motive for positive wealth taxation. Proposition 3 assumes no weight on capitalists, $\gamma = 0$, to ensure that desired redistribution runs in this direction. When $\gamma > 0$ the same results obtain as long as the desire for redistribution continues to run from capitalists towards workers. In contrast, when γ is high enough the desired redistribution flips from workers to capitalists. When this occurs, the optimum naturally involves negative tax rates, to benefit capitalists.

We verify these points numerically. Figure 3 illustrates the situation by fixing $\sigma = 1.25$ and varying the weight γ. Since initial capital is set at the zero-tax steady state, k^*, the direction of desired redistribution flips exactly at $\gamma^* = u'(c^*)/U'(C^*)$. At this value of γ, the planner is indifferent between redistributing towards workers or capitalists at the zero-tax steady state (k^*, c^*, C^*).[20] When $\sigma > 1$ and $\gamma > \gamma^*$ the solution converges to the highest sustainable capital k^g, the highest solution to $\frac{1}{\beta}k + g = f(k) + (1-\delta)k$, rather than k_g, the lowest solution to the same equation.

A deeper understanding of the dynamics can be grasped by noting that the planning problem is recursive in the state variable (k_t, C_{t-1}). It is then possible to study the dynamics for this state variable locally, around the zero tax steady state, by linearizing the first-order conditions (2). We do so for a continuous-time version of the model, to ensure that our results are comparable to Kemp et al. (1993). The details are contained in the appendix. We obtain the following characterization.

Proposition 5. *For a continuous-time version of the model,*

[20] Rather than displaying γ in the legend for Figure 3, 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 $C = \frac{1-\beta}{\beta}k$. The case $\gamma = \gamma^*$ corresponds to 0 in this transformation.

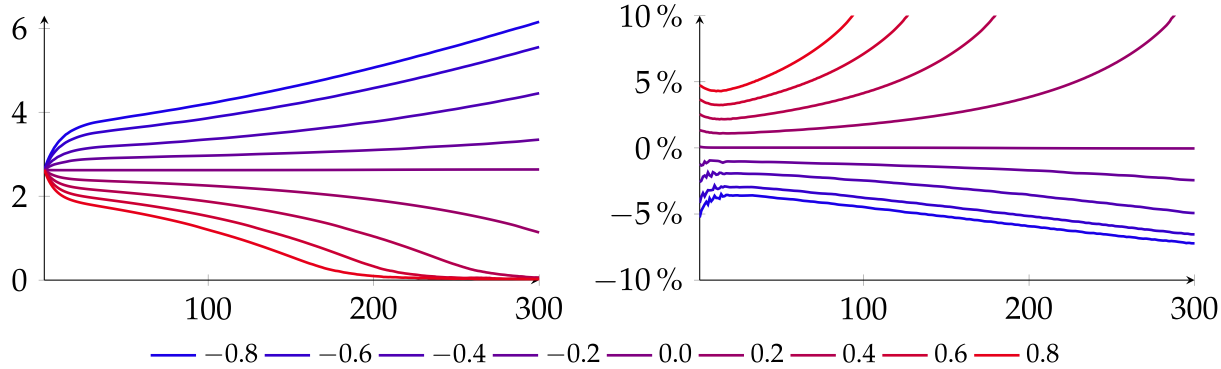

Figure 3: Optimal time paths over 300 years for capital stock (left panel) and wealth taxes (right panel) for various redistribution preferences (zero represents no desire for redistribution; see footnote 20).

(a) if $\sigma > 1$, the zero-tax steady state is locally saddle-path stable;

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

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

The first two points confirm our theoretical and numerical observations. For $\sigma > 1$ the solution is saddle-path stable, explaining why it does not converge to the zero-tax steady state—except for the knife-edged cases where there is no desire for redistribution, in which case the tax rate is zero throughout. For $\sigma < 1$ the solution converges to the zero tax steady state whenever redistribution towards workers is desirable. This lends theoretical support to our numerical findings for $\sigma < 1$, discussed earlier and illustrated in Figure 1.

The third point raises a distinct possibility which is not our focus: the system may become unstable or feature cyclical dynamics. This is consistent with Kemp et al. (1993), who also studied the linearized system around the zero-tax steady state. They reported the potential for local instability and cycles, applying the Hopf Bifurcation Theorem. Proposition 5 clarifies that a necessary condition for this dynamic behavior is $\sigma < 1$ and $\gamma > \gamma^*$. The latter condition is equivalent to a desire to redistribute away from workers towards capitalists. We have instead focused on low values of γ that ensure that desired redistribution runs from capitalists to workers. For this reason, our results are completely distinct to those in Kemp et al. (1993).

3 Representative Agent Ramsey

In the previous section we worked with the two-class model without government debt from Judd (1985). Chamley (1986), in contrast, studied a representative agent Ramsey model with unconstrained government debt; Judd (1999) adopted the same assumptions. This section presents results for such representative agent frameworks.

We first consider situations where the upper bounds on capital taxation do not bind in the long run. We then prove that these bounds may, in fact, bind indefinitely.

3.1 First Best or Zero Taxation of Zero Wealth?

In this subsection, we first review the discrete-time model and zero capital tax steady state result in Chamley (1986) and then present a new result. We show that if the economy settles down to a steady state where the bounds on the capital tax are not binding, then the tax on capital must be zero. This result holds for general recursive preferences that, unlike time-additive utility, allow the rate of impatience to vary. Non-additive utility constituted an important element in Chamley (1986), to ensure that zero-tax results were not driven by an "infinite long-run elasticity of savings".[21] However, we also show that other implications emerge away from additive utility. In particular, if the economy converges to a zero-tax steady state there are two possibilities. Either private wealth has been wiped out, in which case nothing remains to be taxed, or the tax on labor also falls to zero, in which case capital income and labor income are treated symmetrically. These implications paint a very different picture, one that is not favorable to the usual interpretation of zero capital tax results.

Preferences. We write the representative agent's utility as $\mathcal{V}(U_0, U_1, \ldots)$ with per period utility $U_t = U(c_t, n_t)$ depending on consumption c_t and labor supply n_t. Assume that utility $\mathcal{V}$ is increasing in every argument and satisfies a Koopmans (1960) recursion

$$V_t = W(U_t, V_{t+1}) \tag{5a}$$

$$V_t = \mathcal{V}(U_t, U_{t+1}, \ldots) \tag{5b}$$

$$U_t = U(c_t, n_t). \tag{5c}$$

[21] 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.

Here $W(U, V')$ is an aggregator function. We assume that both $U(c, n)$ and $W(U, V')$ are twice continuously differentiable, with $W_U, W_V, U_c > 0$ and $U_n < 0$. Consumption and leisure are taken to be 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. It will prove useful below to note that the strict monotonicity of $\mathcal{V}$ immediately implies that $\bar{\beta}(V) \in (0, 1)$ at any steady state with utility V.[22]

Technology. The economy is subject to the sequence of resource constraints

$$c_t + k_{t+1} + g_t = F(k_t, n_t) + (1 - \delta)k_t \qquad t = 0, 1, \ldots \tag{6}$$

where F is a concave, differentiable and constant returns to scale production function taking as inputs labor n_t and capital k_t, and the parameter $\delta \in [0, 1]$ is the depreciation rate of capital. The sequence for government consumption, g_t, is given exogenously.

Markets and Taxes. Labor and capital markets are perfectly competitive, yielding before tax wages and rates of return given by $w_t^* = F_n(k_t, n_t)$ and $R_t^* = F_k(k_t, n_t) + 1 - \delta$.

The agent maximizes utility subject to the sequence of budget constraints

$$c_0 + a_1 \leq w_0 n_0 + R_0 k_0 + R_0^b b_0,$$
$$c_t + a_{t+1} \leq w_t n_t + R_t a_t \qquad t = 1, 2, \ldots,$$

and the No Ponzi condition $\frac{a_{t+1}}{R_1 R_2 \cdots R_t} \to 0$. The agent takes as given the after-tax wage w_t and the after-tax gross rates of return, R_t. Total assets $a_t = k_t + b_t$ are composed of capital k_t and government debt b_t; with perfect foresight, both must yield the same return

[22] A positive marginal change dU in the constant per period utility stream increases steady state utility by some constant $d\mathcal{V}$. By virtue of (5a) this implies $d\mathcal{V} = W_U dU + W_V d\mathcal{V}$, which yields a contradiction unless $W_V < 1$.

in equilibrium for all $t = 1, 2, \ldots$, so only total wealth matters for the agent; this is not true for the initial period, where we allow possibly different returns on capital and debt. The after-tax wage and return relate to their before-tax counterparts by $w_t = (1 - \tau_t^n) w_t^*$ and $R_t = (1 - \tau_t)(R_t^* - 1) + 1$ (here it is more convenient to work with a tax rate on net returns than on gross returns).

Importantly, we allow for a bound on the capital tax rate: $\tau_t \leq \bar{\tau}$ for some $\bar{\tau} > 0$. As is well understood, without upper bounds on capital taxation the solution involves extraordinarily high initial capital taxation, typically complete expropriation, unless the first best is achieved first. Taxing initial capital 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, $\tau_t \leq \bar{\tau}$. In some cases we follow Chamley (1986) and assume $\bar{\tau} = 1$; in this case the constraint amounts to the restriction that $R_t \geq 1$.[23]

Planning problem. The *implementability condition* for this economy is

$$\sum_{t=0}^{\infty} (\mathcal{V}_{ct} c_t + \mathcal{V}_{nt} n_t) = \mathcal{V}_{c0} \left(R_0 k_0 + R_0^b b_0 \right), \tag{7}$$

whose derivation is standard. In the additive separable utility case $\mathcal{V}_{ct} = \beta^t U_{ct}$ and $\mathcal{V}_{nt} = \beta^t U_{nt}$ and expression (7) reduces to the standard implementability condition popularized by Lucas and Stokey (1983) and Chari et al. (1994). Given R_0 and R_0^b, any allocation satisfying the implementability condition and the resource constraint (7) can be sustained as a competitive equilibrium for some sequence of prices and taxes.[24]

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

$$\mathcal{V}_{ct} = R_{t+1} \mathcal{V}_{ct+1}, \tag{8a}$$

$$R_t = (1 - \tau_t)(F_{kt} - \delta) + 1, \tag{8b}$$

$$\tau_t \leq \bar{\tau}. \tag{8c}$$

The planning problem maximizes $\mathcal{V}(U_0, U_1, \ldots)$ subject to (6), (7) and (8). In addition, we take R_0^b as given. 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

[23] 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.

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

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). *Suppose the optimum converges to an interior steady state where the constraints on capital taxation are asymptotically slack. Let $\tilde{\Lambda}_t = \mathcal{V}_{ct}\Lambda_t$ denote the multiplier on the resource constraint (6) in period t. Suppose further that the multiplier Λ_t converges to an interior point $\Lambda_t \to \Lambda > 0$. Then the tax on capital converges to zero $\frac{R_t}{R_t^*} \to 1$.*

The proof is straightforward. Consider a sufficiently late period t, so that the bounds on the capital tax rate are no longer binding. Then the first-order condition for k_{t+1} includes only terms from the resource constraint (6) and is simply $\tilde{\Lambda}_t = \tilde{\Lambda}_{t+1}R^*_{t+1}$. Equivalently, using that $\tilde{\Lambda}_t = \mathcal{V}_{ct}\Lambda_t$ we have

$$V_{ct}\Lambda_t = \mathcal{V}_{ct+1}\Lambda_{t+1}R^*_{t+1}.$$

On the other hand the representative agent's Euler equation (8a) is

$$V_{ct} = \mathcal{V}_{ct+1}R_{t+1}.$$

The result follows from combining these last two equations.[25]

The main result of this subsection is stated in the next proposition. Relative to Theorem 3, we make no assumptions on multipliers and prove that the steady-state tax rate is zero. More importantly, we derive new implications of reaching an interior steady state.

Proposition 6. *Suppose the optimal allocation converges to an interior steady state and assume the bounds on capital tax rates are asymptotically slack. Then the tax on capital is asymptotically zero. In addition, if the discount factor is locally non-constant at the steady state, so that $\bar{\beta}'(V) \neq 0$, then either*

[25] Chamley (1986) actually worked with the particular bound $\bar{\tau} = 1$, implying a constraint on returns $R_t \geq 1$. For $\bar{\tau} = 1$ it is enough to assume that the multiplier Λ_t converges in the limit and there is no need to require the bounds on capital taxation not to bind. The reason is that in this case the constraints imposed by (8) do not involve k_{t+1}, so the argument above goes through unchanged. This is essentially the form that Theorem 1 in Chamley (1986) takes, although the assumption of converging multipliers is not stated explicitly, but imposed within the proof.

In fact, with $\bar{\tau} = 1$, as long as the multiplier Λ_t converges, one does not even need to assume the allocation converges to arrive at the zero-tax conclusion. This is essentially the argument used by Judd (1999). However, the problem is that one cannot guarantee that the multiplier converges. We shall discuss this in subsection 3.3.

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

(b) *the allocation converges to the first-best, with a zero tax rate on labor.*

This result shows that at any interior steady state where the bounds on capital taxes do not bind, the tax on capital is zero; this much basically echoes Chamley (1986), or our rendering in Theorem 3. However, as long as the rate of impatience is not locally constant, so that $\bar{\beta}'(V) \neq 0$, the proposition also shows that this zero tax result comes with other implications. There are two possibilities. In the first possibility, the capital income tax base has been driven to zero—perhaps as a result of heavy taxation along the transition. In the second possibility, the government has accumulated enough wealth—perhaps aided by heavy taxation of wealth along the transition—to finance itself without taxes, so the economy attains the first best. Thus, capital taxes are zero, but the same is true for labor taxes.

To sum up, if the economy converges to an interior steady state, then either both labor and capital are treated symmetrically or there remains no wealth to be taxed. Both of these implications do not sit well with the usual interpretation of the zero capital tax result. To be sure, in the special (but commonly adopted) case of additive separable utility one can justify the usual interpretation where private wealth is spared from taxation and labor bears the entire burden. However, this is no longer possible when the rate of impatience is not constant. In this sense, the usual interpretation describes a knife edged situation.

3.2 Long Run Capital Taxes Binding at Upper Bound

We now show that the bounds on capital tax rates may bind forever, contradicting a claim by Chamley (1986). This claim has been echoed throughout the literature, e.g. by Judd (1999), Atkeson et al. (1999) and others.

For our present purposes, and following Chamley (1986) and Judd (1999), it is convenient to work with a continuous-time version of the model and restrict attention to additively separable preferences,[26]

$$\int_0^\infty e^{-\rho t} U(c_t, n_t) dt. \tag{9a}$$

$$U(c,n) = u(c) - v(n) \qquad \text{with} \qquad u(c) = \frac{c^{1-\sigma}}{1-\sigma}, \quad v(n) = \frac{n^{1+\zeta}}{1+\zeta}, \tag{9b}$$

[26]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.

where $\sigma, \zeta > 0$. Following Chamley (1986), we adopt an iso-elastic utility function over consumption; this is important to ensure 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

$$c_t + \dot{k}_t + g = f(k_t, n_t) - \delta k_t, \tag{10}$$

where f is concave, homogeneous of degree one and differentiable. For simplicity, government consumption is taken to be constant at $g > 0$. We denote the before-tax net interest rate by $r_t^* = f_k(k_t, n_t) - \delta$. The implementability condition is now

$$\int_0^\infty e^{-\rho t} \left(u'(c_t)c_t - v'(n_t)n_t \right) = u'(c_0)a_0, \tag{11}$$

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

$$\dot{\theta}_t = \theta_t(\rho - r_t), \tag{12a}$$

$$r_t = (1 - \tau_t)(f_k(k_t, n_t) - \delta), \tag{12b}$$

$$\tau_t \leq \bar{\tau}, \tag{12c}$$

for some bound $\bar{\tau} > 0$, where $\theta_t = u'(c_t)$ denotes the marginal utility of consumption. The planning problem maximizes (9a) subject to (10), (11) and (12).

Chamley (1986, Theorem 2, pg. 615) formulated the following claim regarding the path for capital tax rates.[27]

Claim. *Suppose $\bar{\tau} = 1$ and that preferences are given by (9). 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) $T < \infty$.

We do not dispute (a) and (b). 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)."[28] Our next result shows that there is nothing absurd about this within the

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

[28] 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."

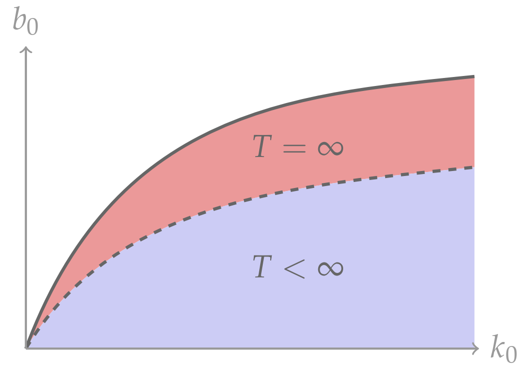

Figure 4: Graphical representation of Proposition 7.

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 7. *Suppose $\bar{\tau} = 1$ and preferences are given by (9) with $\sigma > 1$. Fix any initial capital $k_0 > 0$. There exist $\underline{b} < \bar{b}$ such that for initial government debt $b_0 \in [\underline{b}, \bar{b}]$ the optimum has $\tau_t = \bar{\tau}$ for all $t \geq 0$, while for $b_0 > \bar{b}$ there is no equilibrium.*

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 $T = \infty$ even 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.

Proposition 8. *Suppose that $\bar{\tau} = 1$, that preferences are given by (9) 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.*

Under the conditions stated in the proposition the solution converges to zero in a homogeneous, constant growth rate fashion. This explicit example illustrates that convergence takes place, but not to an interior steady state. This latter property is more general: at least with additively separable utility, whenever indefinite taxation of capital is optimal, $T = \infty$, no interior steady state exists.

To see why this is the case consider first the case with $\bar{\tau} = 1$. Then the after tax interest rate is zero whenever the bound is binding. Since the agent discounts the future positively this prevents a steady state. In contrast, when $\bar{\tau} < 1$ the before-tax interest rate may be positive and the after tax interest rate equal to the discount rate, $(1 - \bar{\tau})r^* = \rho$, the condition for constant consumption. This suggests the possibility of a steady state. However, we must also verify whether labor, in addition to consumption, remains constant. This, in turn, requires a constant labor tax. Yet, one can show that under the assumptions of Proposition 7, but allowing $\bar{\tau} < 1$, we must have

$$\partial_t \tau_t^n = (1 - \tau_t^n)\tau_t r_t^*,$$

implying that the labor tax strictly rises over time whenever the capital tax is positive, $\tau_t > 0$. This rules out a steady state. Intuitively, the capital tax inevitably distorts the path for consumption, but the optimum attempts to undo the intertemporal distortion in labor by varying the tax on labor. We conjecture that the imposition of an upper bound on labor taxes solves the problem of an ever-increasing path for labor taxes, leading to the existence of *interior* steady states with positive capital taxation.

3.3 Revisiting Judd (1999)

Up to this point we have focused on the Chamley-Judd zero-tax results. A follow-up 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. It provides a variant of the result in Chamley (1986) without requiring the allocation to converge to a steady state. The paper also offers a connection between capital taxation and rising consumption taxes to provide an intuition for zero-tax results. Let us consider each of these two points in turn.

Bounded Multipliers and Zero Average Capital Taxes. The main result in Judd (1999) can be restated using our continuous-time setup from Section 3.2. With $\bar{\tau} = 1$, the planning problem maximizes (9a) subject to (10), (11), (12a), and (12b). Let $\hat{\Lambda}_t = \theta_t \Lambda_t$ denote the co-state for capital, that is, the current value multiplier on equation (10), satisfying $\dot{\hat{\Lambda}}_t = \rho \hat{\Lambda}_t - r_t^* \hat{\Lambda}_t$. Using that $\dot{\hat{\Lambda}}_t / \hat{\Lambda}_t = \dot{\theta}_t / \theta_t + \dot{\Lambda}_t / \Lambda_t$ and $\dot{\theta}_t / \theta_t = \rho - r_t$ we obtain

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

If Λ_t converges then $r_t - r_t^* \to 0$. Thus, the Chamley (1986) steady state result actually follows by postulating the convergence of Λ_t, without assuming convergence of the allocation. Judd (1999, pg. 13, Theorem 6) goes down this route, but assumes that the endogenous multiplier Λ_t remains in a bounded interval, instead of assuming that it converges.

Theorem 4 (Judd, 1999). *Let $\theta_t \Lambda_t$ denote the (current value) co-state for capital in equation (10) 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,

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

In particular, under the conditions of this theorem, the optimum cannot converge to a steady state with a positive tax on capital.[29] More generally, the condition requires departures of r_t from r_t^* to average zero.

Note that our proof proceeded without any optimality condition except the one for capital k_t.[30] In particular, we did not invoke first-order conditions for the interest rate r_t nor for the tax rate on capital τ_t. Naturally, this poses two questions. Do the bounds on Λ_t essentially assume the result? And are the bounds on Λ_t consistent with an optimum?

Regarding the first question, we can say the following. The multiplier $e^{-\rho t}\hat{\Lambda}_t$ represents the planner's (time 0) social marginal value of resources at time t. Thus,

$$\text{MRS}_{\text{t,t+s}}^{\text{Social}} = e^{-\rho s}\frac{\hat{\Lambda}_{t+s}}{\hat{\Lambda}_t} = e^{-\int_0^s r_{t+\tilde{s}}^* d\tilde{s}}$$

represents the marginal rate of substitution between t and $t+s$, which, given the assumption $\overline{\tau} = 1$, is equated to the marginal rate of transformation. The private agent's marginal

[29] 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.

[30] 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.

rate of substitution is

$$\text{MRS}^{\text{Private}}_{\text{t,t+s}} = e^{-\rho s}\frac{\theta_{t+s}}{\theta_t} = e^{-\int_0^s r_{t+\tilde{s}} d\tilde{s}},$$

where θ_t represents marginal utility. It follows, by definition, that

$$\text{MRS}^{\text{Social}}_{\text{t,t+s}} = \frac{\Lambda_{t+s}}{\Lambda_t} \cdot \text{MRS}^{\text{Private}}_{\text{t,t+s}}.$$

This expressions shows that the rate of growth in Λ_t is, by definition, equal to the wedge between social and private marginal rates of substitution. Thus, the wedge $\frac{\Lambda_{t+s}}{\Lambda_t} = e^{\int_0^s (r_{t+\tilde{s}} - r^*_{t+\tilde{s}}) d\tilde{s}}$ is the only source of nonzero taxes. Whenever Λ_t is constant, social and private MRSs coincide and the intertemporal wedge is zero, $r_t = r^*_t$; if Λ_t is enclosed in a bounded interval, the same conclusion holds on average.

These calculations afford an answer to the first question posed above: assuming the (average) rate of growth of Λ_t is zero is tantamount to assuming the (average) zero long-run tax conclusion. We already have an answer to the second question, whether the bounds are consistent with an optimum, since Proposition 7 showed that indefinite taxation may be optimal.

Corollary. *At the optimum described in Proposition 7 we have that $\Lambda_t \to 0$ as $t \to \infty$. Thus, in this case the assumption on the endogenous multiplier Λ_t adopted in Judd (1999) is violated.*

There is no guarantee that the endogenous object Λ_t remains bounded away from zero, as assumed by Judd (1999), making Theorem 4 inapplicable.

Exploding Consumption Taxes. Judd (1999) also offers an intuitive interpretation for the Chamley-Judd result based on the observation that an indefinite tax on capital is equivalent to an ever-increasing tax on consumption. This casts indefinite taxation of capital as a villain, since rising and unbounded taxes on consumption appear to contradict standard commodity tax principles, as enunciated by Diamond and Mirrlees (1971), Atkinson and Stiglitz (1972) and others.

The equivalence between capital taxation and a rising path for consumption taxes is useful. It explains why prolonging capital taxation comes at an efficiency cost, since it distorts the consumption path. If the marginal cost of this distortion were increasing in T and approached infinity as $T \to \infty$ this would give a strong economic rationale against indefinite taxation of capital. We now show that this is not the case: the marginal cost remains bounded, even as $T \to \infty$. This explains why a corner solution with $T = \infty$ may be optimal.

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 $r^*, w^* > 0$.

Proposition 9. *Suppose utility is given by (9), with $\sigma > 1$. Suppose technology is linear. Then the solution to the planning problem can be obtained by solving to the following static problem:*

$$\begin{aligned} \max_{T,c,n} \quad & u(c) - v(n), \\ s.t. \quad & (1 + \psi(T))\, c + G = k_0 + \omega n, \\ & u'(c)c - v'(n)n = (1 - \tau(T))u'(c)a_0, \end{aligned} \tag{13}$$

where $\omega > 0$ is proportional to w^*; G is the present value of government consumption; and, c and n are measures of lifetime consumption and labor supply, respectively. The functions ψ and τ are increasing with $\psi(0) = \tau(0) = 0$; ψ is bounded away from infinity and τ is bounded away from 1. Moreover, the marginal trade-off between costs (ψ) and benefits (τ) from extending capital taxation

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

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, $\frac{d\psi}{d\tau} < \infty$, 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). Their guiding principles are, therefore, ill adapted to these settings. In particular, formulas based on local elasticities do not apply, without further modification.

Effectively, a bound on capital taxation restricts the path for the consumption tax to lie below a straight line going through the origin. In the short run, the consumption tax is constrained to be near zero; to compensate, it is optimal to set higher consumption taxes in the future. As a result, it may be optimal to set consumption taxes as high as possible at all times. This is equivalent to indefinite capital taxation.

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

$$(1-\sigma) \sum_{t=0}^{\infty} \beta^t U(C_t) = U'(C_0) R_0 (k_0 + b_0). \tag{14}$$

Anticipated Confiscatory Taxation. For $\sigma > 1$ the left hand side in equation (14) 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 10. *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, 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 11. *Consider the two-class model from Section 2 but with unrestricted government bonds and in continuous time. 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, Chamley (1986) and Judd (1985). Our findings contradict well-established results and their standard interpretations. We showed that, provided the intertemporal elasticity of substitution (IES) is less than one, the long run tax on capital is actually positive. Empirically, an IES below one is considered most plausible.

Why were the proper conclusions missed by Judd (1985), Chamley (1986) and many others? Among other things, these papers assume that the endogenous multipliers associated with the planning problem converge. Although this seems natural, we have shown that this is not necessarily true at the optimum. In fact, on closer examination it is evident that presuming the convergence of multipliers is equivalent to the assumption that the intertemporal rates of substitution of the planner and the agent are equal. This then implies that no intertemporal distortion or tax is required. Consequently, analyses based on these assumptions amount to little more than assuming zero long-run taxes.

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 if debt is low enough.[31] In this paper we refrain from making any such claim, one way or another. We confined our attention to the original theoretical zero-tax results, widely perceived as delivering ironclad conclusions that are independent of parameter values or initial conditions. Based on our results, we have found little basis for such an interpretation.

References

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

[31] Any quantitative exercise could also evaluate the welfare gains from different policies. For example, even when $T < \infty$ is optimal, the optimal value of T may be very high and indefinite taxation, $T = \infty$, may closely approximate the optimum. One can also compare various non-optimal simple policies, such as never taxing capital versus always taxing capital at a fixed rate.

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.

Milgrom, Paul and Ilya Segal, "Envelope theorems for arbitrary choice sets," *Econometrica*, 2002, *70* (2), 583–601.

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 Recursive Formulation of (1a)

In our numerical simulations, we use a recursive representation of the Judd (1985) economy. 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

$$\begin{aligned} V(k, C_-) = &\max_{c \geq 0, (k', C) \in A} \{u(c) + \gamma U(C) + \beta V(k', C)\} \\ & c + C + k' + g = f(k) + (1-\delta)k \\ & \beta U'(C)(C + k') = U'(C_-)k \\ & c, C, k' \geq 0. \end{aligned} \tag{15}$$

Here, A is the *feasible* set, that is, states (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. At $t = 0$, capital k_0 is given, so 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}. If V is differentiable, the first order condition is

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

Since one can show that $\mu_t = V_C(k_t, C_{t-1})U''(C_{t-1})k_t$, this is akin to the condition $\mu_0 = 0$ in equation (2a).[32]

B 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 in that part are somewhat cumbersome and lengthy, so they are relegated to the end of this section to ensure greater readability. Finally, in the third part,

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

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

subject to

$$\begin{aligned} C_0 + k_1 &= R_0 k_0 \\ c_0 + C_0 + k_1 &= f(k_0) + (1-\delta)k_0 \\ c_0, C_0, k_1 &\geq 0. \end{aligned}$$

This alternative gives rise to similar results.

we use our geometric results to prove Proposition 3. Readers interested only in the main steps of the proof are advised to jump straight to the third part.

B.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 (15) of problem (1a) will sometimes be abbreviated by z, and a set of states by Z. The total state space is denoted by $Z_{\text{all}} \subset \mathbb{R}^2_+$ and is defined below. It will prove useful at times to express the set of constraints in (15) as

$$k' = x - C_- \left(\frac{\beta x}{k}\right)^{1/\sigma} \tag{16a}$$

$$C = C_- \left(\frac{\beta x}{k}\right)^{1/\sigma} \tag{16b}$$

$$C_-^{\sigma/(\sigma-1)} \left(\frac{\beta}{k}\right)^{1/(\sigma-1)} \leq x \leq f(k) + (1-\delta)k - g, \tag{16c}$$

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) $z_{t+1} \in \Gamma(z_t)$ for all $t \geq 0$, 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 $C_- = (1-\beta)/\beta k$. For very low and high capital levels k, steady states turn out to be infeasible, but notice that all others are *self-generating*, $z \in \Gamma(z)$. Similarly, a set Z is called *self-generating* if every $z \in Z$ is generated by a sequence in Z. Denote by Z^* (= A in the notation above) 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

$$k = \underbrace{f(k) + (1-\delta)k - g}_{\equiv F(k)} - \frac{1-\beta}{\beta} k. \tag{17}$$

Demanding that $k^g > k_g$ is tantamount to specifying $F'(k^g) < 1/\beta < F'(k_g)$. Equation (17) 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 (17) need not have two solutions, not even a single one, in which case government con-

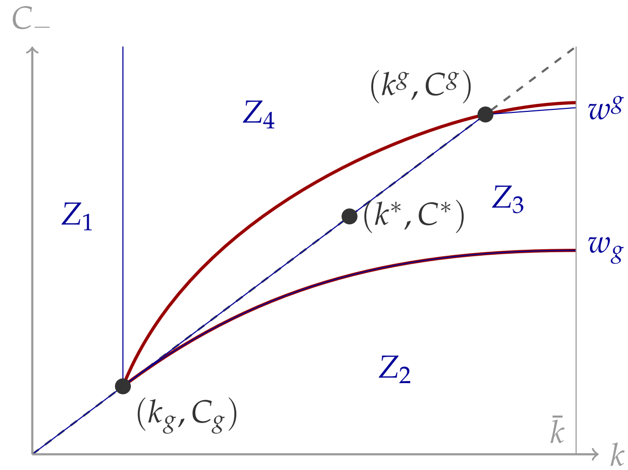

Figure 5: The state space Z_{all}, including the feasible set Z^* (between the two red curves), and all sets Z_i (separated by the blue curves). The point (k^*, C^*) is the zero-tax steady state. Showing that this is the qualitative shape of the feasible set Z^* is an integral part of the proof.

sumption 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 (17) is violated for any $k \notin [k_g, k^g]$.

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 outline of this proof is as follows. In Section B.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 B.3.

B.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

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

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

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

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

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 B.4.1. □

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

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

Proof. See Subsection B.4.2. □

Lemma 2 finishes the characterization of all feasible states with $C_- \leq w^g(k)$. 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 $\bar{w}(k)$ such that states with $C_- > w^g(k)$ are feasible if and only if $C_- \leq \bar{w}(k)$.

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 (16a) and (16b), 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,

$$Tv(k) = \sup\{C_- \mid \exists x \in (0, F(k)] : v(k'(x,k,C_-)) \geq C(x,k,C_-)\}, \tag{18}$$

where recall that $F(k) = f(k) + (1-\delta)k - g$, as in (17). 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 B.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 B.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 B.4.5. □

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

Proof. See Subsection B.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 feasible 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 B.4.7. □

B.3 Proof of Proposition 3

Armed with the results from Section B.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 $\{k_{t+1}, C_t\}$ *optimal path*, if the 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 the 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 Lemma 1 in Section B.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 $c_{t_s} \geq \underline{c}$ for all s. Along these indices, the FOC for capital (2d) implies

$$\underbrace{u'(c_{t_s})}_{\leq u'(\underline{c})}(f'(k_{t_s}) + (1-\delta)) = \frac{1}{\beta} \underbrace{u'(c_{t_s - 1})}_{\geq 0} + \underbrace{U'(c_{t_s - 1})}_{\text{bounded}} \cdot \underbrace{(\mu_{t_s} - \mu_{t_s - 1})}_{\geq \text{const} \cdot \mu_{t_s - 1} \to \infty},$$

and so $k_{t_s} \to 0$ for $s \to \infty$, which is impossible within the feasible set Z^* because it violates $k \geq k_g$ (see Lemma 1). □

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 an 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 $c_{ss} > 0$. Note that such an interior steady state can be reached without leaving the interior of the feasible set, since by Lemmas 2 and 7, hitting the upper or lower boundary once means convergence to a non-interior steady state.[33]

[33]Note that hitting the right boundary at $k = \bar{k}$ (other than with k_0) is of course not feasible due to depreciation.

Now take an interior state (k, C_-) and suppose the optimal control was $c = 0$. Then, by the FOC for capital (2d),[34] 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 of states with optimal control $c = 0$, $\{c = 0\}$ for short, lies on the boundary of Z^*. By Lemmas 2 and 5 this means that $\{c = 0\}$ is exactly equal to the top boundary $\{C_- = \bar{w}(k)\}$. An optimal path which approaches $\{c = 0\}$ must then share the same limiting behavior as states in the set $\{c = 0\}$.[35] By virtue of Lemma 7, it must either converge to (k_g, C_g) or (k^g, C^g). □

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.

Lemma 10. *If an optimal path $\{k_{t+1}, C_t\}$ converges to (k^g, C^g), then the value function V is locally decreasing in C at each 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 $t \geq T$.[36] 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,

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

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

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

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

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

[35]The formal reason for this is that by Berge's Maximum Theorem, the optimal policy c is upper hemicontinuous in the state. Because $c = 0$ along the boundary, and the path converges to the boundary, its policy c converges arbitrarily close to 0. Therefore, it can only converge to where states in $\{c = 0\}$ are converging.

[36]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 $F'(k^g) < 1/\beta$.

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. □

And finally, Lemma 11 proves Proposition 3.

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

Proof. By Lemma 9 it is sufficient to prove that an optimal path does not converge to (k^g, C^g). Suppose the contrary held and there was an 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[37]

$$d\epsilon_t = \left(1 - \frac{1}{\sigma}\frac{C_t}{x_t}\right)^{-1} d\epsilon_{t-1}. \tag{19}$$

Observe that (19) is precisely the relation which ensures that the variation satisfies all the constraints of the system (in particular (16b) of which (19) 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, an optimal path converges to (k_g, C_g). □

[37] Notice that $x_t = C_t + k_{t+1} \geq C_t$ by definition of x_t, and $\sigma > 1$. Hence this expression is well defined.

B.4 Proofs of Auxiliary Lemmas

B.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 (16c). 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 $(C_-/C_g)^\sigma < k/k_g$.[38] 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.[39] Then,

$$h(k', C) = \frac{k'}{C^\sigma} = \frac{k'}{C_-^\sigma \beta x/k} = \frac{k}{C_-^\sigma}\underbrace{\frac{k'}{\beta(k'+C)}}_{>1} > h(k, C_-). \tag{20}$$

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 (20) 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.[40] However, the lowest feasible steady state is (k_g, C_g) and since $(C_-/C_g)^\sigma < k/k_g$,

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

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

[39] Note that if $C \geq (1-\beta)/\beta k'$, then $k' < k_g$. The reason is as follows: The constraints (16a) and (16b) 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 $C \geq (1-\beta)/\beta k'$, 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 $\sigma > 1$).

[40] 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.

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, $Z_2^* = \emptyset$. □

B.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 $(C_-/C^g)^{\sigma} \leq k/k^g$.[41] 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 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 $C_- < (1-\beta)/\beta k$.[42] 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, $C < (1-\beta)/\beta k'$,[43] and so we can recycle equation (20) 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

[41] These inequalities hold for all $k \geq k_g$. The proofs are analogous to the proofs in footnotes 38 and 42.

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

[43] This holds because by the IC constraint (1c), $\beta(k' + C)/C^{\sigma} = k_g/C_g^{\sigma}$ 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.

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. □

B.4.3 Proof of Lemma 3

Proof. First note that T can be rewritten as

$$Tv(k) = \max\{C_- \mid v(k'(F(k), k, C_-)) = C(F(k), k, C_-)\}. \tag{21}$$

There are two ways in which (21) differs from (18):

- Suppose that the supremum in (18) 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

$$\Phi: x \mapsto \underbrace{\psi\left(k'(x,k,C_-), C(x,k,C_-)\right)}_{\text{constant in } x} - \underbrace{\psi\left(k'(x,k,C_-), v(k'(x,k,C_-))\right)}_{\text{decreasing in } x} \tag{22}$$

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

- Suppose that the supremum in (18) 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 (21) 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 $C > C_g$.[44]

[44] $C > C_g$ 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 $C > C_g$.

- To show that $Tv(k)$ is increasing note that $\psi(k'(\dots), C(\dots)) = \beta^{-1}k/C_-^\sigma$ 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 $k \in [k_g, \bar{k}]$.[45]

- 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

$$\psi(k_1', C_1) = \beta^{-1}\frac{k_1}{Tv(k_1)^\sigma} < \beta^{-1}\frac{k_2}{Tv(k_2)^\sigma} = \psi(k_2', C_2). \tag{23}$$

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 (23). 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 $k'(\dots)$ is decreasing and $C(\dots)$ is increasing in C_-, it follows that $k'(F(k^g), k^g, C_-) < k^g$ but $C(F(k^g), k^g, C_-) >$

[45] 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.

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.

□

B.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 $T^i w^g(k) \geq T^j w^g(k)$.[46] □

B.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, $\bar{w}$ must be weakly increasing, $\bar{w}(k_g) = C_g$, $\bar{w}(k^g) = C^g$, and $\psi(k, \bar{w}(k))$ 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 $\bar{w}(k_1)$ and $\bar{w}(k_2)$. 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 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 (21) – this implies that $v(k') < C$ for any $c > 0$, by a similar logic as in (22)). □

B.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 $\bar{w}$) for $k_{t+1} \leq k^g$. If $k_{t+1} > k^g$ it must be the case that

[46] 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 $Z^{(i)} \supseteq Z^{(j)}$.

$k_{s+1} > k^g$ for all $s < t$ as well.[47] But then, using $x_t \le F(k_t) < k_t/\beta$,

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

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 $k_t \le \bar{k}$). 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_-) 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 $h(\hat{k}_{t+1}, \hat{C}_t) \ge 0$, and similarly the second equation is increasing in $F(\hat{k}_t)$ if $\hat{k}_{t+1} \ge 0$. By induction over t, if $h(\hat{k}_t, \hat{C}_{t-1}) \ge h(k_t, C_{t-1})$ and $\hat{k}_t \ge k_t$ (induction hypothesis), then, because $F(\hat{k}_t) \ge x_t$,

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

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

- if $\psi(k', C) \ge \psi(k^g, C^g)$, 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 $\{C \le \bar{w}(k)\}$. 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 $k' \le F(k) - C_- \left(\frac{\beta F(k)}{k}\right)^{1/\sigma} < F(k^g) - C^g \left(\frac{\beta F(k^g)}{k^g}\right)^{1/\sigma} = k^g$.

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 $\{\hat{k}_{t+1}, \hat{C}_t\}$ 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. □

B.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.[48] 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 $k > k_g$,[49] 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[50]. 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 $\{\hat{k}_{t+1}, \hat{C}_t\}$ denote the path with $c_t = 0$, starting from $(k_T, \bar{w}(k_T))$. 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

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

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

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

$dL_C/dk < 0$, $dL_C/dC_- > 0$. This implies that because $C_{T-1} > \bar{w}(k_T)$ (which must hold or else $C_- \leq \bar{w}(k)$ by construction of $\bar{w}$), $C_t > \hat{C}_t$ and $k_{t+1} > \hat{k}_{t+1}$ for all $t \geq T$. Moreover, borrowing from equation (20), 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,

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

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, $(k_{t+1}, C_t) \not\to (k^g, C^g)$. □

C Numerical Method

To solve the Bellman equation (15) we must first compute the feasible set A (or Z^* in the notation above). We restrict the range of capital to a closed interval $[\underline{k}, \bar{k}]$ with $\underline{k} \geq k_g$. This leads us to seek a subset $A^{\underline{k}} \subset A$ of the feasible set 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 $A_0, A_1, A_2 \ldots$ 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 sequence $V_0, V_1, \ldots$ 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.

D Proof of Proposition 4

First, we define the following object,

$$\omega_\tau = \frac{dW_\tau}{dk_{\tau+1}} = \sum_{\tau' \geq \tau+1} \beta^{\tau'-\tau} u'(c_{\tau'})(F'(k_{\tau'}) - R_{\tau'}) \left(\prod_{s=\tau+1}^{\tau'-1} S_{I,s} R_s \right), \tag{24}$$

which corresponds to the response in welfare W_τ, 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

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

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

$$\omega_t S_{I,t} - u'(c_t) = -\frac{1}{k_t} \sum_{\tau=0}^{t-1} \beta^{\tau-t} \omega_\tau S_{\tau,R_t}. \tag{25}$$

By optimization over the initial interest rate R_0, we find the condition

$$\omega_0 S_{I,0} k_0 - u'(c_0) k_0 = 0. \tag{26}$$

Notice that $S_{I,0} > 0$ and so $\omega_0 \in (0, \infty)$. Since the ω_τ satisfy the recursion

$$\omega_\tau = u'(c_{\tau+1})(F'(k_{\tau+1}) - R_{\tau+1}) + \beta S_{I,\tau+1} R_{\tau+1} \omega_{\tau+1},$$

and since it is easy to see that $R_{\tau+1} > 0$ for all τ,[51] it also follows that ω_τ is finite for all τ. Then, due to the recursive nature of (25), if $\omega_\tau > 0$ for $\tau < t$,

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

In particular, using the initial condition (26), this proves by induction that

$$\omega_t S_{I,t} - u'(c_t) \geq 0 \quad \text{for all } t > 0. \tag{27}$$

Now suppose the economy were converging to an interior steady state with non-positive limit tax (either zero or negative), that is, $\Delta_t \equiv F'(k_t) - R_t$ converges to a non-positive number, $c_t \to c > 0$, and $S_{I,t}R_t \to S_I R > 0$. Note that it is immediate by (24) that if Δ_t converges to a negative number, then ω_t must eventually become negative—contradicting (27). Hence suppose $\Delta_t \to 0$. Distinguish two cases.

Case I: Suppose $\prod_{s=1}^{\tau}(\beta S_{I,s} R_s)$ is unbounded or converging to a number in $(0, \infty)$. Then, because ω_0 is finite, we have that the partial sums converge to zero,

$$\bar{\omega}_\tau \equiv \sum_{\tau' \geq \tau+1} \beta u'(c_{\tau'})(F'(k_{\tau'}) - R_{\tau'}) \prod_{s=1}^{\tau'-1} (\beta S_{I,s} R_s) \to 0, \quad \text{as } \tau \to \infty.$$

Hence,

$$\omega_\tau = \left(\prod_{s=1}^{\tau} (\beta S_{I,s} R_s) \right)^{-1} \bar{\omega}_\tau \to 0,$$

contradicting the fact that ω_t is bounded away from zero by $u'(c)/S_I$. In fact notice that when $\beta S_I R > 1$, then notice that by (24) ω_0 can never be finite unless $\Delta_t \to 0$. Thus we proved that $\beta S_I R > 1$ is not compatible with any interior steady state.

Case II: Now suppose $\beta S_I R < 1$. In this case, we show convergence of ω_τ to zero directly. Fix $\epsilon > 0$. Let τ large enough such that $\beta S_{I,s} R_s < b$ for some $b < 1$ and that $|u'(c_{\tau'})\Delta_{\tau'}| < \epsilon(1-b)$. Then,

$$|\omega_\tau| \leq \sum_{\tau'=\tau+1} \epsilon(1-b) b^{\tau'-1-\tau} = \epsilon.$$

Again, this contradicts the fact that ω_t is bounded away from zero by $u'(c)/S_I$.

[51] Otherwise capital would be zero forever after due to $S(0, \dots) = 0$, a contradiction to the allocation converging to an interior steady state.

E Derivation of the Inverse Elasticity Rule (4) and Proof of the Corollary

Consider equation (25). Because $\beta S_I R < 1$, ω_τ converges to

$$\omega = \frac{\beta}{1 - \beta S_I R}(F'(k) - R)u'(c).$$

Suppose $\omega = 0$, or equivalently a zero limit tax $\mathcal{T} = 0$, and regard equation (25). We make the additional convergence assumption

$$\sum_{\tau=1}^{t} \beta^{-\tau} \frac{\omega_{t-\tau} k_{t-\tau}}{\omega_t k_t} \epsilon_{S_{t-\tau}, R_t} \to \sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau} \in [-\infty, \infty], \quad \text{as } t \to \infty, \tag{28}$$

which amounts to first taking the limit of the summands as $t \to \infty$, and then taking the limit of the series, instead of considering both limits simultaneously. Under this order of limits assumption, we can characterize the limit of equation (25) as $t \to \infty$,

$$\underbrace{S_{I,t} - \frac{u'(c_t)}{\omega_t}}_{\to \pm\infty} = -\underbrace{\sum_{\tau=1}^{t} \beta^{-\tau} \frac{\omega_{t-\tau} k_{t-\tau}}{\omega_t k_t} \epsilon_{S_{t-\tau}, R_t}}_{\to \sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau}}.$$

This implies that $\sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau}$ is either plus or minus infinity, which is compatible with the inverse elasticity formula. Next consider the case $\omega \neq 0$. Again, by taking the limit of (25) as $t \to \infty$ and using the condition (28), we find

$$S_I - \frac{u'(c)}{\omega} = -\sum_{\tau=1}^{\infty} \beta^{-\tau} \epsilon_{S,\tau},$$

which can be rewritten as

$$\frac{\beta S_I R}{1 - \beta S_I R}(F'(k) - R) - R = -\frac{1}{1 - \beta S_I R}(F'(k) - R) \sum_{\tau=1}^{\infty} \beta^{-\tau+1} \epsilon_{S,\tau}.$$

Note that $F'(k) - R = \frac{\mathcal{T}}{1-\mathcal{T}} R$. Therefore, we can rearrange the condition to

$$\begin{aligned} \frac{\beta S_I R}{1 - \beta S_I R} - \frac{1 - \mathcal{T}}{\mathcal{T}} &= -\frac{1}{1 - \beta S_I R} \sum_{\tau=1}^{\infty} \beta^{-\tau+1} \epsilon_{S,\tau} \\ \Rightarrow \quad \mathcal{T} &= \frac{1 - \beta R S_I}{1 + \sum_{t=1}^{\infty} \beta^{-t+1} \epsilon_{S,t}}. \end{aligned}$$

Proof of the Corollary. Notice that by Proposition 4 the limit tax rate is positive, $\mathcal{T} > 0$, conditional on convergence to an interior steady state. If now the inverse elasticity formula implies a negative tax rate, then either the regularity condition for the inverse elasticity rule is not satisfied or the allocation does not converge to an interior steady state.

F Infinite Sum of Elasticities with Recursive Utility

Proposition 12. *Suppose capitalists have recursive preferences represented by* (5a) *(see Section 3.1, with $U = c$) then at any zero tax steady state*

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

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

Proof. For this exercise, suppose capitalists' utility is characterized by the recursion $V_t = W(C_t, V_{t+1})$, assuming W is twice continuously differentiable and strictly increasing in both arguments. Suppose after-tax interest rates R_t converge to some $R > 0$ and C_t, a_t converge to positive values. Note that because utility is strictly increasing in a permanent increase in consumption at the steady state, we have $W_V \in (0, 1)$ (see also footnote **??** below).

The conditions for optimality are then,

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

The first equation is the recursion for utility V_t and the second equation is the Euler equation. In particular, note that the latter implies that $RW_V = 1$ at the steady state. Linearizing these equations around the steady state (denoted without time subscripts) yields,

$$W_V\, dV_{t+1} = -W_C R\, da_t + W_C\, da_{t+1} + dV_t - W_C a\, dR_t \tag{29}$$

and

$$\begin{aligned}(RW_C W_{VC} - RW_{CC} - W_{CC})\, da_{t+1} + W_{CC}\, da_{t+2} - (W_V W_C + W_{CC} a)\, dR_{t+1} \\ + (W_{CV} - RW_C W_{VV})\, dV_{t+1} - W_{CV}\, dV_{t+2} \\ = (R^2 W_C W_{VC} - W_{CC} R)\, da_t + (RW_C W_{VC} a - W_{CC} a)\, dR_t,\end{aligned} \tag{30}$$

where all derivatives are evaluated at the steady state $((R-1)\, a, V)$. We solve (29) and (30) by the method of undetermined coefficients, guessing

$$da_{t+1} = \overline{\lambda}\, da_t + \sum_{s=0}^{\infty} \theta_s\, dR_{t+s} \tag{31a}$$

$$dV_t = W_C R\, da_t + (W_C a) \sum_{s=0}^{\infty} W_V^s\, dR_{t+s}. \tag{31b}$$

The form of equation (31a) is what is required by the Envelope condition. We are left to find $\overline{\lambda}$ and the sequence $\{\theta_s\}$. Substituting the guesses (31a) and (31b) into (30), 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 for $\overline{\lambda}$,

$$\overline{\lambda}^2 + \left(\frac{2RW_C W_{VC} - W_{CC}(1+R) - R^2 W_C^2 W_{VV}}{W_{CC} - W_{VC} W_C R} \right) \overline{\lambda} + R = 0. \tag{32}$$

Note that in the additive separable case (when $W(C, V)$ is linear in V), $\bar{\lambda} = 1$ is a solution. Setting the coefficient on dR_t to zero gives

$$\begin{aligned}
\theta_0 &= \frac{(RW_C W_{VC} - W_{CC})\, a}{2RW_C W_{VC} - W_{CC}(1+R) + (W_{CC} - RW_{CV} W_C)\, \overline{\lambda} - R^2 W_C^2 W_{VV}} \\
&= \frac{\overline{\lambda}\,(RW_C W_{VC} - W_{CC})\, a}{-R\,(W_{CC} - W_{VC} W_C R)} = \overline{\lambda} \frac{a}{R}.
\end{aligned}$$

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

$$\theta_1 = W_V \overline{\lambda} \theta_0 + \overline{\lambda} W_V \frac{W_V^2 + \left(\frac{W_{CC}}{W_C} + R^* W_C W_{VV} - W_{CV} \right) W_V a}{W_{VC} - \frac{W_V}{W_C} W_{CC}},$$

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

$$\theta_s = W_V \overline{\lambda} \theta_{s-1} + \overline{\lambda}\,(W_V)^s\,(1 - W_V)\, \frac{W_{VC} + \frac{W_C}{1 - W_V} W_{VV}}{W_{VC} - \frac{W_V}{W_C} W_{CC}} a,$$

for $s = 2, 3, \ldots$. The result then 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. □

G Linearized Dynamics and Proof of Proposition 5

A natural way to prove Proposition 5 would be to linearize our first order conditions in (2), and to solve forward for the multipliers μ_t and λ_t using transversality conditions, arriving at an approximate law of motion of the form

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

To maximize similarity with Kemp et al. (1993), however, we do not take that route; rather we start with the continuous time problem, derive its first order conditions and linearize them around the zero tax steady state. The problem in continuous time is

$$\max \int_0^\infty e^{-\rho t} \left(u(c_t) + \gamma U(C_t) \right) dt$$
$$\text{s.t. } c_t + C_t + g + \dot{k}_t = f(k_t) - \delta k_t$$
$$\dot{C}_t = \frac{C_t}{\sigma} \left(\frac{f(k_t)}{k_t} - \delta - \frac{c_t}{k_t} - \rho \right).$$

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

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

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

$$\lim_{t \to \infty} e^{-\rho t} q_t C_t = 0 \quad \text{and} \quad \lim_{t \to \infty} e^{-\rho t} p_t k_t = 0. \tag{33}$$

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,

$$\dot{x}_t = J(x_t - x^*), \tag{34}$$

where the 4×4 matrix J can be written as

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

with 2×2 matrices

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

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

$$\det J = (1 - \sigma) \underbrace{\frac{f'' u'}{u''}}_{>0} \frac{\rho^2}{\sigma^2}, \tag{35}$$

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,

$$\lambda_{1-4} = \frac{\rho}{2} \pm \left[\left(\frac{\rho}{2} \right)^2 - \frac{\delta}{2} \pm \frac{1}{2} \left(\delta^2 - 4 \det J \right)^{1/2} \right]^{1/2}, \tag{36}$$

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,

$$\delta = \frac{\rho}{\sigma} \frac{u' - \gamma U'}{u'' k^*} - \frac{f'' u'}{u''}. \tag{37}$$

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 (34) 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* (36) *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

$$Re\lambda_1 = \lambda_1 > \rho > Re\lambda_2 \geq \rho/2 \geq Re\lambda_4 > 0 > \lambda_3 = Re\lambda_3. \tag{38}$$

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*

$$Re\lambda_1, Re\lambda_2 > \rho > 0 > Re\lambda_4, Re\lambda_3. \tag{39}$$

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 $\text{Re}\left(\sqrt{x+b}\right) \geq \text{Re}\left(\sqrt{x-b}\right)$.[52] From there, it is straightforward to see that $\text{Re}\lambda_1 \geq \text{Re}\lambda_2$ and $\text{Re}\lambda_4 \geq \text{Re}\lambda_3$. Finally $\text{Re}\lambda_2 \geq \rho/2 \geq \text{Re}\lambda_4$ holds according to our convention of square roots having nonnegative real parts.

(b) The negativity of $\det J$ follows immediately from (35). 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 (38) holds, using monotonicity of $\text{Re}\sqrt{x}$ for real numbers x.

(c) The signs of $\det J$ and δ follow immediately from (35) and (37). In this case, $-\delta/2 \pm 1/2\text{Re}\left(\delta^2 - 4\det J\right)^{1/2} > 0$ proving (39).

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

H Proof of Proposition 6

In this proof, we first exploit the recursiveness of the utility $\mathcal{V}$ to recast the IC constraint (7) entirely in terms of V_t and $W(U, V')$. Then, using the first order conditions, we are able to characterize the long-run steady state.

[52] 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.

Let $\beta_t \equiv \prod_{s=0}^{t-1} W_{Vs}$. Using the definition of the aggregator in (3) this implies that $\mathcal{V}_{ct} = \beta_t W_{Ut} U_{ct}$ and $\mathcal{V}_{nt} = \beta_t W_{Ut} U_{nt}$. Thus the IC constraint (7) can be rewritten as

$$\sum_{t=0}^{\infty} \beta_t W_{Ut}(U_{ct}c_t + U_{nt}n_t) = W_{U0}U_{c0}\left(R_0 k_0 + R_0^b b_0\right), \tag{40}$$

and the planning problem becomes

$$\begin{aligned} &\max V_0 \\ &\text{s.t. } V_t = \mathcal{V}(U(c_t, n_t), U(c_{t+1}, n_{t+1}), \ldots) \\ &\quad\quad \text{RC (6), IC (40).} \end{aligned} \tag{41}$$

To state the first order conditions, define $A_t = \frac{\partial}{\partial V_{t+1}} \sum_{s=0}^{\infty} \beta_s W_{Us}(U_{cs}c_s + U_{ns}n_s)$ and $B_t = \sum_{s=0}^{\infty} \frac{\partial(\beta_s W_{Us})}{\partial U_t}(U_{cs}c_s + U_{ns}n_s)$. Let χ_t be the current value multiplier on the Koopmans constraint (41), λ_t the current value multiplier on the resource constraint (6), and μ the multiplier on the IC constraint (40). Defining $\nu_t \equiv \sum_{s=0}^{t} \frac{\beta_t}{\beta_s}\chi_s$ the first order conditions then take the form

$$\begin{aligned} 1 + \nu_0 &= 0 \\ -\nu_t + \nu_{t+1} + \mu\frac{A_t}{\beta_{t+1}} &= 0 \\ -\nu_t W_{Ut} U_{ct} + \mu W_{Ut}\left(U_{ct} + U_{cc,t}c_t + U_{nc,t}n_t\right) + \mu\frac{B_t}{\beta_t}U_{ct} &= \lambda_t \\ \nu_t W_{Ut} U_{nt} - \mu W_{Ut}\left(U_{nt} + U_{cn,t}c_t + U_{nn,t}n_t\right) - \mu\frac{B_t}{\beta_t}U_{nt} &= \lambda_t f_{nt} \\ -\lambda_t + \lambda_{t+1} W_{Vt} f_{kt+1} &= 0, \end{aligned}$$

where we defined $f(k, n) = F(k, n) + (1-\delta)k$ and the notation X_{zt} stands for the derivative of quantity X with respect to z, evaluated at time t. Now suppose the allocation converges to an interior steady state in c, k, and n. Then U_t and V_t converge, as well as their first and second derivatives (when evaluated at c_t, k_t, and n_t). Similarly, the representative agent's assets a_t converge to a value a, which can be characterized using a time $t+1$ version of the IC constraint,

$$\begin{aligned} a = \lim_{t\to\infty} a_{t+1} &= \lim_{t\to\infty} \left(W_{Ut+1}U_{ct+1}\beta_{t+1}R_{t+1}\right)^{-1} \sum_{s=t+1}^{\infty} \beta_s W_{Us}(U_{cs}c_s + U_{ns}n_s) \\ &= \left((1-\beta)U_c R\right)^{-1}\left(U_c c + U_n n\right), \end{aligned}$$

where $\beta \equiv \bar{\beta}(V) = W_V \in (0,1)$ (see footnote 22). Using this representation, we see that

A_t/β_{t+1} converges as well, to some limit A,

$$A_t = \beta_t W_{UV,t}(U_{ct}c_t + U_{nt}n_t) + \beta_t W_{VV,t} \underbrace{\beta_{t+1}^{-1} \sum_{s=t+1}^{\infty} \beta_s W_{Us}(U_{cs}c_s + U_{ns}n_s)}_{W_{Ut+1}U_{ct+1}R_{t+1}a_{t+1} \to W_U U_c R a}$$

$$\begin{aligned} \Rightarrow \quad \frac{A_t}{\beta_{t+1}} &\to \frac{\beta_U}{\beta}(U_c c + U_n n) + \frac{\beta_V}{\beta} W_U U_c R a \\ &= \left(\frac{1-\beta}{W_U}\beta_U + \beta_V \right) \frac{1}{\beta} W_U U_c R a = \frac{\bar{\beta}'(V)}{\beta} W_U U_c R a \equiv A. \end{aligned} \tag{42}$$

where we defined $\beta_X \equiv W_{VX}$ and $X = U, V$. Similarly, we can show that B_t/β_t converges to some finite value B. Taking the limits of quantities in the first order conditions above, we thus find

$$-\nu_t + \nu_{t+1} + \mu A = 0 \tag{43a}$$

$$-\nu_t + \mu \left(1 + \frac{U_{cc}c}{U_c} + \frac{U_{nc}n}{U_c} \right) + \mu \frac{B}{W_U} = \lambda_t \frac{1}{W_U U_c} \tag{43b}$$

$$-\nu_t + \mu \left(1 + \frac{U_{cn}c}{U_n} + \frac{U_{nn}n}{U_n} \right) + \mu \frac{B}{W_U} = -\lambda_t \frac{f_n}{W_U U_n} \tag{43c}$$

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

Note that

$$\beta f_k - 1 = \frac{\lambda_t}{\lambda_{t+1}} - 1 = -\frac{W_U U_c}{\lambda_{t+1}} \mu A. \tag{44}$$

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 equations (43b) and (43c) we find

$$\lambda_t \frac{f_n}{W_U U_n} \tau^n = \mu \left(\frac{U_{cc}c}{U_c} + \frac{U_{nc}n}{U_c} - \frac{U_{cn}c}{U_n} - \frac{U_{nn}n}{U_n} \right), \tag{45}$$

where $\tau^n \equiv 1 + \frac{U_n}{U_c f_n}$ is the steady state tax on labor. By normality of consumption and labor the term in brackets is negative, $\frac{U_{cc}c}{U_c} + \frac{U_{nc}n}{U_c} - \frac{U_{cn}c}{U_n} - \frac{U_{nn}n}{U_n} < 0$.

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 $\tau^n = 0$. By (43a) and $\mu > 0$, this requires $A \leq 0$.
- Case B: $\nu_t \to \nu \in \mathbb{R}$, then $\lambda_t \to \lambda$ by (43b). By (43a) and $\mu > 0$, this requires $A = 0$. There are two subcases to consider. If $\lambda \neq 0$ then $\tau^n \neq 0$ is possible. If instead $\lambda = 0$, then (45) implies 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 steady state with $\tau^n = 0$. This case requires $A \geq 0$.

What are the condition for Case B with $\tau^n > 0$? This requires $A = 0$, which, according to (42), implies that we must have either $\bar{\beta}'(V) = 0$ or $a = 0$. In sum, this proves that if $\bar{\beta}'(V) \neq 0$, then at any interior steady state of the problem, either the agent's assets are zero, $a = 0$, or the economy is at first best, $\tau^n = 0$.

I Proof of Proposition 7

For the whole proof we fix some positive initial level of capital $k_0 > 0$. The problem under scrutiny is

$$V(b_0) \equiv \max_{\{c_t, n_t, k_t\}} \int_0^\infty e^{-\rho t} \left(u(c_t) - v(n_t) \right), \tag{46a}$$

$$\dot{c}_t \geq -\frac{\rho}{\sigma} c_t, \tag{46b}$$

$$c_t + g_t + \dot{k}_t = f(k_t, n_t) - \delta k_t, \tag{46c}$$

$$\int_0^\infty e^{-\rho t} \left(u'(c_t) c_t - v'(n_t) n_t \right) \geq u'(c_0)(k_0 + b_0), \tag{46d}$$

where recall that $u(c) = c^{1-\sigma}/(1-\sigma)$ and $v(n) = n^{1+\zeta}/(1+\zeta)$. Note that the value function $V(b_0)$ is decreasing in b_0. Problem (46a) has the following necessary first order conditions

$$\Phi_v^W v'(n_t) = \lambda_t f_n(k_t, n_t), \tag{47a}$$

$$\dot{\eta}_t - \rho \eta_t = \eta_t \frac{\rho}{\sigma} + \lambda_t - \Phi_u^W u'(c_t), \tag{47b}$$

$$\dot{\lambda}_t = (\rho - r_t^*) \lambda_t, \tag{47c}$$

$$\eta_0 = -\mu \sigma c_0^{-\sigma-1}(k_0 + b_0), \tag{47d}$$

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 $f_k - \delta$.[53] Here, μ is the multiplier on the IC constraint (46d), λ_t is the multiplier of the resource constraint (46c), and η_t denotes the costate of consumption c_t. In this proof we take parts (a) and (b) of the claim on page 25 as given (see proof in Chamley (1986, Theorem 2, pg. 615)). In particular, we take for granted that the optimal capital tax policy is bang-bang, with $\tau_t = \bar{\tau}$ for $t < T$ and $\tau_t = 0$ for $t > T$, where $T \in [0, \infty]$.

If $\eta_t < 0$, then constraint (46b) is binding. Further, if at $T < \infty$ we have $\eta_t = 0$ for

[53] Note that it can be shown that $r_t^* > 0$ by a standard argument: If ever $f_k(k_t, n_t)$ were to drop below δ, then a marginal variation involving a marginal reduction in labor and capital growth $\dot{k}_t$ in the time prior to negative interest rates relax the implementability condition (46d) and resource constraint (46c) while improving the objective (46a).

$t \in [T, T+\varepsilon)$ 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

$$\eta_t < 0 \quad \text{for all } t, \tag{48}$$

$$e^{-\rho t}\eta_t c_t \to 0. \tag{49}$$

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 13. *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. In that case we already noted that $\lambda_T = \Phi_u^W u'(c_T)$ implying that $\lambda_T < 0$. This contradicts the FOC for labor (47a) at $t = T$. Therefore, $T = \infty$. □

It is convenient to characterize a restricted problem, where T is required to be infinite. Effectively, this implies that constraint (46b) holds with equality throughout and the path of c_t is entirely characterized by c_0. To this end we define the minimum discounted sum of labor disutilities needed to sustain this path $\{c_t\}$ as

$$\begin{aligned} \tilde{v}(c_0) \equiv \min_{\{c_t, n_t, k_t\}} & \int_0^\infty e^{-\rho t} v(n_t) \\ \text{s.t. } & c_t + g_t + \dot{k}_t \leq f(k_t, n_t) - \delta k_t \\ & c_t = c_0 e^{-\rho/\sigma t}. \end{aligned}$$

Notice that $\tilde{v}$ is bounded away from zero, and differentiable, strictly increasing and convex in c_0. Next, define the restricted problem

$$V_\infty(b_0) \equiv \max_{c_0 > 0} u(c_0)\frac{\sigma}{\rho} - \tilde{v}(c_0) \tag{50a}$$

$$c_0\frac{\sigma}{\rho} - c_0^\sigma(1+\zeta)\tilde{v}(c_0) \geq k_0 + b_0. \tag{50b}$$

We obtained (50a) from the original problem (46a) by requiring that $T = \infty$ and using the definition of $\tilde{v}$. Also, we divided the IC condition (46d) by $u'(c_0)$ to arrive at (50b). Note that, by construction, $V_\infty \leq V$, but whenever $T = \infty$ is optimal, $V_\infty = V$. Our next result characterizes the restricted value function V_∞.

Lemma 14. *There exists a level of initial debt $\bar{b}$ such that a solution to the restricted planner's problem (50a) exists for all $b_0 \leq \bar{b}$. The restricted value function $V_\infty : (-\infty, \bar{b}] \to \mathbb{R}$ is weakly decreasing, concave, and differentiable. Moreover, $\lim_{b \nearrow \bar{b}} V'_\infty(b) = -\infty$.*

Proof. First, define

$$\bar{b} \equiv \max_{c_0} c_0 \frac{\sigma}{\rho} - (1+\zeta) c_0^{\sigma} \tilde{v}(c_0) - k_0. \tag{51}$$

Given the constraint (50b), the constraint set is non-empty if and only if $b_0 \leq \bar{b}$. Then, notice that $V_\infty(b_0)$ is weakly decreasing because the constraint set for c_0 is decreasing in b_0. Also, the constraint (50b) describes a convex set in (b_0, c_0) space, and the objective (50a) is strictly concave. Therefore, V_∞ is concave and the maximizer $c_0^*(b_0)$ for a given b_0 is unique.

Due to the properties of $\tilde{v}$ mentioned above, $\bar{c} \equiv c_0^*(\bar{b}) > 0$ is the unique maximizer of (51), satisfying the following first order condition,

$$\frac{\sigma}{\rho} \bar{c}^{-\sigma} = (1+\zeta)\sigma \bar{c}^{-1} \tilde{v}(\bar{c}) + (1+\zeta)\tilde{v}'(\bar{c}).$$

Notice that at $c_0 = \bar{c}$, the derivative of the objective function is positive,

$$u'(\bar{c})\frac{\sigma}{\rho} - \tilde{v}'(\bar{c}) = (1+\zeta)\sigma \bar{c}^{-1}\tilde{v}(\bar{c}) + \zeta \tilde{v}'(\bar{c}) > 0,$$

implying that $c_0^*(b_0) > \bar{c}$ for all $b_0 < \bar{b}$. Let $\tilde{\mu}$ be the Lagrange multiplier on the constraint (50b) in the restricted problem (50a). By the necessary first order condition for c_0 we have that

$$u'(c^*)\frac{\sigma}{\rho} - \tilde{v}'(c^*) + \tilde{\mu}\left(\frac{\sigma}{\rho}(c^*)^{-\sigma} - (1+\zeta)\sigma(c^*)^{-\sigma}\tilde{v}(c^*) - (1+\zeta)\tilde{v}'(c^*)\right) = 0, \tag{52}$$

for $c^* = c^*(b_0)$ and any $b_0 < \bar{b}$. This shows that there is always a unique multiplier $\tilde{\mu}$, characterized by (52) since the factor multiplying $\tilde{\mu}$ is nonzero as $c_0^*(b_0) > \bar{c}$ for all $b_0 < \bar{b}$. Therefore we can apply the Envelope Theorem and obtain that the restricted value function $V_\infty(b_0)$ is differentiable at every $b_0 < \bar{b}$, with $V_\infty'(b_0) = -\tilde{\mu}$.

Now, note that there exists a unique maximizer for any Lagrangian of the form,

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

with $\tilde{\mu} \in [0, \infty)$, even for very large values of $\tilde{\mu}$. Thus there also exists a value of b_0 satisfying the constraint (50b), so the image of V_∞' must span all values $(-\infty, 0]$. By concavity, this implies that $\lim_{b \nearrow \bar{b}} V'(b) = -\infty$. □

Lemma 14 provides a characterization of the problem conditional on $T = \infty$. As a corollary, for $b_0 \leq \bar{b}$ the constraint set of the original problem (46a) is nonempty as well. In a next lemma we prove that in fact the constraint set of the original problem is nonempty *if and only if* $b_0 \leq \bar{b}$. Moreover, for $b_0 = \bar{b}$ only allocations with $T = \infty$ are inside this constraint set, implying that $V(\bar{b}) = V_\infty(\bar{b})$.

Lemma 15. *Take $b_0 \in \mathbb{R}$. The constraints (46b), (46c), (46d) define a non-empty set for $\{c_t, n_t, k_t\}$ if and only if $b_0 \leq \bar{b}$. Moreover, if $b_0 = \bar{b}$ then necessarily $T = \infty$, implying*

$V(\bar{b}) = V_\infty(\bar{b})$.

Proof. It suffices to show that the constraint set in the original problem is empty for $b_0 > \bar{b}$, and that $T = \infty$ is necessary for $b_0 = \bar{b}$. We show both by proving that any $b_0 \geq \bar{b}$ is infeasible with $T < \infty$.

Hence fix some $b_0 \geq \bar{b}$ and assume it was achievable with $T < \infty$. Then, the process for consumption at any optimum is governed by

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

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 the fastest possible 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 left hand side of (46d) is strictly decreasing in c_t and n_t, this strictly relaxes the IC constraint. Hence,

$$\int_0^\infty e^{-\rho t} \hat{c}_t^{1-\sigma} - \int_0^\infty e^{-\rho t} v(\hat{n}_t) > \hat{c}_0^{-\sigma}(k_0 + b_0).$$

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

$$\hat{c}_0^{1-\sigma} \frac{\sigma}{\rho} - (1+\zeta)\tilde{v}(\hat{c}_0) > \hat{c}_0^{-\sigma}(k_0 + b_0).$$

By definition of $\bar{b}$ this is a contradiction to $b_0 \geq \bar{b}$. Therefore, $\bar{b}$ is the highest sustainable debt level in the original problem and can only be achieved with $T = \infty$. □

This lemma is very useful because it shows that $V(\bar{b}) = V_\infty(\bar{b})$. Given that $V(b_0) \geq V_\infty(b_0)$ for all b_0 this means that V also becomes infinitely steep close to $\bar{b}$.

To show that there is an interval $[\underline{b}, \bar{b}]$ with $\underline{b} < \bar{b}$ for which $T = \infty$ is optimal, or equivalently $V = V_\infty$, we assume to the contrary that there exists an increasing 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 $(b, \bar{b})$ of $\bar{b}$. To prove a contradiction, we would like to use the Envelope Theorem to link the local behavior of V to what we know about the μ multiplier from Lemma 13. In order to be able to do so, notice that the value function $V(b_0)$ is the value of a convex maximization problem. To see this, rewrite the problem in terms of $u_t \equiv u(c_t)$ and $v_t \equiv v(n_t)$,[54]

$$V(b_0) \equiv \max_{u_t,v_t,k_t} \int_0^\infty e^{-\rho t} \left(u_t - v_t\right), \tag{53}$$
$$\dot{u}_t \geq (\sigma - 1)\frac{\rho}{\sigma} u_t,$$
$$((1-\sigma)u_t)^{-1/(\sigma-1)} + g_t + \dot{k}_t \leq f\left(k_t, ((1+\zeta)v_t)^{1/(1+\zeta)}\right) - \delta k_t,$$
$$\int_0^\infty e^{-\rho t} \left((1-\sigma)u_t - (1+\zeta)v_t\right) \geq ((1-\sigma)u_0)^{\sigma/(\sigma-1)} (k_0 + b_0).$$

Because the resource constraint and IC of this problem are strictly convex, the Lagrangian of this maximization problem is strictly concave. Therefore, there is a globally unique maximizer, which varies smoothly in b_0 by Berge's Maximum Theorem. Applying the Envelope theorem of Milgrom and Segal (2002, Corollary 5), we obtain that V is absolutely continuous, with

$$V(b_0) = V(\bar{b}) + \int_{b_0}^{\bar{b}} \mu(b) c_0(b)^{-\sigma} db,$$

for $b_0 \leq \bar{b}$. By optimality of $T < \infty$ and Lemma 13, the multiplier $\mu(b)$ is bounded from above by $1/(\sigma-1)$ for all $b \in \{V \neq V_\infty\}$; moreover $c_0(b)$, with $b \in [b_n, \bar{b}]$ is 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 $c_0\sigma/\rho - (1+\zeta)c_0^\sigma \tilde{v}(c_0) = k_0 + b_n$, and n is large enough to ensure that $\underline{c}_0$ is positive.[55] Therefore, $\mu(b)c_0(b)^{-\sigma} \leq \frac{\underline{c}_0^{-\sigma}}{\sigma-1}$ for all $b \in \{V \neq V_\infty\} \cap [b_n, \bar{b}]$. It follows that for any $b_0 \in \{V \neq V_\infty\} \cap [b_n, \bar{b}]$ sufficiently close to $\bar{b}$, such that $V'_\infty(b_0) < -\frac{\underline{c}_0^{-\sigma}}{\sigma-1}$,

$$V(b_0) = V(\bar{b}) - \int_{\{V \neq V_\infty\} \cap [b_0, \bar{b}]} \underbrace{\left(-\mu(b)c_0(b)^{-\sigma}\right)}_{> V'_\infty(b)} db - \int_{\{V = V_\infty\} \cap [b_0, \bar{b}]} \underbrace{\left(-\mu(b)c_0(b)^{-\sigma}\right)}_{= V'_\infty(b)} db$$
$$< V(\bar{b}) + \int_{\bar{b}}^{b_0} V'_\infty(\tilde{b}) d\tilde{b} = V_\infty(b_0),$$

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

[54] Note that we can without loss of generality relax the resource constraint given that $\lambda_t > 0$ at the optimum, see (47a).

[55] To see that this is indeed a lower bound, write the IC constraint as $g(c_0, T) = (k_0 + b_0)$. Notice that in the proof of Lemma 15 we showed that $g(\cdot, T)$ can only increase as we move to $T = \infty$ (and simultaneously optimize over labor with $\tilde{v}$), where $g(c_0, \infty) = c_0\frac{\sigma}{\rho} - (1+\zeta)c_0^\sigma \tilde{v}(c_0)$ is strictly concave and $g(0, \infty) = 0$. Therefore, for $b_0 \geq b_n$, the optimal choice c_0 in the original problem must be bounded from below by the smaller of the two solutions to $g(c_0, \infty) = k_0 + b_n$.

J Proof of Proposition 8

We proceed by solving the necessary first order conditions to problem (46a) and invoking a transversality condition. Noting that the problem is strictly convex, see (53), 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 $\Phi_v^W \equiv 1 + \mu(1+\zeta)$. Solving the necessary FOCs,

$$\begin{aligned} \Phi_v^W v'(n_t) &= \lambda_t f_n(k_t, n_t) \\ c_t + \dot{k}_t &= f(k_t, n_t) - \delta k_t \\ \int e^{-\rho t}\left(c_t^{1-\sigma} - n_t^{1+\zeta}\right) &= c_0^{-\sigma}(k_0 + b_0), \end{aligned} \tag{54}$$

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 $w^* = f_n(k_0, n_0)$,

$$\begin{aligned} r^* &= \zeta\frac{\rho}{\sigma} + \rho \\ w^* &= f_n\left(1, g\left(\zeta\frac{\rho}{\sigma} + \rho + \delta\right)\right) \\ n_0 &= k_0 g\left(\zeta\frac{\rho}{\sigma} + \rho + \delta\right) \\ c_0 &= k_0\left[f\left(1, g\left(\zeta\frac{\rho}{\sigma} + \rho + \delta\right)\right) - \delta + \frac{\rho}{\sigma}\right] \\ b_0 &= c_0\frac{\sigma}{\rho} - \frac{1}{\rho + (1+\zeta)\rho/\sigma} c_0^{\sigma} n_0^{1+\zeta} - k_0. \end{aligned}$$

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, $\dot{\eta}_t - \rho\eta_t = \eta_t\frac{\rho}{\sigma} + \lambda_t - \Phi_u^W u'(c_t)$, and the transversality condition, $e^{-\rho t}\eta_t c_t \to 0$,

$$\eta_t = -\frac{\lambda_0}{\rho + (1+\zeta)\rho/\sigma} e^{-\zeta\rho/\sigma t} + \frac{\sigma}{\rho}\Phi_u^W c_0^{-\sigma} e^{\rho t}.$$

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

$$\begin{aligned} \eta_0 &= -\frac{\lambda_0}{\rho + (1+\zeta)\rho/\sigma} + \frac{\sigma}{\rho}\Phi_u^W c_0^{-\sigma} \\ \eta_0 &= -\mu\sigma c_0^{-\sigma-1}(k_0 + b_0) \\ \Phi_v^W n_0^{\zeta} &= \lambda_0 w^*, \end{aligned}$$

and the inequality $\Phi_u^W \leq 0$ 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[56]

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

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.

K Proof of Proposition 9

First, we show that the planner's problem is equivalent to (13). Then we show that the functions $\psi(T)$ and $\tau(T)$ 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,

$$\begin{aligned} \max \int e^{-\rho t}\left(u(c_t) - v(n_t)\right) & \quad (55) \\ \text{s.t. } \dot{c} \geq c\frac{1}{\sigma}((1-\bar{\tau})r^* - \rho) & \\ \int e^{-r^* t}(c_t - w^* n_t) + G = k_0 & \\ \int e^{-\rho t}\left[(1-\sigma)u(c_t) - (1+\zeta)v(n_t)\right] \geq u'(c_0)a_0, & \end{aligned}$$

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,

$$n_t = n_0 e^{-(r^*-\rho)t/\zeta}. \quad (56)$$

Part (a) and (b) of the claim on page 25 imply the existence of $T \in [0, \infty]$ such that

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

$\tau_t = \bar{\tau}$ 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

$$c_t = c_0 e^{-\frac{\rho-\bar{r}}{\sigma}t + \frac{r^*-\bar{r}}{\sigma}(t-T)^+}. \tag{57}$$

Substituting equations (56) and (57) into (55), the planner's problem simplifies to,

$$\begin{aligned} \max_{T,c_0,\bar{n}} \quad & \psi_1(T)u(c_0) - \psi_3 v(n_0) \\ \text{s.t. } & \psi_2(T)(\chi^*)^{-1}c_0 + G = k_0 + \psi_3 w^* n_0 \\ & \psi_1(T)u'(c_0)c_0 - \psi_3 v'(n_0)n_0 = \chi^* u'(c_0)a_0, \end{aligned} \tag{58}$$

where $\psi_1(T) = \frac{\chi^*}{\chi}\left(1-e^{-\chi T}\right) + e^{-\chi T}$, $\psi_2(T) = \frac{\chi^*}{\hat{\chi}}\left(1-e^{-\hat{\chi} T}\right) + e^{-\hat{\chi} T}$, $\psi_3 = \chi^*\left(r^* + \frac{r^*-\rho}{\zeta}\right)^{-1}$ 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

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

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 $\omega n \equiv w^*\psi_3^{\zeta/(1+\zeta)}n$. Here, we note that by definition, ψ is bounded away from infinity and τ is bounded away from 1. Then, we can rewrite problem (58) as

$$\begin{aligned} \max_{T,c,n} \quad & u(c) - v(n) \\ \text{s.t. } & (1+\psi(T))c + G = k_0 + \omega n \\ & u'(c)c - v'(n)n = (1-\tau(T))u'(c)a_0, \end{aligned}$$

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 ∞.

L Proof of Proposition 11

We proceed as in the first part of the proof of Proposition 7. As in Section 2, labor supply is inelastic at $n_t = 1$. The problem is then

$$\max \int_0^\infty e^{-\rho t} u(c_t), \tag{59a}$$

$$\dot{C}_t \geq -\frac{\rho}{\sigma} C_t, \tag{59b}$$

$$c_t + C_t + \dot{k}_t = f(k_t) - \delta k_t, \tag{59c}$$

$$\int_0^\infty e^{-\rho t} u'(C_t) C_t \geq u'(C_0)(k_0 + b_0). \tag{59d}$$

Problem (59a) has the following necessary first order conditions

$$\dot{\eta}_t - \rho \eta_t = \eta_t \frac{\rho}{\sigma} + \lambda_t - \Phi_u^W U'(C_t), \tag{60a}$$

$$\dot{\lambda}_t = (\rho - f'(k_t) + \delta)\lambda_t, \tag{60b}$$

$$\eta_0 = -\mu\sigma C_0^{-\sigma-1}(k_0 + b_0), \tag{60c}$$

$$u'(c_t) = \lambda_t \tag{60d}$$

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 (59d), λ_t is the multiplier of the resource constraint (59c), and η_t denotes the costate of capitalists' consumption C_t. If $\eta_t < 0$, then constraint (59b) is binding.

Suppose $T < \infty$, in which case we have $\eta_T = 0$ and $\dot{\eta}_T = 0$ (see Lemma .13 in the proof of Proposition 7 for a similar argument). Using the law of motion for η_t, equation (60a), this implies that $\lambda_T = \Phi_u^W U'(C_T) < 0$. This contradictions $\lambda_T = u'(c_T) > 0$ from (60d). Thus, $T = \infty$.