A Robust Theory of Optimal Capital Taxation*

Lukas Mayr†

University of Essex

July 19, 2022

Abstract

I derive a robust condition for the optimality of capital income tax rates that holds across a battery of benchmark macroeconomic models. Applying my theoretical results to US data and disciplining the tax elasticity of wealth with recent quasi-experimental evidence, I find high optimal Rawlsian tax rates of about 90%, because capital tax increases raise the gross return on capital, mitigating the excess burden. At the same time, capital tax hikes depress wages, resulting in lower optimal tax rates from the perspective of households with substantial labor income, the status quo being optimal for households around the 70th income percentile.

JEL Classifications: E62, H21, H24, H25

Keywords: dynamic taxation, Ramsey, sufficient statistics, general equilibrium

*I am very grateful to Dominik Sachs, Árpád Ábrahám and Piero Gottardi, for their guidance and support during the process of writing this paper. It benefited from comments of Gian Luca Clementi, Pawel Doligalski, Sebastian Dyrda, Axelle Ferriere, Jonathan Heathcote, Joachim Hubmer, Dirk Krueger, Elif Kubilay, Andreas Müller, Nicola Pavoni, Markus Riegler, Víctor Ríos-Rull, Petr Sedlacek, Kjetil Storesletten, Yikai Wang, Hakki Yazici, as well as seminar participants at Bristol, Essex, EUI, IFS, Konstanz, LMU Munich, UvA, the Dynamic Macroeconomics Workshop in Vigo, the 8th Annual Mannheim Taxation Conference and the RCEA Conference 2022. Any remaining errors are my own.

†E-mail: lukas.mayr@essex.ac.uk

1 Introduction

The current high degree of economic inequality has spurred the public debate on what the appropriate level of redistribution through the tax system should be. Due to its high concentration at the top, the optimal taxation of capital receives particular attention in this context. Yet, despite a large existing economics literature on this topic, it would be a difficult task to summarize its conclusions to policy makers because the policy prescriptions vary greatly across studies and generally depend on the underlying modeling framework within which they are derived.

In this paper, I derive policy prescriptions that are robust. Within a very rich dynamic general equilibrium environment that nests a battery of important benchmark models as special cases (Judd, 1985; Chamley, 1986; Aiyagari, 1994; Piketty and Saez, 2013; Saez and Stantcheva, 2018), I first show theoretically that the welfare effect of a change in the capital tax rate can always be parsimoniously decomposed as the difference between two components: (i) a 'normative' component, to which I refer as the *equity effect*; it maps the choice of social welfare function to the redistributional gain from capital tax increases; and (ii) a component commonly referred to as the *marginal excess burden*, which is independent of the choice of social welfare function and measures the revenue loss through agents' responses in investment and labor supply. I transparently illustrate the contribution of general equilibrium effects to each of the two welfare components. Given the choice of social welfare function, the capital tax rate is then optimal when the difference between these two components is zero. Consequently, I obtain an intuitive and robust condition for the optimality of capital tax rates.

I apply my theoretical results to US income and wealth data and I discipline the tax-elasticity of households' wealth with most recent quasi-experimental evidence. I find that the endogenous response in the gross return to capital substantially reduces the marginal excess burden, resulting in a high optimal Rawlsian tax rate – which maximizes the welfare of households who live exclusively from government transfers – of about 90%. However, due to the wage depressing effect, middle class households, whose main income source are labor earnings, prefer much lower capital tax rates. The status quo tax rate, calibrated to 41.5%, turns out to be optimal for households in about the 70th percentile of the US income distribution.

To obtain these conclusions, I combine the approaches of two – so far rather disconnected – literatures on optimal capital taxation. The first literature studies capital taxation in parameterized dynamic general equilibrium environments. The aforementioned large variation in policy prescriptions across frameworks can be most easily manifested by a short summary of some of its main contributions. The seminal papers by Judd (1985) and Chamley (1986) prescribe zero optimal capital tax rates in the long run, while Atkeson, Chari, and Kehoe (1999) find that this zero capital tax result holds not only assymptotically but more generally, leading them to conclude that taxing capital is "a bad idea".[1] Straub and Werning (2020), on the other hand, show that in the very same environments, under certain conditions on parameters, the optimal capital tax rate may be positive forever. In environments with incomplete markets, Aiyagari (1995) shows theoretically that the optimal long-run capital tax rate is strictly positive, while Conesa, Kitao, and Krueger (2009) find that it is quantitatively large, leading them to conclude that taxing capital is in fact "not a bad idea after all". Similarly, Domeij and Heathcote (2004), taking into account the welfare effects emanating from transitional dynamics, find large welfare losses from the elimination of capital taxes and an optimal tax rate of about 40%, close to the status quo in the US. Most recently, allowing policy instruments to vary with time, Dyrda and Pedroni (2022) as well as Açıkgöz, Hagedorn, Holter, and Wang (2022) find optimal capital tax rates that are very high in the beginning, at their exogenously set upper bound of 100%, before gradually declining to about 20-25% in the very long run.

Partially motivated by the large variation in results across studies, a different, second, literature emerged, which aims to derive conditions for optimal capital tax rates that are invariant to the assumed structure of the economy (Piketty and Saez, 2012, 2013; Golosov, Tsyvinski, and Werquin, 2014; Saez and Stantcheva, 2018). Following the 'sufficient statistics' approach (Saez, 2001; Chetty, 2009), the welfare effects derived in this literature are not expressed in terms of deep (unmeasured) model primitives but in terms of elasticities that can be directly estimated in the data and thus do not rely on precise modelling assumptions. The optimal tax formulas obtained in these studies highlight the equity-efficiency trade-off around which much of the public debate evolves. Specifically, these studies provide a transparent mapping from social welfare weights to the corresponding 'optimal' tax rates. However, in contrast to the papers in the parametric macroeconomics literature, they assume that factor prices are exogenously given and consequently invariant to tax changes.

My framework keeps the dynamic general equilibrium structure of the macroeconomics literature. Specifically, capital and labor enter complementary in production and factor markets are required to clear, implying that wages and interest rates respond endogenously to tax changes. At the same time, in line with the goal of suffi-

[1] Jones, Manuelli, and Rossi (1997) and Chari, Nicolini, and Teles (2020) provide further generalizations and refinements of the zero capital tax result in complete markets environments. Similarly, Robert Lucas finds that eliminating capital income taxes altogether would lead to "the largest genuinely free lunch [he has] seen in 25 years in business" (Lucas, 1990, p.314).

cient statistics literature, I derive robust tax formulas that are not expressed in terms of model dependent primitives but in terms of elasticities, factor shares, and some distributional statistics that can be directly estimated in the data. The robustness of my theoretical and quantitative results is partially achieved by requiring the policy instruments to be time-invariant. Specifically, I assume an exogenous labor tax schedule, which can take an (almost) arbitrary functional form. The capital tax is linear and allowed to be changed once and for all. The revenue raised through these taxes is partially used to finance an exogenously given spending requirement. Every revenue in excess of this requirement is redistributed lump-sum and equally to all agents.

The exogenous factor price case is nested in my framework as the special case, where capital and labor are perfect substitutes (Piketty and Saez, 2013; Saez and Stantcheva, 2018). In this case the demand for capital is perfectly elastic such that the whole incidence of capital tax increases is borne by the suppliers of capital, that is by savers and investors. As mentioned above, the welfare effect of a marginal increase in the capital tax rate is the difference of two components. In the case with exogenous factor prices, the *equity* effect simply aggregates the mechanical change in net capital- and transfer income of each agent, weighted by the chosen collection of social welfare weights. The *marginal excess burden* is mostly or – absent income effects on labor supply – fully given by the loss in revenue due to a reduction in agents' investment. The optimal tax formula hence balances a classical equity-efficiency trade-off.

By contrast, when capital and labor are complements, a lower capital stock increases the marginal product of capital but reduces the marginal product of labor. Consequently, firms' factor demand is no longer perfectly elastic. Specifically, an increase in the capital income tax rate that reduces the supply of capital ceteris paribus causes an excess demand for capital and an excess supply of labor. To restore market clearing, in general equilibrium the (gross) return on capital increases, while wages fall. These price responses, in turn, redistribute across agents with different income compositions, contributing to the *equity effect*. Specifically, they reduce the income of a large middle class, whose main income source are wages, a 'trickle down' mechanism that reduces optimal capital tax rates if the planner assigns high weight on wage workers.

Importantly, the endogenous price responses also affect the *marginal excess burden* in three ways: First, and most importantly, increasing (gross) interest rates mitigate the fall in the net return to capital and thus moderate the investment decline. This force reduces the excess burden. Second, falling wages affect labor supply; if the substitution effect dominates the income effect, they reduce labor supply, leading to lower labor income tax revenue and thus to an increase in the excess burden. Finally, the factor price changes themselves have a direct impact on revenue. Specifically, when capital is taxed at a higher rate than labor, the simultaneous rise in interest rate and fall in wages increases revenue and thus reduces the marginal excess burden.

The statistics entering these effects can be either directly taken from the data or can be expressed in terms of other statistics for which there exists readily available evidence. An important statistic of the latter category is the net-of-tax elasticity of the equilibrium capital stock. This statistic is a "policy elasticity" in the sense of Hendren (2016). It summarizes the overall reaction of the equilibrium capital stock to tax changes, taking all responses to simultaneous changes in transfers and factor prices into account. I discipline this unmeasured policy elasticity by deriving a one-to-one mapping to recent quasi-experimental estimates of net-of-tax-elasticities of individual wealth, which capture agents' savings responses to tax changes holding transfers and prices fixed (Jakobsen, Jakobsen, Kleven, and Zucman, 2020).

I then move on to a quantitative application of my theoretical results. Using US income and wealth data from the Survey of Consumer Finances 2019, I apply my condition to a sequence of social welfare functions, each of which concentrates the whole welfare weight at one particular percentile of the total income distribution. I compare the policy prescriptions of my optimality condition with the standard condition of the nested exogenous price framework.

I find strong quantitative and qualitative discrepancies between the two cases. Assuming that factor prices are exogenous, one finds that the status quo capital taxes in the US are close to optimal, perhaps slightly too low, for a large part of the US population, about the bottom 60 percent of the income distribution. Absent responses in wages, the welfare effects of capital tax changes are quite homogeneous within this part of the population since even those around the 60th percentile earn very little capital income, which is concentrated among the very high earners.

By contrast, taking into account the endogeneity of factor prices, one finds that the bottom 60 percent of the US income distribution would experience high gains from capital tax increases because endogenous factor price responses significantly reduce the marginal excess burden, from 88 to only 13 cents per dollar of revenue raised mechanically. Most of this reduction is due to the fact that in general equilibrium the above mentioned rise in the gross return to capital significantly mitigates the investment decline. Specifically, the (average discounted) net-of-tax-elasticity of the aggregate capital stock declines from 1.24 in partial- to 0.39 in general equilibrium. Consequently, I find a high optimal Rawlsian tax rate – which maximizes the welfare of households who finance their consumption exclusively through government transfers – of above

90%.[2] Furthermore, rather than being homogeneous, due to their depressing effect on wages, households' welfare gains from capital tax increases are strongly declining in their labor income. For households around the 70th income percentile, the negative effect of the decline in wages just offsets the positive effect of higher government transfers, rendering the satus quo tax rate about optimal for these households.

I contribute to both of the aforementioned literatures. With regards to the 'sufficient statistics' literature, the just described discrepancies reveal that truly robust optimality conditions for capital tax rates require to account for the endogeneity of factor prices. The second self-proclaimed goal of this literature is to "better connect the theory of optimal capital taxation to the policy debate" (Stantcheva, 2020, p.9.21f). However, by treating factor prices as exogenous, this literature assumes away a mechanism that is emphasized in almost every public policy debate on capital taxation. In particular, proponents of 'trickle down' theory advocate the idea that low taxes on the rich may actually benefit the poor. In line with the mechanism described above, they argue that lowering taxes on capital encourages investment, which in turn increases the demand for labor. The resulting increase in wages would benefit poorer households, who tend to receive predominantly labor income. Given that the practical relevance of this mechanism is subject to extensive political discussion, a theory that aims to connect well to the public debate should capture it.

With regards to the macroeconomics literature, I contribute by deriving policy prescriptions that are transparent and robust across a variety of benchmark models. Furthermore, rather than imposing a particular single welfare objective – papers in the macroeconomics literature typically employ the utilitarian objective – I consider a general set of social welfare functions and thus provide a transparent mapping from redistributional preferences to 'optimal' capital tax rates. This provides a more complete picture and more comprehensively informs the policy debate. Finally, my theoretical results inform the calibration of parameterized models by identifying statistics that are relevant for the welfare impact of capital tax changes but to date largely neglected in the parametric macroeconomics literature on optimal capital taxation. For example, virtually all of the papers studying optimal capital taxation in dynamic general equi-

---

[2] Given that the elasticities, which are estimated locally around the current tax system, may change with such large tax increases, I complement my sufficient statistics analysis with a global solution method using a nested parametric model that is calibrated to (locally) replicate all the relevant statistics. I then compute optimal tax rates using global solution methods, taking into account transitional dynamics (as in Domeij and Heathcote, 2004). For most social welfare functions, which prescribe optimal tax rates in the range of 10-85%, the sufficient statistics approach approximates the so obtained tax rates remarkably well. However, outside this range there are slight discrepancies. For example, the optimal Rawlsian tax rate according to the sufficient statistics approach is 98%, while the (more accurate) global solution method prescribes a Rawlsian tax rate of 91%. See Section 5.3 and Appendix E for details.

librium models (implicitly) assume a capital-labor substitution elasticity equal to one.[3] I show that the optimal capital tax rate crucially depends on precisely this elasticity, empirical estimates of which span a broad range with most of it significantly lower than one.[4]

Related Literature

My paper relates to various other strands of the public finance and macroeconomics literature. The importance of endogenous factor price responses is emphasized in a growing recent literature that studies optimal income taxation in frameworks where output is produced with complementary production factors (Rothschild and Scheuer, 2013; Scheuer, 2014; Ales, Kurnaz, and Sleet, 2015; Scheuer and Werning, 2017; Sachs, Tsyvinski, and Werquin, 2020). While in my framework these factors are capital and labor, the latter of which is perfectly substitutable across agents, these models abstract from capital and instead consider different types of labor input that are imperfectly substitutable. All of these papers study optimal income taxation in static Mirrleesian environments, abstracting from the dynamic accumulation process of production factors, in particular of capital.

The New Dynamic Public Finance (NDPF) literature instead considers dynamic Mirrleesian settings with savings (Golosov, Kocherlakota, and Tsyvinski, 2003; Farhi and Werning, 2013). Slavik and Yazici (2014) take explicit account of the complementarity between different types of capital and labor and the implications of general equilibrium spillover effects on wages for optimal capital taxes. In related settings, Thümmel (2020) and Guerreiro, Rebelo, and Teles (2022) study the optimal taxation of robots. Contrary to the present paper, in their numerical applications these papers specify and calibrate concrete parametric functions for model primitives. They do not derive formulas in terms of estimable sufficient statistics as my paper does.

Finally, a paper very similar in spirit to mine is the one by Badel and Huggett (2017), who derive a robust formula for revenue maximizing income tax rates. As the present paper, they find important interaction effects of one tax rate with other tax bases requiring to adjust standard formulas that neglect these interactions.

[3] An exception is Kina, Slavik, and Yazici (2020), who study optimal capital taxation in a framework with capital-skill complementarity.

[4] See Antras (2004), Chirinko (2008) and, more recently, Gechert, Havranek, Irsova, and Kolcunova (2022).

2 The Framework

The theoretical results I derive, hold for a very general framework. Both, to save space and for pedagogical reasons, I present a special case of this model in the main text. Specifically, the model I present here is simplified as much as possible such that it still captures all the relevant economic effects of capital tax changes. In particular, the local welfare effects as well as the optimality condition for capital tax rates will be identical to the one in the more general framework for a large set of social welfare functions, including the most often used utilitarian objective.

The model in the main text nests the economic environments studied in the seminal papers of Judd (1985) and Chamley (1986) as special cases. In addition to their models, here agents have heterogeneous labor productivity. The more general model, in which both working- and investment ability are subject to uninsurable idiosyncratic risk, can be found in Appendix B.

2.1 Households

There is a continuum of infinitely lived agents (dynasties) of measure one. In the simplified model, agents differ only in their initial wealth endowment k_0 and in their working ability $\eta \in [\underline{\eta}, \bar{\eta}]$. Furthermore, in contrast to the more general framework in Appendix B, here ability is assumed to be perfectly persistent. I denote the joint distribution over initial individual states by $\Gamma(k_0, \eta)$.

Given their initial endowment k_0, agents solve

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

subject to the sequence of budget constraints for $t \in \{0, 1, 2, ...\}$,

$$k_{t+1} + c_t = k_t + (1 - \tau_{k,t}) \underbrace{r_t k_t}_{y_t^k(k_0,\eta)} + \underbrace{w_t \eta l_t}_{y_t^l(k_0,\eta)} - \tau_l(w_t \eta l_t) + T_t.$$

Existing capital is denoted by k_t. The income from capital $y_t^k(k_0, \eta)$ is taxed at a linear rate $\tau_{k,t}$, which has a time index because I allow the planner to perform a one-off change in this rate. The agents' gross labor income $y_t^l(k_0, \eta)$ is the product of the wage rate w_t, her working ability η and her labor supply l_t. The function $\tau_l(.)$ is assumed to be twice continuously differentiable and maps gross labor income into labor tax payments. Finally, T_t denotes a lump-sum transfer from the government, which has a time index since I require it to adjust in response to capital tax changes in a way to ensure government budget balance.

The utility function satisfies the following standard assumption.

Assumption 1. *The Bernoulli utility function $u(.,.)$ is twice continuously differentiable in both arguments. For all $(c,l) \geq 0$ it satisfies the conditions $u_c(c,l) > 0$, $u_{cc}(c,l) < 0$, $u_l(c,l) \leq 0$ and $u_{ll}(c,l) < 0$.*

2.2 Firms

In the simplified model, there is a representative price-taking firm, which maximizes profits by choosing capital K_t and labor L_t

$$\max_{K_t \geq 0, L_t \geq 0} \{F(K_t, L_t) - (r_t + \delta)K_t - w_t L_t\},$$

where $\delta \in (0,1)$ is the depreciation rate of capital and the technology $F(.)$ satisfies the following assumption.

Assumption 2. *Denote by k and l effective capital and effective labor, respectively. The production function $F(k,l)$ is twice continuously differentiable and has constant returns to scale. It satisfies for all $(k,l) \geq 0$ the conditions $F_k(k,l) > 0$, $F_l(k,l) > 0$, $F_{kk}(k,l) \leq 0$, $F_{ll}(k,l) \leq 0$ and $F_{kl}(k,l) \geq 0$.*

Equilibrium factor prices are characterized by firms' optimal demand for capital and labor, in particular

$$F_k(K_t, L_t) - \delta = r_t \quad \text{and} \quad F_l(K_t, L_t) = w_t.$$

The sufficient statistics literature on optimal capital taxation assumes that factor prices are invariant to policy changes. Within a general equilibrium framework, where output is produced with capital and labor, this can only be rationalized if the two production factors are assumed to be perfect substitutes. My model captures this special case. Specifically, when $F_{kl}(k,l) = 0$ for all (k,l) my model collapses to the frameworks of Piketty and Saez (2013) and Saez and Stantcheva (2018), allowing for a direct comparison.

2.3 Government

The endogenous policy instruments are a linear capital income tax rate τ_k and a lump-sum transfer T. In addition, there is an exogeneous, twice continuously differentiable labor income tax schedule $\tau_l(.)$ that maps gross labor income into labor tax payments.

Finally, some of the generated tax revenue needs to be allocated to finance a constant stream of government expenditure $G > 0$.

As Saez and Stantcheva (2018), I consider different pre-announcement periods. Specifically, at time $t = 0$ the government announces a change in the capital income tax rate τ_k, which comes into effect after the announcement period $t^a \geq 0$ passed. Formally, the capital income tax rate in period t is given by

$$\tau_{k,t} = \begin{cases} \tau_k^b & \text{for } t < t^a \\ \tau_k^r & \text{for } t \geq t^a, \end{cases}$$

where τ_k^b denotes the pre-existing tax rate in place before the reform, while τ_k^r denotes the tax rate after the reform comes into effect. As is standard in the literature, the government is assumed to be able to commit not to change the tax rate again. Agents have perfect foresight from time $t = 0$ on. During the transition to a new steady state, the transfer T_t is required to adjust in order to ensure period-by-period government budget clearing.

Of course, the 'optimal' allocation which can be achieved depends on the policy instruments which the government has at its' disposal. In particular, different admissible sets of policy instruments may lead to different 'optimal' allocations. The restrictions I impose on the policy instruments have the advantage of resulting in policy prescriptions that are largely invariant to the underlying modelling framework, allowing me to highlight key mechanisms that affect optimal capital tax rates across many different environments and across general sets of economic primitives such as preferences and technology. I view this approach complementary to those which instead allow for very rich sets of policy instruments and study how the policy prescriptions vary across frameworks and what the reasons behind these variations are.[5]

2.4 Equilibrium and Steady State

The equilibrium conditions are standard and a formal definition for the most general framework is presented in Appendix B. Here I focus on the most important elements.

First, both the capital and the labor market need to clear. In particular, in each period $t \in \{0, 1, 2, ...\}$ factor markets need to clear, that is

$$K_t = \int k_t(k_0, \eta) d\Gamma, \qquad \text{and} \qquad L_t = \int \eta l_t(k_0, \eta) d\Gamma.$$

[5] See, for example, Chari et al. (2020), who allow for a very rich set of policy instruments and discuss the conditions on modeling primitives under which the optimal inter-temporal distortion is zero.

Furthermore, the government budget needs to clear,

$$T_t + G = \tau_{k,t} r_t K_t + \int \tau_l(\eta w_t l_t(k_0, \eta)) d\Gamma. \tag{1}$$

In equilibrium, total production of firms $\tilde{Y}_t$ and total household income Y_t are given by, respectively,

$$\tilde{Y}_t = \underbrace{(r_t + \delta)K_t}_{\tilde{Y}_t^k} + \underbrace{w_t L_t}_{Y_t^l} \qquad \text{and} \qquad Y_t = \underbrace{r_t K_t}_{Y_t^k} + \underbrace{w_t L_t}_{Y_t^l}.$$

They differ to the extent that capital depreciates. The distinction between gross- and net factor shares, that is factor shares before and after capital deprecation, is going to be important and shall therefore be made very explicit.

Definition 1. ***Factor Shares.*** *Firms' expenditure shares on capital and labor are defined by, respectively,*

$$\tilde{\alpha}_t^k = \frac{\tilde{Y}_t^k}{\tilde{Y}_t} = \frac{(r_t + \delta)K_t}{(r_t + \delta)K_t + w_t L_t} \qquad \textit{and} \qquad \tilde{\alpha}_t^l = \frac{Y_t^l}{\tilde{Y}_t} = \frac{w_t L_t}{(r_t + \delta)K_t + w_t L_t}.$$

Households' shares of capital and labor income are given by, respectively,

$$\alpha_t^k = \frac{Y_t^k}{Y_t} = \frac{r_t K_t}{r_t K_t + w_t L_t} \qquad \textit{and} \qquad \alpha_t^l = \frac{Y_t^l}{Y_t} = \frac{w_t L_t}{r_t K_t + w_t L_t}.$$

Steady State. In this simplified environment a steady state is simply given when $k_t(k_0, \eta) = k_0$ is time-constant for all (k_0, η).[6] In the analysis below, I follow Saez and Stantcheva (2018) and restrict attention to situations, in which the economy is originally in steady state, as this considerably simplifies the analysis.[7]

Assumption 3. *In period $t = -1$ the economy is in a stationary equilibrium.*

In the following variables without time index refer to their value in the initial steady state.

[6] This is obviously more complicated in the general environment, where productivities are stochastic and time-varying. There, a steady state requires time-invariance of the distribution over individual states. See Definition B.1 in Appendix B.

[7] I refer to their framework in the second part of their paper with concave utility in consumption (their Section 5), which is nested as special case of mine. In the first part of their paper Saez and Stantcheva (2018) assume preferences that are linear in consumption and concave in wealth, implying an immediate jump to the new steady state following a tax change, a behaviour that is inconsistent with the evidence on consumption smoothing (see e.g. Browning and Lusardi, 1996; Browning and Crossley, 2001; Havranek and Sokolova, 2020).

2.5 Special Cases

My general framework nests several important benchmark models as special cases. To some of those I will refer to in the description of the optimality condition as they facilitate the understanding of the most general case. The special cases that are already covered in the simplified model of the main text include:

i. the exogenous factor price model of Section 5 in Saez and Stantcheva (2018): $F_{kl}(K, L) = 0$ for all (K, L);

ii. the dynastic exogenous factor price model of Section 3 in Piketty and Saez (2013): $F_{kl}(K, L) = 0$ for all (K, L);

iii. the neoclassical growth framework of Section 3 in Chamley (1986): degenerate Γ;

iv. the neoclassical growth model with heterogeneous initial wealth in Section 4 of Judd (1985): $\underline{\eta} = \bar{\eta}$;

The more general environment in Appendix B further nests

v. the standard incomplete markets model of Aiyagari (1994);

vi. an incomplete markets model with investment risk on top of labor income risk (Benhabib, Bisin, and Zhu, 2015).

As already discussed, the results presented here carry over to the most general framework for a very broad set of social welfare criteria that includes the utilitarian objective, which is most commonly used in the quantitative macroeconomics literature.

3 Optimal Capital Taxation

Given a collection of Pareto weights $\bar{\omega} = \{\omega(k_0, \eta)\}$ the social planner solves

$$(P) \qquad \max_{\tau_k \leq 1} W(\bar{\omega}) = \max_{\tau_k} \int \omega(k_0, \eta) \sum_{t=0}^{\infty} \beta^t u\big(c_t(k_0, \eta), l_t(k_0, \eta)\big) d\Gamma.$$

Following the sufficient statistics literature, I denote the marginal social welfare weights by $g(k_0, \eta) = \omega(k_0, \eta) u_c(c_0(k_0, \eta), l_0(k_0, \eta))$. Without loss of generality the Pareto weights are normalized such that $\int g(k_0, \eta) d\Gamma = 1$. Hence, $g(k_0, \eta)$ is the planner's relative valuation of a marginal dollar in the hand of agents with characteristics (k_0, η) vs. the equal distribution of this dollar to the whole population.

3.1 Preliminaries

Before studying the welfare effects of tax changes and the condition for the capital tax rate to be optimal, it is useful to define some recurring objects.

Definition 2. ***Income Weighted Marginal Social Welfare Weights.*** *The capital- and labor income weighted marginal social welfare weights are defined by, respectively,*

$$\bar{g}^k = \frac{\int g(k_0,\eta) k_0 d\Gamma}{K} \quad \textit{and} \quad \bar{g}^l = \frac{\int g(k_0,\eta) y^l(k_0,\eta) d\Gamma}{Y^l}.$$

Average marginal social welfare, weighted by labor income and marginal net-of-labor-tax rates is given by

$$\tilde{g}^l = \frac{\int g(k_0,\eta)(1-\tau_l'(y^l(k_0,\eta)))y^l(k_0,\eta)d\Gamma}{(1-\bar{\tau}_l')Y^l},$$

where

$$\bar{\tau}_{l,t}' = \frac{\int y_t^l(k_0,\eta)\tau_l'\left(\eta y_t^l(k_0,\eta)\right)d\Gamma}{Y_t^l}$$

is the labor income weighted average marginal labor tax rate.

The latter definition turns out to be useful when the labor tax schedule is non-linear. However, the intuition of most of the economic effects goes through with linear labor taxes, in which case $\tilde{g}^l = \bar{g}^l$.

Definition 3. ***Policy Elasticities.*** *The elasticity and semi-elasticity of any period-t equilibrium variable x_t with respect to the (reformed) net-of-capital tax rate $1-\tau_k^r$ are given by, respectively,*

$$\epsilon_{x_t,1-\tau_k} = \frac{d\ln x_t}{d\ln(1-\tau_k^r)} \quad \textit{and} \quad \varepsilon_{x_t,1-\tau_k} = \frac{d\ln x_t}{d(1-\tau_k^r)}.$$

The discounted average elasticities and semi-elasticities of x with respect to the (reformed) net-of-capital tax rate $1-\tau_k^r$ are given by, respectively,

$$\bar{\epsilon}_{x,1-\tau_k} = (1-\beta)\sum_{t=0}^{\infty}\beta^t\epsilon_{x_t,1-\tau_k} \quad \textit{and} \quad \bar{\varepsilon}_{x,1-\tau_k} = (1-\beta)\sum_{t=0}^{\infty}\beta^t\varepsilon_{x_t,1-\tau_k}$$

All these elasticities are what Hendren (2016) refers to as "policy elasticities", which measure the causal effect of a concrete policy experiment. For example, $\epsilon_{K_t,1-\tau_k}$ ($\varepsilon_{K_t,1-\tau_k}$) measures the relative change in the *equilibrium* capital stock in period t following an increase in the net-of-tax rate $1-\tau_k$ by one percent (one percentage point).

I define both elasticities ϵ and semi-elasticities ε because sometimes the formulas can be expressed more economically with one and sometimes with the other definition. Note, however, that they can be easily translated since

$$\epsilon_{x,1-\tau_k} = (1-\tau_k)\varepsilon_{x,1-\tau_k}.$$

Since the interpretation of none of the economic effects is qualitatively affected by which of the two concepts one uses, I employ different versions of the same greek letter (ϵ and ε) and I may, in the following, loosely refer to either of them as "elasticity".[8]

3.2 Local Welfare Effects and Globally Optimal Taxes

I now turn to the discussion of tax policy. A main contribution of this paper is the transparent decomposition of the total welfare effect of capital tax changes. Each component has a clear and intuitive economic interpretation. Generally, the separate components can be grouped into positive and normative ones, depending on whether the respective welfare effect is or is not invariant to the choice of welfare weights.

In particular, the change in welfare due to a marginal increase in the capital tax rate can be written as

$$dW = \left[EQ - MEB\right] Y_k d\tau_k.$$

The overall change in welfare is a cardinal measure that is not directly interpretable. I therefore define its components in terms of money metric utilities, as fraction of the additional tax revenue raised "mechanically" each period. Specifically $Y^k d\tau_k$ is the additional tax revenue that would be raised if agents were to keep their investment and labor supply unchanged. The marginal excess burden MEB measures how much, per mechanically raised dollar, the government loses in revenue due to individuals' behavioral responses (and their induced equilibrium effects). The equity effect EQ measures the planner's valuation of the tax induced change in the distribution of utilities. While the marginal excess burden MEB is a purely positive measure, the equity effect EQ depends on the particular choice of social welfare weights. A necessary condition for the existing tax rate τ_k to be optimal is

$$\frac{dW}{d\tau_k} = 0 \iff EQ = MEB,$$

that is the share of the mechanically raised revenue that is lost through changes in

[8] The distinction becomes important when a constant-elasticity assumption in used to extrapolate effects away from the current tax system. I will come back to this issue in Section 5.3.

agents' behaviour needs to be exactly offset by the distributional gain.

3.2.1 Standard Welfare Effects and Optimal Taxation with Exogenous Prices

I discuss first the welfare effects that are present also with exogenous factor prices, a case that is nested in my framework when capital and labor are assumed to be perfect substitutes ($F_{kl} = 0$).

Mechanical Redistribution. In that case the equity effect is exclusively given by the welfare change of mechanical redistribution

$$EQ_M = \beta^{t_a}(1 - \bar{g}^k).$$

Everything else equal, the government redistributes from capital income earners, whom it values by $\bar{g}^k$, to the general population, whom it values by $\bar{g} = 1$. Since redistribution happens only after the pre-announcement period t_a past, the effect is discounted by β^{t_a}.

Excess Burden through Investment Decline. An increase in the capital tax rate discourages investment and thereby reduces capital tax revenue. The effect

$$MEB_K = \tau_k \bar{\varepsilon}_{K,1-\tau_k}$$

measures how much, per mechanical dollar raised, the government loses in capital income tax revenue through through this channel. If labor supply was inelastic, this would be the only effect on the excess burden and the optimality condition would satisfy

$$EQ_M = MEB_K \iff \tau_k = \frac{\beta^{t_a}(1 - \bar{g}^k)}{\bar{\varepsilon}_{K,1-\tau_k}}. \tag{2}$$

If optimal, the capital income tax rate balances a classical equity-efficiency trade-off that follows the standard inverse elasticity rule. Given the redistributive preferences of the planner (captured parsimoniously by $\bar{g}^k$), the optimal tax rate is higher, the lower the (discounted average) net-of-tax-elasticity of capital. The optimal capital tax rate is declining in the announcement periods t_a because, while the redistributive gains are achieved only once the reform comes into effect, households will adjust their savings behaviour prior to that. Specifically, an increase in τ_k will split revenue more equally only from t_a on, but it reduces the raised revenue already from $t = 1$ onwards.

Excess Burden through Labor Supply Response. With elastic labor supply, one needs to correct this condition for the fact that changes in the capital income tax rate may affect labor supply and hence the excess burden of capital taxation. The welfare effect

$$MEB_L = \frac{\alpha^l}{\alpha^k}\bar{\varepsilon}_{L,1-\tau_k}\left[\mathrm{E}_\Gamma[\tau_l'] + \mathrm{Cov}_\Gamma\left(\tau_l', \frac{y^l}{Y^l}\frac{\bar{\varepsilon}_{l,1-\tau_k}}{\bar{\varepsilon}_{L,1-\tau_k}}\right)\right]$$

measures how much, per dollar in capital tax revenue raised mechanically, the government loses in labor income tax revenue. When factor prices are assumed to be constant, and wages hence invariant to capital tax changes, such changes affect labor supply only through potential income effects. Specifically, a reduction in the capital tax rate induces (i) a positive income effect due to the increase in net capital income $(1-\tau_k)y^k$ and (ii) a negative income effect through a reduction in the transfer T. The sign of MEB_L is therefore ambiguous.

With a linear labor income tax code, that is with $\tau_l'(y_l) = \tau_l$ for all y_l, we have $MEB_L = \frac{\alpha^l}{\alpha^k}\bar{\varepsilon}_{L,1-\tau_k}\tau_l$. The revenue effect through changing labor supply depends proportionally on the responsiveness of effective aggregate labor supply, captured by the semi-elasticity $\bar{\varepsilon}_{L,1-\tau_k}$, as well as on the extent to which a given unit of labor supply change translates into a change in revenue. The latter is captured by the product of the labor tax rate τ_l and the ratio of taxable labor over taxable capital income α^l/α^k. In this case the optimality condition boils down to the condition in Saez and Stantcheva (2018),[9]

$$EQ_M = MEB_K + MEB_L \iff \tau_k = \frac{\beta^{t_a}(1-\bar{g}^k) - \bar{\varepsilon}_{L,1-\tau_k}\frac{\alpha^l}{\alpha^k}\tau_l}{\bar{\varepsilon}_{K,1-\tau_k}}. \tag{3}$$

With a nonlinear tax schedule τ_l is replaced with the average marginal labor tax rate $\mathrm{E}_\Gamma[\tau_l']$. Furthermore, since agents with different marginal tax rates may adjust their labor supply differently, one needs to correct for the covariance (with respect to distribution Γ) of marginal tax rates and the labor income weighted elasticities of labor supply.

[9] Compare Propositions 8 and 9 in Saez and Stantcheva (2018) and note that condition (3) is equivalent to

$$\tau_k = \frac{1-\bar{g}^k - \tau_l\frac{Y^l}{Y^k}\beta^{-t^a}\bar{\varepsilon}_{L,1-\tau_k}}{1-\bar{g}^k + \beta^{-t^a}\bar{\varepsilon}_{K,1-\tau_k}}.$$

3.2.2 Additional Welfare Effects and Optimal Taxation with Endogenous Prices

Consider now the more realistic case, in which capital and labor are complements in production ($F_{kl} > 0$) or, equivalently, the substitution elasticity

$$\sigma_t \equiv \frac{d\ln\left(\frac{K_t}{L_t}\right)}{d\ln\left(\frac{F_l(K_t,L_t)}{F_k(K_t,L_t)}\right)} = \frac{F_k(K_t,L_t)F_l(K_t,L_t)}{F(K_t,L_t)F_{kl}(K_t,L_t)}$$

is finite.

In that case the marginal products of the two production factors are no longer invariant to tax changes. In particular, an increase in the capital tax rate, which reduces investment, will increase the marginal product of capital but reduce the marginal product of labor. In turn, this increases the demand for capital but reduces the demand for labor, causing a rise in the equilibrium interest rate but a decline in the equilibrium wage. Depending on whether labor responds positively or negatively to capital tax changes, this change in factor prices may be amplified or mitigated. In any case, the factor price changes are characterized by the following Lemma.[10]

Lemma 1. ***Price Elasticities and the Capital-Labor Substitution Elasticity.*** *Let Assumption 2 be satisfied. Then for all $t \geq 0$ we have*

$$\varepsilon_{r_t,1-\tau_k} = -\frac{\varepsilon_{K_t,1-\tau_k} - \varepsilon_{L_t,1-\tau_k}}{\sigma_t}\tilde{\alpha}_t^k\frac{\alpha_t^l}{\alpha_t^k} \quad \text{and} \quad \varepsilon_{w_t,1-\tau_k} = \frac{\varepsilon_{K_t,1-\tau_k} - \varepsilon_{L_t,1-\tau_k}}{\sigma_t}\tilde{\alpha}_t^k.$$

As a consequence the relative factor price changes are related through

$$\alpha_t^k\varepsilon_{r_t,1-\tau_k} = -\alpha_t^l\varepsilon_{w_t,1-\tau_k}.$$

Proof. See Appendix A.1.1. □

The responsiveness of factor prices is directly proportional to the relative change in the capital-labor ratio $\varepsilon_{K_t,1-\tau_k} - \varepsilon_{L_t,1-\tau_k}$ but indirectly proportional to the substitution elasticity σ_t. Higher complementarity between capital and labor (i.e. a lower σ_t) implies a more inelastic demand for production factors resulting in stronger price movements for any given change in factor supply. Furthermore, the elasticities of equilibrium factor prices are proportional to firms' expenditure share of the other factor. That

[10] Contrary to the case with exogenous prices, labor supply may change even if one rules out income effects. Specifically, the reduction in the equilibrium wage following a capital tax increase induces a substitution effect that reduces labor supply. With general preferences the wage reduction further induces a positive income effect, while the increase in the interest rate induces a negative income effect on labor supply.

is, the interest rate elasticity is proportional to the firms' expenditure share on labor $\tilde{\alpha}_l = 1 - \tilde{\alpha}_k$, while the wage elasticity is proportional to the expenditure share on capital $\tilde{\alpha}_k = 1 - \tilde{\alpha}_l$. Intuitively, if overall the firm spends very little on one factor, it is accepting larger changes in the unit cost of that factor to maintain a certain level of production.

The second part of the Lemma makes explicit that wage increases are accompanied by proportional reductions in the interest rate and vice versa, allowing me to express the welfare effect of factor price changes in terms of only one of these two elasticities.

The Welfare Effect of Changing Factor Prices. Factor price responses indirectly redistribute across agents with different income compositions and impact the marginal excess burden. I call the overall effect of price changes on welfare the *price effect*

$$P = -\frac{1}{\alpha^k}\left[\left[(1-\tau_k)\bar{g}^k + \tau_k\right]\alpha^k \bar{\varepsilon}_{r,1-\tau_k} + \left[(1-\bar{\tau}_l')\tilde{g}^l + \bar{\tau}_l'\right]\alpha^l \bar{\varepsilon}_{w,1-\tau_k}\right].$$

The rise in the interest rate caused by a tax hike of $d\tau_k > 0$ increases total capital income. A fraction $(1-\tau_k)$ of this income increase remains with its earners, while a fraction τ_k is taxed and hence increases the post-government income of all agents to an equal extent. Whenever, the planner values redistribution and there is inequality in capital income, she discounts the former by $\bar{g}^k < 1$. Analogously, the accompanying decline in wages reduces total labor income, a fraction $(1-\bar{\tau}_l')$ of which is borne by its earners, while a fraction $\bar{\tau}_l'$ is borne by the whole population through the reduction in revenue and hence the transfer. In the presence of labor income inequality, the planner discounts the former by $\tilde{g}^l < 1$.[11]

Decomposing the Price Effect. One can hence decompose the price effect into a component that affects the distribution of net income (EQ_P) and a component that affects the marginal excess burden (MEB_P),

$$\begin{aligned} P =& EQ_P - MEB_P \\ =& \underbrace{\frac{\alpha^l}{\alpha^k}\left[(1-\tau_k)\bar{g}^k - (1-\bar{\tau}_l')\tilde{g}^l\right]\bar{\varepsilon}_{w,1-\tau_k}}_{EQ_P} - \underbrace{\frac{\alpha^l}{\alpha^k}\left[\bar{\tau}_l' - \tau_k\right]\bar{\varepsilon}_{w,1-\tau_k}}_{MEB_P}, \end{aligned}$$

where I substituted out the interest elasticity $\bar{\varepsilon}_{r,1-\tau_k}$ using the proportionality result of Lemma 1.

[11] If labor income were to be taxed linearly, we would have $\tilde{g}^l = \bar{g}^l$, that is $\tilde{g}^l$ would be equal to the labor income weighted average marginal social welfare weight. With a progressive labor income tax code and a concave welfare objective, we have $\tilde{g}^l > \bar{g}^l$ as agents with higher marginal social welfare weight tend to have higher marginal retention rates.

The distributional effect of price changes on welfare EQ_P depends on the planners' valuation of proportionally distributing a marginal dollar to the earners of capital income vs. distributing it to the earners of labor income ($\bar{g}^k$ vs. $\tilde{g}^l$). By contrast, the effect of price changes on the excess burden MEB_P is independent of social welfare weights. It is proportional to the difference of the income weighted marginal tax rates on capital and labor income ($\bar{\tau}_l' - \tau_k$). Importantly, if $\tau_k > \bar{\tau}_l'$ the interest rate increase and wage reduction induced by an increase in the capital tax rate have a positive impact on revenue and therefore reduce the marginal excess burden ($MEB_P < 0$).

The following proposition summarizes the welfare decomposition of marginal tax increases and the condition for optimality for the most general case.

Proposition 1. ***Local Welfare Effects and Optimal Capital Tax Rate.*** *Let Assumptions 1 to 3 be satisfied. The effect of a marginal tax increase $d\tau_k > 0$ on social welfare is given by*

$$dW = \Big[\underbrace{EQ_M + EQ_P}_{=EQ} - \underbrace{\left(MEB_K + MEB_L + MEB_P\right)}_{MEB}\Big] Y^k d\tau_k. \tag{4}$$

Consequently, the pre-existing capital income tax rate $\tau_k < 1$ is optimal only if it satisfies

$$\tau_k = \frac{\beta^{t^a}(1 - \bar{g}^k) - MEB_L + P}{\bar{\varepsilon}_{K,1-\tau_k}}. \tag{5}$$

Proof. See Appendix A.1.2.[12] □

3.3 Illustrating the Optimality Condition

To obtain a better understanding of condition (5), I will next discuss the optimal tax rates for several important special cases. For simplicity, I focus on the case with a linear labor income tax code, that is where $\tau_l(y^l) = \tau_l y^l$ for all y^l.

Homogenous Labor Income. As mentioned above, assuming $\underline{\eta} = \bar{\eta}$, a linear labor tax code and inelastic labor supply ($\bar{\varepsilon}_{L,1-\tau_k} = 0$), my framework nests the model of Judd (1985), in which agents have heterogeneous initial wealth and therefore heterogeneous capital income, while labor income is homogeneous.[13] In this case $\tilde{g}^l = \bar{g}^l = 1$ and the

[12] Note that since the framework presented in the main text is a special case of the more general model in the Appendix, this proposition follows in principle directly from Proposition B.1 in Appendix B.2.2, which derives the same condition for the general framework. However, for the reader's convenience, I present a separate (easier) proof for the model of the main text.

[13] Note that my model only nests the framework of Section 4 in Judd (1985). However, the economic environment in Section 3 of Judd (1985), where a part of the population is excluded from the capital market, would yield the same tax formula.

condition becomes

$$\tau_k = \frac{\beta^{t_a}(1-\bar{g}^k)}{\bar{\varepsilon}_{K,1-\tau_k}} \underbrace{-\frac{\alpha^l}{\alpha^k}\frac{\tilde{\alpha}^k}{\sigma}(1-\bar{g}^k)(1-\tau_k)}_{<0}. \tag{6}$$

The first term is the same as in condition (2) with exogenous prices. The additional term captures standard 'trickle-down' theory. A decrease in investment reduces wages and increases capital returns. Since labor income is equally distributed while capital income is unequally distributed across agents, the wage decrease associated with an increase in capital taxes reduces social welfare and thus the optimal capital income tax rate. The paper of Judd (1985) is mostly known for its famous result that capital income taxes should be zero in the long run. However, the result refers to the infinite future, derived in settings where the Ramsey planner is able to commit to a path of time-varying taxes. In the short run, also Judd (1985) finds positive optimal capital tax rates although "redistributive capital taxation is severely limited in its effectiveness since it depresses wages" (Judd, 1985, p.59). When taxes are required to be time-invariant, this depressing effect on wages is transparently captured by the additional negative term in the optimality condition (6).

Note that the additional term is independent of the labor income tax rate τ_l because, absent wage heterogeneity, any dollar of labor taxes paid ends up back in the hand of the agent through an equal increase in her lump-sum transfer. Hence, the wage decrease, in and by itself, has an equally negative effect on all agents' disposable income. However, it is discounted by $(1-\bar{g}^k)$ because the associated increase in the interest rate benefits the earners of capital income, whom the government values by $\bar{g}^k$.

With elastic labor supply the condition becomes

$$\tau_k = \frac{\beta^{t_a}(1-\bar{g}^k) - \bar{\varepsilon}_{L,1-\tau_k}\frac{\alpha^l}{\alpha^k}\tau_l}{\bar{\varepsilon}_{K,1-\tau_k}} - \frac{\alpha^l}{\alpha^k}\frac{\tilde{\alpha}^k}{\sigma}\left(1-\frac{\bar{\varepsilon}_{L,1-\tau_k}}{\bar{\varepsilon}_{K,1-\tau_k}}\right)(1-\bar{g}^k)(1-\tau_k). \tag{7}$$

The first term is analogous to the one in condition (3) and has the same interpretation. If labor supply declines in response to a capital tax increase, that is if $\bar{\varepsilon}_{L,1-\tau_k} > 0$, the excess burden of capital taxes is increased $MEB_L > 0$, reducing the optimal capital tax rate. However, at the same time the decline in labor supply mitigates the responsiveness in equilibrium factor prices and therefore the depressive effect of capital taxes on wages, a force that is increasing the optimal capital tax rate. The opposite is the case when $\bar{\varepsilon}_{L,1-\tau_k} < 0$.

The optimality of a zero long-run capital tax in Judd (1985) – as the one in Chamley (1986) – was recently revisited by Straub and Werning (2020), who show that, depending on parameters, optimal tax rates on capital may remain positive forever. All three papers assume that the Ramsey planner is able to set the whole path of time-varying capital taxes. By contrast, as Piketty and Saez (2013) or Saez and Stantcheva (2018), I only allow for a one-off change in the capital tax rate, after which it is required to remain constant. While time-invariant taxes naturally result in a lower welfare optimum than what could be achieved with time-varying taxes, this restriction induces a robust solution.[14]

Heterogeneous Labor- and Capital Income. In reality, both capital- and labor income are heterogeneous, in which case the optimal capital tax rate is implicitly given by

$$\tau_k = \frac{\beta^{t_a}\left(1-\bar{g}^k\right) - \bar{\varepsilon}_{L,1-\tau_k}\frac{\alpha^l}{\alpha^k}\tau_l}{\bar{\varepsilon}_{K,1-\tau_k}} + \frac{\alpha^l}{\alpha^k}\frac{\tilde{\alpha}^k}{\sigma}\left(1-\frac{\bar{\varepsilon}_{L,1-\tau_k}}{\bar{\varepsilon}_{K,1-\tau_k}}\right)\Big[\underbrace{(1-\tau_k)\bar{g}^k}_{>0}\underbrace{-(1-\tau_l)\bar{g}^l}_{<0}\underbrace{+\tau_k-\tau_l}_{?}\Big].$$

The three terms in the squared bracket capture the welfare impact of tax induced factor price changes. The first term $(1-\tau_k)\bar{g}^k > 0$ captures the gains through agents' net capital income increase, the second term $-(1-\tau_l)\bar{g}^l < 0$ captures the loss through agents' net labor income decline, and the third term $\tau_k - \tau_l$ captures the effect of increasing interest rates and the associated reduction of wages on government revenue. This last effect is positive if and only if $\tau_k > \tau_l$.

In the data capital income is much more concentrated than labor income. For example, the poorer half of the US population earns basically zero capital income but has substantial labor income. One may hence conclude that for the same reason as above the wage reductions and interest increases accompanying capital tax increases should reduce social welfare and thus call for a lower taxation of capital. The formula shows why such an interpretation is wrong, or at least incomplete. Consider a planner who cares only about the very lowest earners with neither labor- nor capital income ($\bar{g}^k = \bar{g}^l = 0$). Since price changes do not have any impact on these agents' net income the distributional loss is zero, $EQ_P = 0$. However, at the same time these price changes affect the marginal excess burden whenever capital and labor are taxed differently. If $\tau_k > \tau_l$ they reduce the excess burden, $MEB_P < 0$, and thus call for a higher taxation

[14] As demonstrated by Straub and Werning (2020), with time-varying taxes, along the optimal trajectory, the economy often does not converge to an interior steady state, in which case the optimal capital tax remains positive forever. Benhabib and Szőke (2021) derive conditions for positive long-run capital taxes that are compatible with an interior steady state but where the capital tax rate remains at its allowed upper limit, that is at the corner solution.

of capital.[15]

Idiosyncratic Risk. As mentioned above – for a large set of social welfare functions – the optimality condition (5) holds for a much more general framework (presented in Appendix B), which nests models with uninsurable idiosyncratic risks to labor- and capital income. In particular, it also holds for the standard incomplete markets model of Aiyagari (1994). The generality of this condition may seem surprising in light of the result in Aiyagari (1995), who shows that the optimal long-run capital income tax rate in the standard incomplete markets model is positive, contrasting Judd (1985) and Chamley (1986), who find zero optimal long-run tax rates in frameworks with complete markets. The reason is again the time-invariance restriction on capital tax rates imposed in the present paper. With such a restriction in place, equation (5) provides a testable condition for optimality. That means, irrespective of whether the true data generating process is better explained by one nested framework or another, as long as the elasticities, factor shares and distributional statistics on the right hand side of the condition are estimated accurately, the condition tells you whether the capital tax rate is optimal, and if not, in which direction it should be adjusted.[16]

4 Recovering Unmeasured Policy Elasticities

In the interest of space, I from now on restrict the analysis to unannounced reforms ($t^a = 0$) and to preferences that do not exhibit income effects on labor supply. While these assumptions are not necessary to perform the analysis, they considerably simplify the exposition.

Assumption 4. *Preferences exhibit no income effects on labor supply, that is the Bernoulli utility function is of the form*

$$u(c,l) = U(c - v(l)),$$

where $U'(.) > 0$, $U''(.) < 0$, $v(.)' \geq 0$ *and* $v(.)'' > 0$.

The equilibrium factor elasticities $\bar{\varepsilon}_{K,1-\tau_k}$ and $\bar{\varepsilon}_{L,1-\tau_k}$ are endogenous objects that de-

[15] There is a close analogy between this analysis and the one in Sachs et al. (2020), who study tax incidence and optimal income taxation in a static Mirrlees environment with complementary labor types. In their environment, an increase in the progressivity of the tax system increases the wages of top earners, whose labor input becomes more scarce, and decreases the wages of lower earners. Hence, the endogenous wage responses of further increasing the progressivity of an already progressive tax system positively impacts government revenue.

[16] In the following Section, I link the unmeasured net-of-tax elasticities of the equilibrium capital stock to actually estimated capital supply elasticities. I create this link via the deterministic model of the main text. I discuss in Appendix C.3 why idiosyncratic risk, and the associated precautionary savings motive, does not matter much quantitatively in this context.

pend in particular on the substitution elasticity σ. As discussed above, absent income effects, labor supply is unaffected by capital tax changes ($\bar{\varepsilon}_{L,1-\tau_k} = 0$) when labor and capital are assumed to be perfect substitutes ($\sigma = \infty$). However, whenever the substitution elasticity is finite ($\sigma < 0$) an increase in the capital tax rate reduces wages, which induces a negative substitution effect on labor supply ($\bar{\varepsilon}_{L,1-\tau_k} > 0$).

More importantly, the more complementary capital and labor are, the stronger the increase in the gross interest rate following a capital tax hike. This increase in the interest rate mitigates the mechanical drop in the net return on capital and thus has a moderating effect on the net-of-tax (semi-)elasticity of the equilibrium capital stock. In sum, the sequencies of equilibrium capital elasticities $\{\varepsilon_{K_t,1-\tau_k}\}_{t=1}^{\infty}$ and equilibrium labor elasticities $\{\varepsilon_{L_t,1-\tau_k}\}_{t=1}^{\infty}$ are "policy elasticities" in the sense of Hendren (2016). They measure the causal effect of the concrete policy experiment performed in this paper and they therefore capture all simultaneous equilibrium responses. Such elasticities are hard, if not impossible, to estimate directly.

However, by exploiting agents' optimality conditions together with market clearing, one can recover these unmeasured policy elasticities from pure factor supply elasticities that are actually estimated. The interested reader can find the detailed description of this methodology in Appendix C, while I only briefly summarize it here.

4.1 Labor Supply

Abstracting from income effects, labor supply in any given period is only affected through contemporaneous changes in the wage. Consider an individual with characteristics (k_0, η). The policy elasticity of her labor supply is given by

$$\epsilon_{l_t(k_0,\eta),1-\tau_k} = \tilde{\epsilon}_{l_t(k_0,\eta),w_t}\epsilon_{w_t,1-\tau_k},$$

where $\tilde{\epsilon}_{l_t(k_0,\eta),w_t}$ denotes the pure supply elasticity, which measures the relative change in labor supplied in period t if only the wage in that same period changes, keeping all other prices, taxes and transfers fixed. Denoting by γ_l the Frisch elasticity and by

$$p(y^l) = -\frac{\partial \ln(1-\tau_l'(y^l))}{\partial \ln(y^l)} = \frac{y^l \tau_l''(y^l)}{1-\tau_l'(y^l)}$$

the local rate of labor tax progressivity, the wage elasticity of labor supply is given by

$$\tilde{\epsilon}_{l_t(k_0,\eta),w_t} = \frac{\partial \ln l_t(k_0,\eta)}{\partial \ln w_t} = \frac{\gamma_l\big(1-p(y^l(k_0,\eta))\big)}{1+\gamma_l p(y^l(k_0,\eta))}.$$

With a linear labor tax schedule ($p(y^l) = 0$ for all y^l) a one percent increase in wages ceteris paribus increases labor supply by $\tilde{\epsilon}_{l_t(k_0,\eta),w_t} = \gamma_l$ percent. That is, the labor supply response of each agent equals the Frisch elasticity. A progressive tax schedule ($p(y^l) > 0$) dampens the positive substitution effect of the wage increase on labor supply and therefore the labor supply elasticity. Aggregating up $\tilde{\epsilon}_{l_t(k_0,\eta),w_t}$ gives the pure supply elasticity of aggregate labor $\tilde{\epsilon}_{L_t,w_t}$, which is related to the policy elasticity in an analogous way as the individual supply elasticities. Consequently, one can substitute

$$\epsilon_{w_t,1-\tau_k} = \frac{\tilde{\epsilon}_{L_t,w_t}}{\epsilon_{L_t,1-\tau_k}}$$

into the equation in Lemma 1 and solve for $\epsilon_{L_t,1-\tau_k}$ to obtain

$$\epsilon_{L_t,1-\tau_k} = \frac{\tilde{\alpha}^k \frac{\tilde{\epsilon}_{L_t,w_t}}{\sigma}}{1 + \tilde{\alpha}^k \frac{\tilde{\epsilon}_{L_t,w_t}}{\sigma}} \epsilon_{K_t,1-\tau_k}.$$

The net-of-capital-tax-rate elasticity of equilibrium aggregate labor supply is therefore directly proportional to the net-of-capital-tax-rate elasticity of the equilibrium capital stock. A reduction in the capital tax rate increases investment and hence the capital stock. This increases the marginal product of capital and hence wages, thereby also encouraging labor supply. Observe that $\varepsilon_{L_t,1-\tau_k} < \varepsilon_{K_t,1-\tau_k}$. With an infinite substitution elasticity, as in Piketty and Saez (2013), Golosov et al. (2014) or Saez and Stantcheva (2018), we have $\varepsilon_{L_t,1-\tau_k} = 0$. In that case wages are invariant to capital tax changes, which in the absence of income effects implies that also labor supply is unaffected.

In any case, the path of equilibrium effective labor elasticities $\{\epsilon_{L_t,1-\tau_k}\}_{t=1}^{\infty}$ can be recovered from the path of equilibrium capital elasticities $\{\epsilon_{K_t,1-\tau_k}\}_{t=1}^{\infty}$. The main difficulty is to obtain the latter, an issue to which I turn next.

4.2 Capital Supply

One can recover the path of elasticities of the equilibrium capital stock from the path of pure capital supply elasticities with respect to the net-of-capital tax rate,

$$\tilde{\epsilon}_{K_t,1-\tau_k} = \frac{\partial \ln K_t}{\partial \ln(1-\tau_k)}.$$

$\tilde{\epsilon}_{K_t,1-\tau_k}$ is again a *ceteris-paribus* elasticity (hence the partial derivative) that measures the relative change in capital supply (wealth) with respect to a change in the net-of-capital tax rate *keeping all other variables fixed*. In particular, it measures households' response in wealth accumulation if only the tax rate would change but both prices and transfers would remain constant. This elasticity is therefore the treatment effect of an experiment that changes the capital tax rate for a small part of the population, whose behaviour has a negligible influence on the government budget and on equilibrium prices.

Though sparse, we do have evidence on the elasticity of capital supply. Arguably the best currently available estimates are those by Jakobsen, Jakobsen, Kleven, and Zucman (2020), who use administrative Danish data. Their study is partially motivated by the paper of Saez and Stantcheva (2018) and the therein stated lack of evidence for this crucial elasticity. The authors exploit natural experiments emanating from a 1989 wealth tax reform, with which they estimate the elasticity of wealth with respect to wealth taxes for eight years following the reform.[17]

As I prove in Lemma C.3 optimal savings behaviour dictates a tax-elasticity of wealth that is linear in time. Specifically, denoting the intertemporal elasticity of substitution by γ_c optimal savings behavior implies

$$\tilde{\epsilon}_{K_t,1-\tau_k} = t\beta\frac{C}{K}\gamma_c = t\tilde{\epsilon}_{K_1,1-\tau_k}.$$

Consequently, a single observation, taken at any time after the reform, carries enough information to recover the whole path of pure supply elasticities. The evidence in Jakobsen et al. (2020) provides eight such data points, implying that $\tilde{\epsilon}_{K_t,1-\tau_k}$ is, in principle, over-identified. To make use of all available evidence, I hence regress the estimated tax elasticities of wealth on time. The black solid line in Figure 1 depicts the estimates in Jakobsen et al. (2020), the red dotted line the regression. The theoretically predicted linearity in time squares remarkably well with the data.[18]

However, as described above, in our experiment the government transfers $\{T_s\}_{s=0}^{\infty}$ adjust period-by-period in a budget neutral way. Furthermore, whenever capital and labor are imperfect substitutes, a tax change induces factor prices $\{r_s, w_s\}_{s=0}^{\infty}$ to change too. A change in any transfer T_s or wage w_s induces an income effect on capital supply. A change in the (gross) return r_s induces both a substitution and an income effect similar to a change in the capital tax rate. The estimated tax-elasticities of capital supply pin down these income- and substitution effects, such that they are sufficient to recover the capital supply responses to changes in transfers and factor prices, that is

[17] Jakobsen et al. (2020) exploit two quasi-experiments that affected two different subsets of the population. In the main text I use their estimates on households between the 97.6th and the 99.3rd percentile of the wealth distribution. Their elasticity estimates of the top percentile are slightly higher. I report those in Appendix C.

[18] In the second part of their paper Jakobsen et al. (2020) employ a structural life-cycle model to extrapolate the "long-run" wealth elasticity with respect to wealth taxes. They define the "long-run" elasticity as the end-of-life elasticity. However, although the authors argue that there is a high tax-elasticity of bequests, they abstract from the fact that when inheriting more, heirs also accumulate more wealth. As do the heirs of heirs, and so on. That is, the finiteness of their long-run supply elasticity is artificially introduced by the finite time horizon.

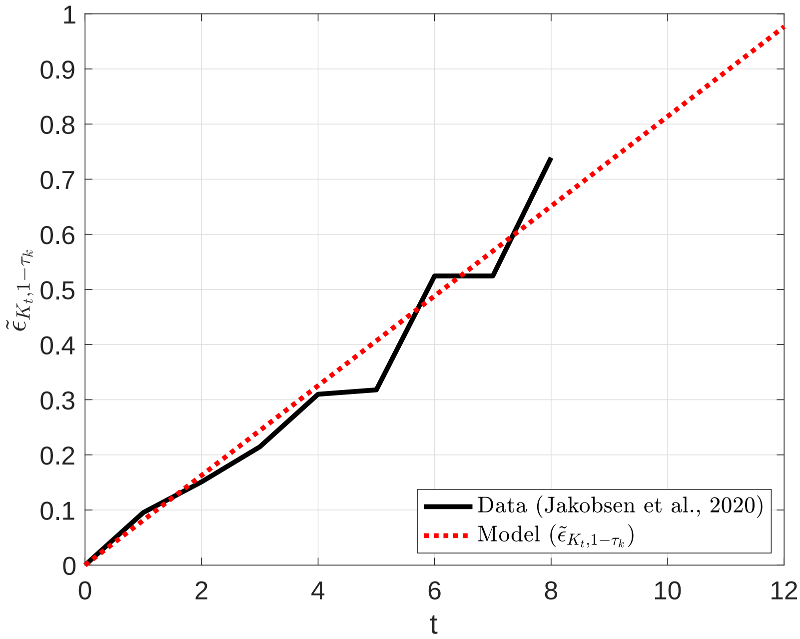

Figure 1: Capital Supply Elasticity: data (solid line) from Jakobsen et al. (2020) Figure V (left panel); treatment on the treated; net-of-wealth-tax elasticities are translated to net-of-capital-tax elasticities using the return of $r = 6.58\%$; model (dotted line), $\tilde{\epsilon}_{K_t,1-\tau_k} = t\tilde{\epsilon}_{K_1,1-\tau_k}$.

$\{\tilde{\epsilon}_{K_t,T_s}, \tilde{\epsilon}_{K_t,w_s}, \tilde{\epsilon}_{K_t,r_s}\}_{s=0}^{\infty}$. The tax-elasticity of the *equilibrium* capital stock can then be obtained by multiplying the various supply responses with changes in transfers and factor prices $\{\epsilon_{T_s,1-\tau_k}, \epsilon_{r_s,1-\tau_k}, \epsilon_{w_s,1-\tau_k}\}_{s=0}^{\infty}$ that are consistent with government budget- and factor market clearing in the respective period. Specifically, for each period $t \geq 0$ the equilibrium capital stock can be decomposed as

$$\epsilon_{K_t,1-\tau_k} = \tilde{\epsilon}_{K_t,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s}\epsilon_{T_s,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,r_s}\epsilon_{r_s,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,w_s}\epsilon_{w_s,1-\tau_k}.$$

Solving this system of linear equations then gives the path of the desired unmeasured policy elasticities $\{\epsilon_{K_t,1-\tau_k}\}_{t=0}^{\infty}$.

Figure 2 plots the path of equilibrium capital elasticities for my benchmark finite substitution elasticity (blue dash-dotted line) as well as for the case where capital and labor are assumed perfect substitutes, that is when prices are assumed to be invariant to tax changes (red dashed line) along with the pure supply elasticities of Figure 1. When the substitution elasticity is infinite ($\sigma = \infty$) the equilibrium capital elasticity grows linearly in time, though the income effect from the budget neutral transfer mitigates the savings response relative to the pure supply elasticity. However, whenever the substitution elasticity is finite ($\sigma < 0$) the policy elasticity converges to a finite level.[19]

[19] In Appendix D, I perform a sensitivity analysis with respect to the range of empirical estimates of σ.

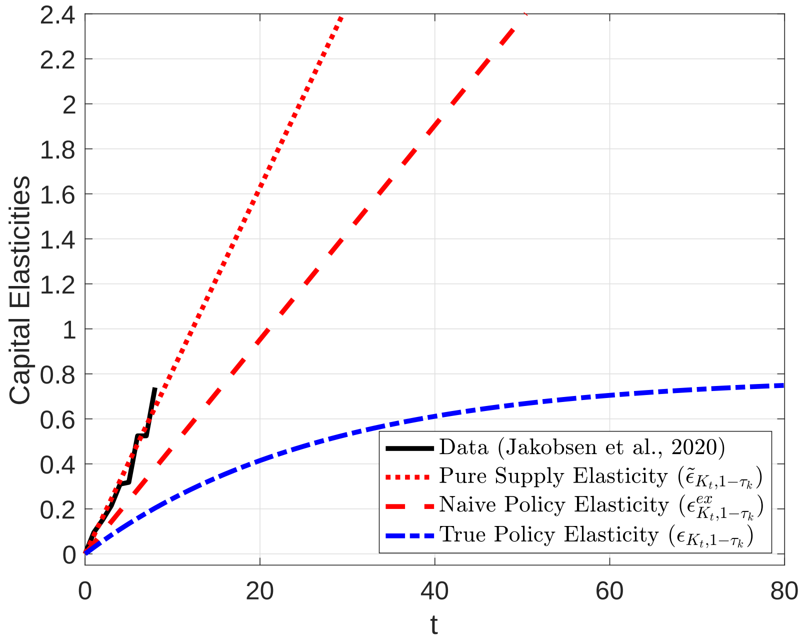

Figure 2: Capital Elasticities: black solid line and red dotted line as in Figure 1; red dashed line ($\epsilon^{ex}_{K_t,1-\tau_k}$): policy elasticity in the exogenous price case ($\sigma = \infty$); blue dash-dotted line line ($\epsilon_{K_t,1-\tau_k}$): policy elasticity with endogenous prices ($\sigma = 0.6$); Frisch elasticity of labor supply $\gamma_l = 0.5$.

The reason for this difference is that whenever the substitution elasticity is finite, a tax cut that induces an increase in capital accumulation reduces the marginal product of capital and thus the rental rate of capital, which firms are willing to pay to investors. This decline in r mitigates the overall increase in the net return $\bar{r} = (1 - \tau_k)r$ and thus moderates the investment increase. Such a mechanism is absent when capital and labor are assumed to be perfect substitutes. In that case, capital increases to infinity for similar reasons as in the *Ak* model of economic growth. The implicit, counter-factual, assumption behind this is that the marginal product of capital is constant.[20] As we will see below, these differences in the net-of-tax elasticities of equilibrium capital will imply strong differences in the marginal excess burden of capital taxation.

5 Quantitative Application

I now move to the quantitative application of my theoretical results. For this means, I first need to collect all the relevant statistics entering my formulas.

[20] Saez and Stantcheva (2018) wrongly attribute the Chamley-Judd zero long run tax result to the infinite long-run capital elasticity ($\lim_{t\to\infty} \epsilon_{K_t,1-\tau_k=\infty}$) in their framework (see their Table 1 and the discussion surrounding it). However, this elasticity only goes to infinity when one assumes perfect substitutability in capital and labor, an assumption neither Judd (1985) nor Chamley (1986) make.

5.1 Values of Sufficient Statistics

This section summarizes the benchmark estimates for the sufficient statistics, which I use in the quantitative analysis of the main text. In Appendix D, I provide a sensitivity analysis with respect to those statistics, for which the empirical range of estimates is rather broad.

Factor Shares. The gross capital share of $\tilde{\alpha}^k = 0.4$ is relatively uncontroversial. However, only few studies estimate net income shares (after capital depreciation). One exception is Rognlie (2015), who finds that the net capital share is 74% of the gross share. Given the gross capital share of $\tilde{\alpha} = 0.4$ this implies $\alpha^k = 0.296$.

Capital- and Labor Income. The weighted marginal social welfare weights $\bar{g}^k$ and $\tilde{g}^l$ depend partially on the chosen welfare objective, but partially also on the distribution of capital and labor income. I use income and wealth data from the Survey of Consumer Finances 2019 (SCF), restricting the sample to prime-age workers (non-retirees of age 64 and younger). Capital income is the return to net worth. While single SCF waves do not include sufficient information on unrealized capital gains, Xavier (2021) estimates returns in the US by combining the SCF waves from 1989 to 2019 with data on private business equity from the US Financial Accounts as well as public equity- and real estate indices. She finds a wealth weighted average annual return of 6.80%. Similarly, in their comprehensive cross-country analysis Jorda, Knoll, Kuvshinov, Schularick, and Taylor (2019) find an annual post 1980 average return on wealth of 6.58% for the US. I use their estimate in my analysis, which is very close to their estimate for Denmark (6.62%), for which we have quasi-experimental evidence on the capital supply elasticity. However, I provide robustness checks with lower and higher capital returns (of $r = 5\%$ and $r = 8\%$, respectively) in Appendix D.1.

Labor income comprises wage income and some of the income generated in privately owned businesses and farms. It is known to be empirically difficult to disentangle the capital- and labor component of the latter. To discipline this choice somewhat, I calibrate that 62% of business and farm income is to be assigned as labor income, such that for the (representative) population in the SCF the capital income share is $\tilde{\alpha}^k = 0.296$ with the benchmark return of $r = 6.58\%$. This implies a reasonable depreciation rate of $\delta = 3.85\%$ per annum.

The left panel of Figure 3 depicts the data in a scatter plot with wealth on the x-axis and labor income on the y-axis. The right panel shows the corresponding Lorenz curves for labor income and wealth. As is well known, wealth is much more concentrated than labor income. For example, while the poorest 60% of US households own only a negligible amount of total US wealth, the bottom 60% of the labor income distribution

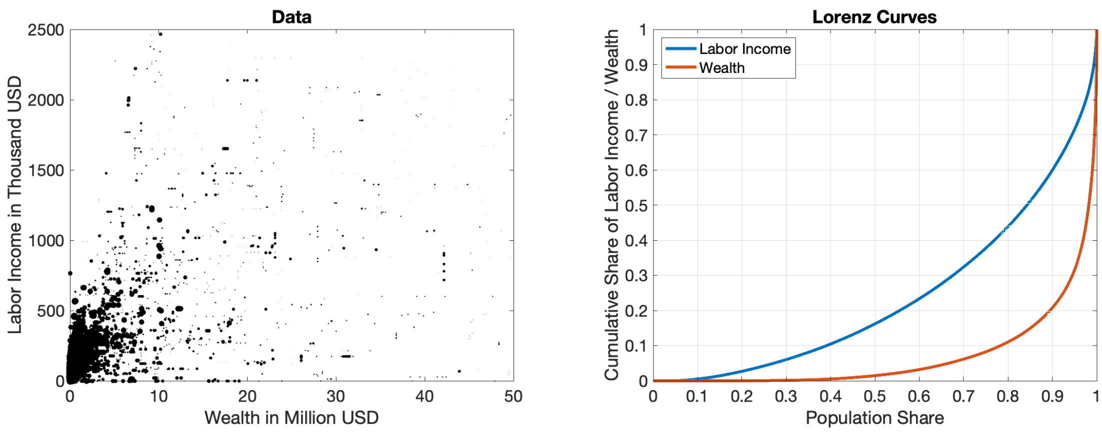

Figure 3: Labor Income and Wealth: data from Survey of Consumer Finances 2019 (SCF); non-retired households aged 64 or younger; 62% of business and farm income assigned to labor income; size of scatter plot markers proportional to SCF sample weights.

earn almost 25% of total US labor income. Similarly, while the richest 10% US households own about 80% of total US wealth, the 10% highest labor income earners, receive 'only' about 40% of total US labor income. The corresponding Gini indices are 0.53 for labor income and 0.86 for wealth. While the depicted Lorenz curves, or at least some points on them, are often calibration targets in parametric quantitative studies, they do not uniquely pin down the joint distribution of labor income and wealth.

An advantage of my approach is that I can apply my formulas directly to the data. In particular, the data points in the scatter plot are the inputs for the computation of $\bar{g}^k$ and $\tilde{g}^l$ that enter my welfare effects. Thus, in contrast to the parametric literature, I do not rely on any approximation of the true joint distribution of income and wealth. Instead, I can use the exact distribution.

Taxes. Income taxes are not reported in the SCF. However, Heathcote, Storesletten, and Violante (2017) document that, properly calibrated, the mapping

$$\tau_l(y^l) = y^l - (1-\tau_0)(y^l)^{1-p} \tag{8}$$

provides an exceptionally good fit of the data. I hence use this mapping and their progressivity parameter of $p = 0.181$. Their level parameter τ_0 is neither scale invariant, nor reported in their study. I calibrate it such that the labor income weighted average marginal tax rate is $\bar{\tau}_l' = 0.225$, as estimated by Trabandt and Uhlig (2012). This calibration gives an average marginal labor tax of $\mathrm{E}_\Gamma[\tau_l'] = 0.07$. From the latter study, I also take the capital income tax rate in the status quo, $\tau_k = 0.415$.

Labor Supply Elasticities. I take a benchmark Frisch elasticity of labor supply of $\gamma_l = 0.5$. However, to account for the variation in empirical estimates estimates, I provide a sensitivity analysis in Appendix D. The constant Frisch elasticity together with the constant rate of progressivity tax schedule imply a zero co-variance between the marginal labor tax rate and the net-of-capital-tax elasticity of equilibrium labor supply, that is $\text{Cov}_\Gamma(\tau_l', \bar{\epsilon}_{l,1-\tau_k}) = 0$.[21]

Substitution Elasticity. Empirical estimates of σ are vast. I use the most recent estimate of $\sigma = 0.6$ from Oberfield and Raval (2021) as my benchmark value. For comparability, in the main text I also report to the case of capital and labor being perfect substitutes ($\sigma = \infty$). Furthermore, in Appendix D.2, I report the results for the whole range of empirical estimates of σ.

Discounted Average Tax-Elasticities of Capital and Labor. As explained above, the discounted average elasticities of equilibrium capital and labor are endogenous to σ and γ_l. With the benchmark values of $\sigma = 0.6$ and $\gamma_l = 0.5$ we have $\bar{\epsilon}_{K,1-\tau_k} = 0.39$ and $\bar{\epsilon}_{L,1-\tau_k} = 0.08$, while under the assumption of exogenous prices, that is with $\sigma = \infty$, we have $\bar{\epsilon}_{K,1-\tau_k} = 1.24$ and $\bar{\epsilon}_{L,1-\tau_k} = 0$, irrespective of γ_l. I report these elasticities for different combinations of substitution- and Frisch elasticities in Table D.3 in Appendix D.2.

5.2 Welfare Decomposition of Local Tax Changes

In this section I perform a local welfare decomposition. That is, given the status quo tax system, I consider a marginal increase in the capital tax rate $d\tau_k > 0$ and I compute the various components affecting the overall welfare change

$$dW = \Big[\underbrace{EQ_M + EQ_P}_{=EQ} - \underbrace{\big(MEB_K + MEB_L + MEB_P\big)}_{MEB}\Big] Y^k d\tau_k.$$

5.2.1 The Marginal Excess Burden

I first consider the components that are invariant to the choice of the social welfare function, that is the components that affect the marginal excess burden (MEB) of capital taxation. Table 1 summarizes the three components of MEB. The first line covers the case where capital and labor are perfect substitutes and factor prices are

[21] Specifically, the individual net-of-capital-tax elasticities of equilibrium labor supply in period t are given by

$$\epsilon_{l_t(k_0,\eta),1-\tau_k} = \tilde{\epsilon}_{l_t(k_0,\eta),w_t} \epsilon_{w_t,1-\tau_k} = \frac{\gamma_l(1-p)}{1+\gamma_l p} \epsilon_{w_t,1-\tau_k}$$

and are thus independent of the agents' characteristics (k_0, η) and hence constant across the population.

therefore constant. The exclusion of income effects implies that the decomposition is trivial in this case. Specifically, absent changes in the equilibrium wage, a change in the capital tax rate will not affect labor supply and hence keep labor income tax revenue constant ($MEB_L = 0$). Furthermore, assuming away factor price changes implies that also $MEB_P = 0$. Consequently, the total marginal excess burden consists exclusively of the revenue loss due to a reduction in agents' savings. This revenue loss of $MEB = MEB_K = 0.88$, however, is substantial. For each mechanical dollar raised the government loses 88 cent due to the behavioral investment decline. Absent counteracting responses in the equilibrium interest rate the net-of-tax elasticity of the equilibrium capital stock is large (see Figure 2), implying a large tax distortion.

	MEB_K	MEB_L	MEB_P	MEB
Exogenous prices ($\sigma = \infty$)	0.8775	0.0000	0.000	0.8775
Endogenous prices ($\sigma = 0.6$)	0.2589	0.0196	-0.1497	0.1287

Table 1: Decomposition of the Marginal Excess Burden: Components of the marginal excess burden (MEB); numbers in dollar per mechanical dollar in capital tax revenue raised; MEB_K: loss in capital income tax revenue due to lower savings; MEB_L: loss in labor income tax revenue due to lower labor supply; MEB_P: revenue impact of changing factor prices due to differential taxation of capital and labor; Frisch elasticity: $\gamma_l = 0.5$

By contrast, when capital and labor are complements ($\sigma = 0.6$, second line) a rise in the capital tax rate increases the gross return to capital. This mitigates the equilibrium reduction in the net return $(1 - \tau_k)r$, which has a moderating effect on the investment decline. Consequently, the capital tax revenue loss coming from the investment reduction, $MEB_K = 0.26$, is much lower than in the case with exogenous prices. The tax induced reduction in the wage lowers labor supply and thus negatively affects labor income tax revenue. However, this effect of $MEB_L = 0.02$ is quite small. More importantly, the factor price changes have a direct revenue impact. Specifically, since in the status quo capital is taxed at a higher average rate than labor ($\tau_k > \bar{\tau}_l'$), the increase in the gross interest rate causes higher capital tax revenue gains than the reduction in wages causes labor tax revenue losses. The overall price effect of $MEB_P = -0.15$ significantly reduces the excess burden of capital taxation. The total excess burden in the realistic case is $MEB = 0.13$, only about one quarter of the excess burden one obtains when naively assuming constant prices.

5.2.2 The Equity Effect

An increase in the capital tax rate affects the distribution of disposable income. The welfare assessment of this redistribution requires a normative judgement, that is a stand on how to value the relative consumption of different individuals. To be agnostic and to explore the welfare effects across the distribution, I follow the strategy of Piketty and Saez (2013) by considering hundred different social welfare functions, each of which concentrates the whole weight in a specific percentile of the total gross income distribution. Formally, for each social welfare function indexed by $pct \in \{1, 2, ..., 100\}$, the corresponding marginal social welfare weights are given by

$$g_{pct}(k_0, \eta) = \begin{cases} 100 & \text{if } \Gamma_y(y(k_0, \eta)) \in [pct - 1, pct) \\ 0 & \text{else,} \end{cases}$$

where Γ_y is the distribution of gross total income induced by the distribution Γ over states (k_0, η) and the agents' optimal choices. In turn, this implies for each of these welfare objectives that the capital (labor) income weighted marginal social welfare weight for the social welfare function pct is simply the average capital (labor) income within percentile pct of the income distribution divided by the mean capital (labor) income of the whole population,

$$\bar{g}^k_{pct} = \int_{\Gamma_y(y(k_0,\eta)) \in [pct-1, pct)} \frac{y^k(k_0, \eta)}{Y^k} d\Gamma \quad \text{and} \quad \bar{g}^l_{pct} = \int_{\Gamma_y(y(k_0,\eta)) \in [pct-1, pct)} \frac{y^l(k_0, \eta)}{Y^l} d\Gamma$$

Similarly, $\tilde{g}^l = \tilde{g}^l_{pct}$ is the (relative) average labor income weighted by the marginal retention rate $(1 - \tau_l'(y^l))$ within percentile pct.

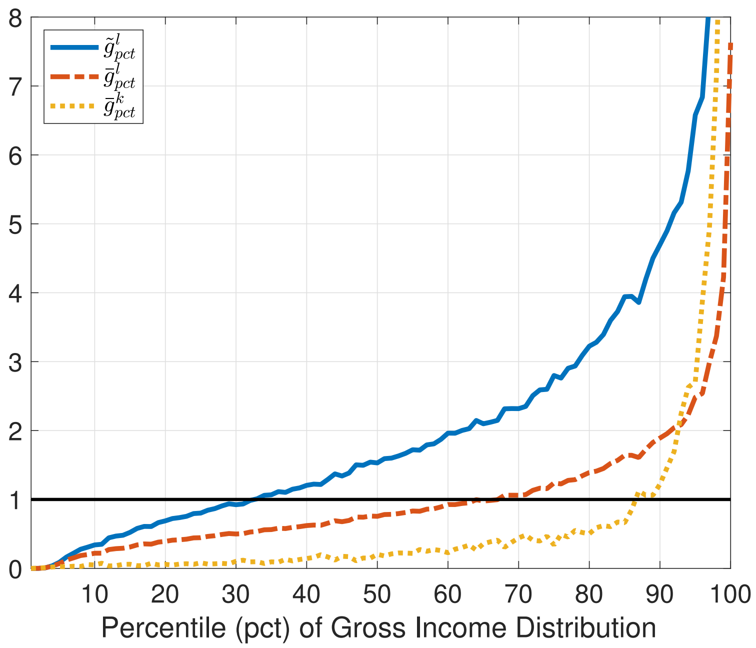

Figure 4: Weighted Average Marginal Social Welfare: $\bar{g}^k_{pct}$ ($\bar{g}^l_{pct}$): average capital (labor) income in percentile pct of the total income distribution as fraction of capital (labor) income in the whole population; $\tilde{g}^l_{pct}$: average net-of-marginal-tax-weighted labor income in percentile pct relative to average in population.

Figure 4 depicts $\bar{g}^k_{pct}$, $\bar{g}^l_{pct}$ and $\tilde{g}^l_{pct}$, with the percentiles pct on the x-axis. Since gross income is naturally positively correlated with its components, both the labor- and the capital income weighted average marginal social welfare weight are increasing in pct. Observe that the 87th percentile of the gross income distribution owns average wealth ($\bar{g}^k_{87} = 1$), while the 65th percentile of the gross income distribution earns about average labor income ($\bar{g}^l_{65} = 1$). Due to the progressivity of the labor tax code, however, already the 32th percentile of the income distribution earns the net-of-marginal-tax-rate weighted average labor income ($\tilde{g}^l_{32} = 1$).

Figure 5 plots the components of the equity effect EQ for each of these social welfare functions, where the values pct on the horizontal axis again refer to the welfare function that concentrates all welfare weight in percentile pct of the total income distribution. The left and right panels capture the cases with, respectively, exogenous and endogenous prices.

Mechanical Redistribution. The solid black line depicts the mechanical effect

$$EQ_M = 1 - \bar{g}^k,$$

which measures the change in welfare if agents' consumption, savings and labor supply were to be fixed at their pre-reform level. Naturally, the mechanical effect is identical across the two cases, since absent behavioral responses, prices are unaffected. The lower thirty percent of the gross income distribution do not earn any significant capital income, implying that a planner who only values those individuals does not discount the mechanically raised dollar, that is $\bar{g}^k_{pct} \approx 0$ for $pct < 30$. As one moves up the total gross income distribution, households earn more and more capital income, implying that $\bar{g}^k_{pct}$ is increasing in the percentile pct. However, since capital income is concentrated at the very top, the decline in the mechanical effect EQ_M is relatively modest until about income percentile 85, from which on households tend to have more substantial wealth and hence capital income. The skewness of the wealth distribution implies that the mean capital income is much higher than the median. Specifically, the 87th percentile of the gross income distribution earns about the average capital income, which is the reason why the mechanical effect crosses the x-axis at the 87th percentile.

Effect of Redistributing Factor Price Changes. The overall equity effect comprises of the mechanical effect and the redistributional effect of factor price changes,

$$EQ = EQ_M + EQ_P,$$

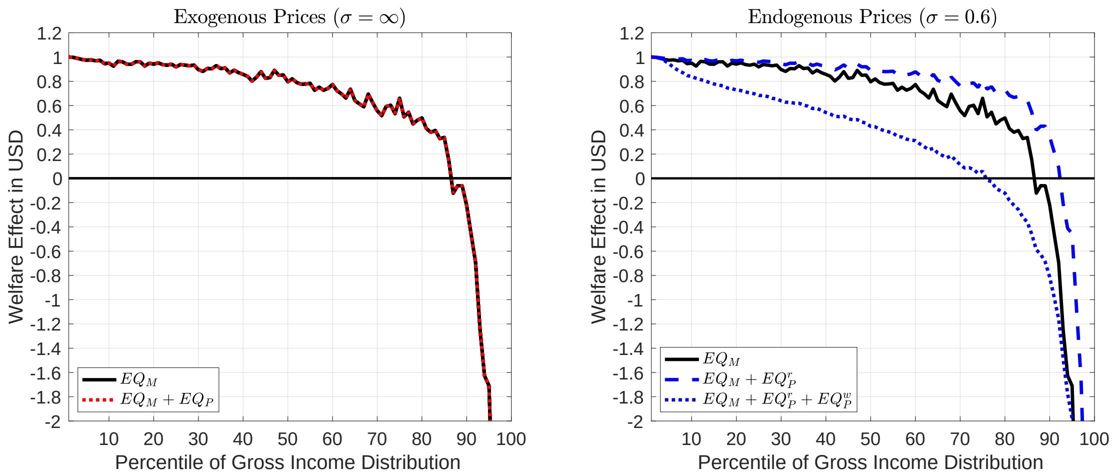

Figure 5: The Equity Effect: different substitution elasticities σ and Frisch elasticities γ_l; in USD per dollar of revenue mechanically raised; EQ_M: mechanical effect (red solid line, same for all σ), EQ_P : redistributional effect of factor price changes; value p on x-axis corresponds to the social welfare function that concentrates the whole welfare weight at percentile p of the total gross income distribution.

where the latter is given by

$$EQ_P = \underbrace{-(1-\tau_k)\bar{g}^k \bar{\varepsilon}_{r,1-\tau_k}}_{EQ_P^r \geq 0} \underbrace{-\alpha^l/\alpha^k(1-\bar{\tau}_l')\tilde{g}^l \bar{\varepsilon}_{w,1-\tau_k}}_{EQ_P^w \leq 0}.$$

I decompose this effect further into the welfare gain from the increase in interest rates ($EQ_P^r \geq 0$ since $\bar{\varepsilon}_{r,1-\tau_k} \leq 0$) and the welfare loss from a reduction in wages ($EQ_P^w \leq 0$ since $\bar{\varepsilon}_{w,1-\tau_k} \geq 0$).

When capital and labor are perfect substitutes (left panel of Figure 5), prices do not change, that is $\varepsilon_{w,1-\tau_k} = \varepsilon_{r,1-\tau_k} = 0$. Consequently, the price effect is zero ($EQ_P = 0$) and the total equity effect comprises only of the mechanical effect, $EQ = EQ_M$.

By contrast, whenever capital and labor are complements (right panel of Figure 5), a marginal increase in the capital tax rate reduces wages and increases interest rates, which redistributes from households with mostly wage income to households with mostly capital income. The dashed line adds to the mechanical effect the welfare gain from increases in the (gross) return to capital r. The additional effect is almost zero for the lower 30 percent of the income distribution, who earn basically zero capital income. Since the increase in the gross interest rate mitigates the reduction in the net return this effect is increasingly positive as further up the income distribution households receive more and more capital income. Yet, since capital income is concentrated at the very top the additional gain is relatively modest throughout the bottom 80 percent of the income distribution and starts to get significant only for the very highest earners.

Finally, the dotted line adds the welfare loss from the reduction in wages. Households at the very bottom of the income distribution do not have any earned income, that is neither capital- nor labor income. Hence for the bottom percentile the redistributive price effect is zero. However, the wage reduction implies welfare losses for households in the broad middle class, who finance their consumption predominantly with their net labor income. Furthermore, the adverse wage effect dominates the positive interest rate effect for all but the top 3 percent of the income distribution, a consequence of the fact that capital income is much more concentrated than labor income. Consequently, the total equity effect (the dotted line) is lower than the mechanical effect for all welfare functions but those that value the top three percent of earners. In sum, the decline in wages significantly reduces the mechanical redistributional gains from higher capital taxes for a very large middle class, say the middle 90% of the gross income distribution.

5.2.3 The Total Welfare Change

Simply adding the marginal excess burden to the equity effect gives the total welfare change per mechanically raised dollar in revenue. Figure 6 again depicts the case of exogenous prices on the left and the case with exogenous prices on the right. The dotted lines depict the equity effect EQ and are identical to the total equity effect in Figure 5. The solid lines add the marginal excess burden. As discussed above, the marginal excess burden is independent of the choice of social welfare function. Hence adding it results simply in a parallel downward downward shift along the y-axis.

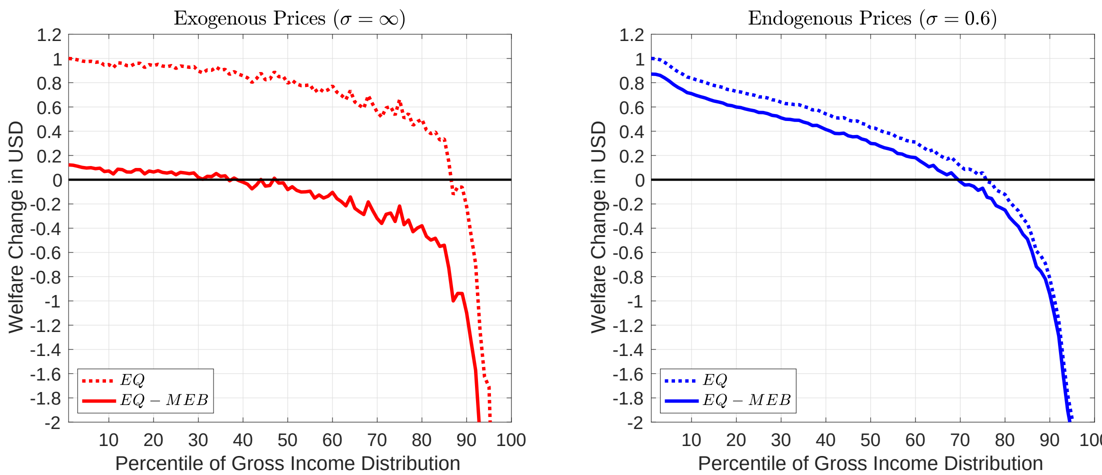

Figure 6: Welfare Change: in USD per dollar of revenue mechanically raised; EQ: equity effect, MEB: marginal excess burden; value p on x-axis corresponds to the social welfare function that concentrates the whole welfare weight at percentile p of the total gross income distribution; Frisch elasticity of labor supply: $\gamma_l = 0.5$.

In the case of exogenous prices (left panel), the marginal excess burden $MEB = 0.88$ is very large. Consequently, even for welfare objectives that value agents with about zero capital income, say the bottom thirty percent of the income distribution, the total welfare gains from capital tax increases is very small, while it is about zero for households between income percentiles 30 and 50.

Consider next the case with endogenous prices (right panel). As discussed above, the depressive effect of capital tax increases on wages causes the equity effect to decline much sharper as one moves up the income distribution. However, the much smaller excess burden of $MEB = 0.13$ implies that despite these depressive wage effect, the lower half of the income distribution experiences significant gains from capital tax increases, in stark contrast to the case with exogenous prices. The 69th income percentile finds the current tax rate just optimal. In Appendix D, I show that these results are quite robust to the whole range of estimates for the substitution elasticity and the wage elasticity of labor supply as well as to different assumptions on the return to capital. Specifically, for the whole empirical range of estimates would the bottom 60 percent of the total income distribution gain from increases in the capital tax rate, while the top 30 percent would lose.

5.3 Optimal Capital Tax Rates

In the absence of income effects on labor supply, the tax rate that satisfies condition (5) is given by

$$\tau_k = \frac{1 - \bar{g}^k - \frac{\alpha^l}{\alpha^k}\left(\bar{\epsilon}_{L,1-\tau_k}\mathrm{E}_\Gamma[\tau_l'] + \left[\tilde{g}^l - \bar{g}^k + (1-\tilde{g}^l)\bar{\tau}_l'\right]\bar{\epsilon}_{w,1-\tau_k}\right)}{\bar{\epsilon}_{K,1-\tau_k} + (1-\bar{g}^k)\left(1+\bar{\epsilon}_{r,1-\tau_k}\right)}, \tag{9}$$

while the corresponding tax rate assuming exogenous prices is

$$\tau_k^{ex} = \frac{1 - \bar{g}^k}{\bar{\epsilon}_{K,1-\tau_k} + (1-\bar{g}^k)}. \tag{10}$$

Figure 7 depicts these tax rates for each of the social welfare functions. The solid red line depicts the case if one assumed that prices were exogenous (τ_k^{ex}), while the solid blue line takes into account the endogeneity of factor prices (τ_k).

It has become the standard in the sufficient statistics literature to interpret these as the "optimal" tax rates (Piketty and Saez, 2012, 2013; Saez and Stantcheva, 2018). However, such an interpretation implicitly assumes that the (endogenous) statistics that enter the right hand side of the formula are invariant to tax changes.[22] It might thus lead to distorted policy prescriptions, especially when the so obtained "optimal" tax rate is far away from the existing one, around which the statistics are measured.[23] To address this problem, I also implement another, complementary, analysis. Specifically, I calibrate a parametric version of the model such that it replicates (around the status quo tax system) all the statistics entering the above formula. Using global solutions methods analogous to the ones used in the parametric macroeconomics literature (Domeij and Heathcote, 2004; Conesa et al., 2009; etc.) I then compute the full transitional dynamics following one-off tax changes in τ_k. For each of these reforms, I compute the welfare change for all individuals and hence obtain the 'truly' optimal capital tax rate from the perspective of households in each percentile of the income distribution. These tax rates are depicted by the dashed blue line in Figure 7. See Appendix E for more details on this procedure.

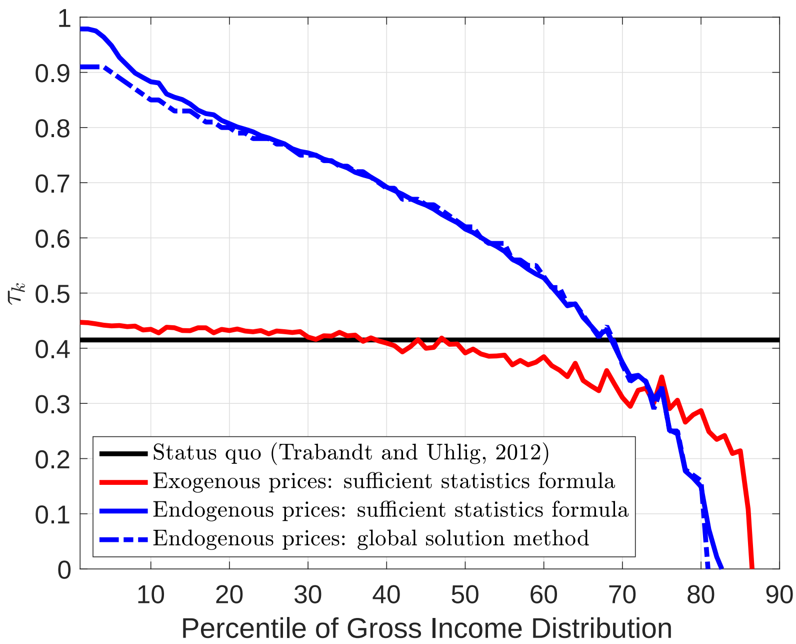

Figure 7: Optimal Capital Tax Rates: value p on the x-axis corresponds to the social welfare function that concentrates the whole welfare weight at percentile p of the total gross income distribution; capital-labor substitution elasticities $\sigma = 0.6$ (endogenous prices) and $\sigma = \infty$ (exogenous prices); benchmark Frisch elasticity of labor supply ($\gamma_l = 0.5$).

In any case, we observe that the sufficient statistics formula provides a very good approximation of the optimal tax rates. Only for objectives, which maximize the welfare

[22] Note that the distinction between semi-elasticities and elasticities matters for such an analysis. In line with the sufficient statistics literature, I assume that the elasticities $\bar{e}_{x,1-\tau_k}$, rather than the semi-elasticities $\bar{\varepsilon}_{x,1-\tau_k}$, are constant. As shown in Figure 7, this approximates the tax rates obtained through global numerical optimization methods quite well.

[23] For a more thorough discussion on this issue see, for example, Section 3.3 in Kleven (2021).

of the bottom 10 percent, are the discrepancies somewhat significant. For example, while the 'true' optimal Rawlsian tax rate is around 91%, the sufficient statistic formula suggests that it is close to 98%.

Otherwise, we observe a similar pattern as with the local welfare analysis above. Specifically, while the exogenous price case suggests that the current tax rate is approximately optimal for the bottom 60% of the income distribution, taking into account price endogeneity renders current tax rates too low for this part of the population. The reason is, as discussed above, the significantly lower excess burden. Furthermore, contrary to the exogenous price case, optimal tax rates are strongly decreasing in income when prices are endogenous. For example, with exogenous prices, the "optimal" tax rate from the perspective of the very bottom of the income distribution is around 44%, only 6-7% higher than what households in percentile 60 of the income distribution would find optimal. By contrast, with endogenous prices, the corresponding optimal tax rates decrease by almost 40%, from above 90% (Rawlsian) to about 53% (welfare objective maximizing welfare of percentile 60). As explained above, the main reason is the depressing effect of capital tax increases on wages. As one moves up the income distribution the net income loss due to the decrease in wages tends to become more and more important relative to the gain in transfer income. Households around the 70th percentile find the current tax rate approximately optimal, while higher income households would like to see capital tax reductions. Appendix D shows that these results are quite robust to different assumptions on the return to capital as well as to the range of empirical estimates for the capital-labor substitution elasticity and the wage-elasticity of labor supply.

6 Conclusion

In this paper, I derive an intuitive, testable and robust condition for the optimality of capital income tax rates. I apply my theoretical results to US income and wealth data and find that the majority of the US population, at least the bottom 60% of the income distribution, would benefit from significant capital tax increases relative to the status quo. Due to their depressing effect on wages, however, the desired capital tax rates across this part of the population are strongly declining in labor income.

While the condition I derive holds for a battery of standard macroeconomic models, an interesting further generalization would be to allow for heterogeneous labor skill types that exhibit different degrees of substitutability with capital. Since empirically high-skilled labor exhibits higher complementarity with capital than low-skilled labor (Krusell, Ohanian, Ríos-Rull, and Violante, 2000), in such an environment capital tax increases should reduce the wages of low earners by less than those of higher earners.

Relative to the quantitative results in this paper this should further increase optimal capital tax rates for welfare objectives that assign high weight on households in the lower middle class, i.e. households who finance their consumption mostly through wages – rather than government transfers – but who do not earn particularly much.

References

Açıkgöz, O., M. Hagedorn, H. Holter, and Y. Wang (2022). The Optimum Quantity of Captial and Debt. *Working Paper*.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *Quaterly Journal of Economics 109(3)*, 659–684.

Aiyagari, S. R. (1995). Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting. *Journal of Political Economy 103(6)*, 1158–1175.

Ales, L., M. Kurnaz, and C. Sleet (2015). Technical Change, Wage Inequality, and Taxes. *American Economic Review 105(10)*, 3061–3101.

Angeletos, G. M. (2007). Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic Dynamics 10(1)*, 1–30.

Antras, P. (2004). Is the U.S. Aggregate Production Function Cobb-Douglas? *B.E. Journal of Macroeconomics 4(1)*, 1–36.

Atkeson, A., V. Chari, and P. Kehoe (1999). Taxing Capital Income: A Bad Idea. *Federal Reserve Bank of Minneapolis Quarterly Review*, 3–17.

Badel, A. and M. Huggett (2017). The Sufficient Statistics Approach: Predicting the Top of the Laffer Curve. *Journal of Monetary Economics 87*, 1–12.

Benhabib, J., A. Bisin, and S. Zhu (2011). The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents. *Econometrica 79(1)*, 123–157.

Benhabib, J., A. Bisin, and S. Zhu (2015). The Wealth Distribution in Bewley Economies with Capital Income Risk. *Journal of Economic Theory 159*, 489–515.

Benhabib, J. and B. Szőke (2021). Optimal Positive Capital Taxes at Interior Steady States. *American Economic Journal: Macroeconomics 13(1)*, 114–150.

Browning, M. and T. Crossley (2001). The Life-Cycle Model of Consumption and Saving. *Journal of Economic Perspectives 15(3)*, 3–22.

Browning, M. and A. Lusardi (1996). Household Saving: Micro Theories and Micro Facts. *Journal of Economic Literature 34(4)*, 1797–1855.

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

Chari, V., P. Nicolini, and P. Teles (2020). Optimal Capital Taxation Revisited. *Journal of Monetary Economics 116*, 147–165.

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Structural and Reduced-Form Methods. *Annual Review of Economics 1(1)*, 451–488.

Chirinko, R. (2008). σ: The Long and the Short of It. *Journal of Macroeconomics 30(2)*, 671–686.

Conesa, J. C., S. Kitao, and D. Krueger (2009). Taxing Capital: Not a Bad Idea After All. *American Economic Review 99(1)*, 25–48.

Domeij, D. and J. Heathcote (2004). On the Distributional Effects of Reducing Capital Taxes. *International Economic Review 45(2)*, 523–554.

Dyrda, S. and M. Pedroni (2022). Optimal Fiscal Policy in a Model with Uninsurable Idiosyncratic Shocks. *Review of Economic Studies*, forthcoming.

Farhi, E. and I. Werning (2013). Insurance and Taxation over the Life Cycle. *Review of Economic Studies 80(2)*, 596–635.

Gechert, S., T. Havranek, Z. Irsova, and D. Kolcunova (2022). Measuring Capital-Labor Substitution: The Importance of Method Choice and Publication Bias. *Review of Economic Dynamics 45*, 55–82.

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003). Optimal Indirect Capital Taxation. *Review of Economic Studies 70(3)*, 569–587.

Golosov, M., A. Tsyvinski, and N. Werquin (2014). A Variational Approach to the Analysis of Tax Systems. *NBER Working Paper No. 20780*.

Guerreiro, J., S. Rebelo, and P. Teles (2022). Should Robots be Taxed? *Review of Economic Studies 89(1)*, 279–311.

Havranek, T. and D. Sokolova (2020). Do Consumers Really Follow a Rule of Thumb? Three Thousand Estimates from 144 Studies Say "Probably Not". *Review of Economic Dynamics 35*, 97–122.

Heathcote, J., K. Storesletten, and G. Violante (2017). Optimal Tax Progressivity: An Analytical Framework. *Quarterly Journal of Economics 132(4)*, 1693–1754.

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

Jakobsen, K., K. Jakobsen, H. Kleven, and G. Zucman (2020). Wealth Taxation and Wealth Accumulation: Theory and Evidence From Denmark. *Quarterly Journal of Economics 135(1)*, 329–388.

Jones, L., R. Manuelli, and P. Rossi (1997). On the Optimal Taxation of Capital Income. *Journal of Economic Theory 73*, 93–117.

Jorda, O., K. Knoll, D. Kuvshinov, M. Schularick, and A. Taylor (2019). The Rate of Return on Everything, 1870 to 2015. *Quarterly Journal of Economics 134(3)*, 1225–1298.

Judd, K. (1985). Redistributive Taxation in a Simple Foresight Model. *Journal of Public Economics 28(1)*, 59–83.

Kina, O., C. Slavik, and H. Yazici (2020). Redistributive Capital Taxes Revisited. *CESifo Working Paper No. 8627*.

Kleven, H. J. (2021). Sufficient statistics revisited. *Annual Review of Economics 13*(1), 515–538.

Krusell, P., L. Ohanian, V. Ríos-Rull, and G. Violante (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica 68(5)*, 1029–1053.

Krusell, P. and A. Smith (1998). Income and Wealth Heterogeneity in the Macroeconomy. *Journal of Political Economy 106(5)*, 867–896.

Lucas, R. (1990). Supply-Side Economics: An Analytical Review. *Oxford Economic Papers 42(2)*, 293–316.

Oberfield, E. and D. Raval (2021). Micro Data and Macro Technology. *Econometrica 89(2)*, 703–732.

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

Piketty, T. and E. Saez (2012). A Theory of Optimal Capital Taxation. *NBER Working Paper No. 17989*.

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

Rognlie, M. (2015). Deciphering the Fall and Rise in the Net Capital Share: Accumulation or Scarcity? *Brookings Papers on Economic Activity 46(1)*, 1–46.

Rothschild, C. and F. Scheuer (2013). Redistributive Taxation in the Roy Model. *Quarterly Journal of Economics 128(2)*, 623–668.

Sachs, D., A. Tsyvinski, and N. Werquin (2020). Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium. *Econometrica 88(2)*, 469–493.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies 68(1)*, 205–229.

Saez, E. and S. Stantcheva (2018). A Simpler Theory of Optimal Capital Taxation. *Journal of Public Economics 162*, 120–142.

Scheuer, F. (2014). Entrepreneurial Taxation with Endogenous Entry. *American Economic Journal: Economic Policy 6(2)*, 126–163.

Scheuer, F. and I. Werning (2017). The Taxation of Superstarts. *Quarterly Journal of Economics 132(1)*, 211–270.

Slavik, C. and H. Yazici (2014). Machines, Buildings, and Optimal Dynamic Taxes. *Journal of Monetary Economics 66*, 47–61.

Stantcheva, S. (2020). Dynamic Taxation. *Annual Review of Economics 12*, 801–831.

Straub, L. and I. Werning (2020). Positive Long-Run Capital Taxation: Chamley-Judd Revisited. *American Economic Review 110(1)*, 86–119.

Thümmel, U. (2020). Optimal Taxation of Robots. *CESifo Working Paper No. 7317*.

Trabandt, M. and H. Uhlig (2012). How do Laffer Curves Differ across Countries? *Fiscal Policy after the Financial Crisis*, 211–249.

Xavier, I. (2021). Wealth Inequality in the US: The Role of Heterogeneous Returns. *Working Paper*.

A Proofs

A.1 Proofs of Theoretical Results in the Main Text

A.1.1 Proof of Lemma 1

Proof. In equilibrium $r_t = F_k(K_t, L_t) - \delta$ and $w_t = F_l(K_t, L_t)$. Hence

$$\frac{dr_t}{d(1-\tau_k^r)} = F_{kk}(K_t, L_t)\frac{dK_t}{d(1-\tau_k^r)} + F_{kl}(K_t, L_t)\frac{dL_t}{d(1-\tau_k^r)}$$

and

$$\frac{dw_t}{d(1-\tau_k^r)} = F_{kl}(K_t, L_t)\frac{dK_t}{d(1-\tau_k^r)} + F_{ll}(K_t, L_t)\frac{dL_t}{d(1-\tau_k^r)}.$$

Since F is homogeneous of degree one, F_k and F_l are both homogeneous of degree zero, implying that $F_{kk}(K_t, L_t)K_t + F_{kl}(K_t, L_t)L_t = 0$ and $F_{kl}(K_t, L_t)K_t + F_{ll}(K_t, L_t)L_t = 0$. These conditions, in turn, imply that $F_{kk}(K_t, L_t) = -\frac{L_t}{K_t}F_{kl}(K_t, L_t)$ and $F_{ll}(K_t, L_t) = -\frac{K_t}{L_t}F_{kl}(K_t, L_t)$. Plugging these into the conditions above gives

$$\frac{dr_t}{d(1-\tau_k^r)} = \left[-\frac{L_t}{K_t}\frac{dK_t}{d(1-\tau_k^r)} + \frac{dL_t}{d(1-\tau_k^r)}\right]F_{kl}(K_t, L_t)$$

and

$$\frac{dw_t}{d(1-\tau_k^r)} = \left[\frac{dK_t}{d(1-\tau_k^r)} - \frac{K_t}{L_t}\frac{dL_t}{d(1-\tau_k^r)}\right]F_{kl}(K_t, L_t),$$

which is equivalent to

$$\epsilon_{r_t,1-\tau_k} = \frac{L_t F_{kl}(K_t, L_t)}{r_t}\left[-\epsilon_{K_t,1-\tau_k} + \epsilon_{L_t,1-\tau_k}\right]$$

and

$$\epsilon_{w_t,1-\tau_k} = \frac{K_t F_{kl}(K_t, L_t)}{w_t}\left[\epsilon_{K_t,1-\tau_k} - \epsilon_{L_t,1-\tau_k}\right]$$

Using the fact that

$$\sigma_t = \frac{F_k(K_t, L_t)F_l(K_t, L_t)}{F(K_t, L_t)F_{kl}(K_t, L_t)},$$

these expressions are, in turn, is equivalent to the desired expressions

$$\epsilon_{r_t,1-\tau_k} = \frac{\tilde{\alpha}_t^l}{\sigma}\frac{r_t+\delta}{r_t}\left[\epsilon_{L_t,1-\tau_k} - \epsilon_{K_t,1-\tau_k}\right]$$

and

$$\epsilon_{w_t,1-\tau_k} = \frac{\tilde{\alpha}_t^k}{\sigma}\left[\epsilon_{K_t,1-\tau_k} - \epsilon_{L_t,1-\tau_k}\right].$$

Finally, observe that

$$\begin{aligned}
\epsilon_{r_t,1-\tau_k} =& \frac{\tilde{\alpha}_t^l}{\sigma}\frac{(r_t+\delta)K_t}{r_t K_t}\left[\epsilon_{L_t,1-\tau_k} - \epsilon_{K_t,1-\tau_k}\right] \\
=& \frac{\tilde{\alpha}_t^l}{\sigma}\frac{\tilde{\alpha}^k \tilde{Y}}{\alpha^k Y}\left[\epsilon_{L_t,1-\tau_k} - \epsilon_{K_t,1-\tau_k}\right] \\
=& \frac{wL}{Y}\frac{\tilde{\alpha}^k}{\alpha^k}\left[\epsilon_{L_t,1-\tau_k} - \epsilon_{K_t,1-\tau_k}\right] \\
=& \alpha^l \frac{\tilde{\alpha}^k}{\alpha^k}\left[\epsilon_{L_t,1-\tau_k} - \epsilon_{K_t,1-\tau_k}\right].
\end{aligned}$$

□

A.1.2 Proof of Proposition 1

Proof. For convenience, I use the shorthand notation

$$u_{x,t}(k_0,\eta) \equiv u_x\big(c_t(k_0,\eta), l_t(k_0,\eta)\big)$$

for $x \in \{c, l, cc, ll, cl\}$ throughout this proof.

I will also make use of the households' optimality conditions, that is the intra-temporal labor supply condition,

$$\eta w_t\big(1 - \tau_l'(\eta w_t l_t(k_0,\eta))\big)u_{c,t}(k_0,\eta) = -u_{l,t}(k_0,\eta),$$

and the intertemporal Euler equation

$$u_{c,t}(k_0,\eta) = \beta[1 + (1-\tau_{k,t+1})r]u_{c,t+1}(k_0,\eta).$$

The planner solves

$$\max_{\tau_k^r \le 1} W = \max_{\tau_k^r \le 1} \int \omega(k_0,\eta) \sum_{t=0}^{\infty} \beta^t u\big(c_t(k_0,\eta), l_t(k_0,\eta)\big) d\Gamma$$

The first order condition with respect to $1-\tau_k^r$ is

$$\frac{dW}{d(1-\tau_k^r)} = \int \omega(k_0,\eta) \sum_{t=0}^{\infty} \beta^t \left[u_{c,t}(k_0,\eta)\frac{dc_t(k_0,\eta)}{d(1-\tau_k^r)} + u_{l,t}(k_0,\eta)\frac{dl_t(k_0,\eta)}{d(1-\tau_k^r)}\right] d\Gamma = 0.$$

The households' budget constraint is given by

$$c_t(k_0,\eta) + k_{t+1}(k_0,\eta) = [1 + (1-\tau_{k,t})r_t]k_t(k_0,\eta) + \eta_t w_t l_t(k_0,\eta) - \tau_l(\eta_t w_t l_t(k_0,\eta)) + T_t.$$

Deriving with respect to $(1-\tau_k^r)$ gives

$$\begin{aligned}\frac{dc_t(k_0,\eta)}{d(1-\tau_k^r)} =& -\frac{dk_{t+1}(k_0,\eta)}{d(1-\tau_k^r)} + 1_{t\geq t^a} r_t k_t(k_0,\eta) + \big(1 + (1-\tau_{k,t})r_t\big)\frac{dk_t(k_0,\eta)}{d(1-\tau_k^r)} \\ &+ (1-\tau_{k,t})\frac{dr_t}{d(1-\tau_k^r)}k_t(k_0,\eta) + \eta_t w_t\big(1-\tau_l'(\eta_t w_t l_t(k_0,\eta))\big)\frac{dl_t(k_0,\eta)}{d(1-\tau_k^r)} \\ &+ \eta_t l_t(k_0,\eta)\big(1-\tau_l'(\eta_t w_t l_t(k_0,\eta))\big)\frac{dw_t}{d(1-\tau_k^r)} + \frac{dT_t}{d(1-\tau_k^r)},\end{aligned}$$

where $1_{t\geq t^a}$ denotes an indicator function that takes the value one if $t \geq t^a$ and the value zero otherwise. The government transfer in period t is given by

$$T_t + G = \tau_{k,t} r_t K_t + \int \tau_l(\eta w_t l_t(k_0,\eta))d\Gamma.$$

Deriving it with respect to $(1-\tau_k^r)$ gives

$$\begin{aligned}\frac{dT_t}{d(1-\tau_k^r)} =& -r_t K_t 1_{t\geq t^a} + \tau_{k,t} r_t \frac{dK_t}{d(1-\tau_k^r)} + \tau_{k,t} K_t \frac{dr_t}{d(1-\tau_k^r)} \\ &+ \int \tau_l'(\eta w_t l_t(k_0,\eta))\left[\eta w_t \frac{dl_t(k_0,\eta)}{d(1-\tau_k^r)} + \eta l_t(k_0,\eta)\frac{dw_t}{d(1-\tau_k^r)}\right]d\Gamma.\end{aligned}$$

Plugging the expression for $\frac{dc_t(k_0,\eta)}{d(1-\tau_k^r)}$ into the first order condition of the planner then gives

$$\begin{aligned}\frac{dW}{d(1-\tau_k^r)} =& \\ &\sum_{t=0}^{\infty}\beta^t \int \omega(k_0,\eta)u_{c,t}(k_0,\eta)\Big\{r_t(k_t - K_t)1_{t\geq t^a} + (1-\tau_{k,t})\frac{dr_t}{d(1-\tau_k)}k_t(k_0,\eta) + \tau_{k,t}K_t\frac{dr_t}{d(1-\tau_k)} \\ &\quad + \eta l_t(k_0,\eta)\big(1-\tau_l'(\eta w_t l_t(k_0,\eta))\big)\frac{dw_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w_t l_t(k_0,\eta))\eta l_t(k_0,\eta)\frac{dw_t}{d(1-\tau_k^r)}d\Gamma \\ &\quad + \tau_{k,t} r_t \frac{dK_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w_t l_t(k_0,\eta))\eta w_t \frac{dl_t(k_0,\eta)}{d(1-\tau_k^r)}d\Gamma\Big\}d\Gamma = 0.\end{aligned}$$

Note that the derivatives $\frac{dk_t(k_0,\eta)}{d(1-\tau_k^r)}$ and $\frac{dl_t(k_0,\eta)}{d(1-\tau_k^r)}$ all drop out because of households' optimization behavior (envelope conditions).

By Lemma 1 we have

$$\frac{dr_t}{d(1-\tau_k^r)} = \frac{L_t}{K_t}\frac{dw_t}{d(1-\tau_k^r)}.$$

Hence, the first order condition can be further simplified to

$$\frac{dW}{d(1-\tau_k^r)} = \sum_{t=0}^{\infty}\beta^t \int \omega(k_0,\eta)u_{c,t}(k_0,\eta)\Big\{r_t(k_t-K_t)1_{t\geq t^a} + \frac{dw_t}{d(1-\tau_k^r)}\Big[-L_t(1-\tau_{k,t})\frac{k_t(k_0,\eta)}{K_t} - \tau_{k,t}L_t + \eta l_t(k_0,\eta)\big(1-\tau_l'(\eta w_t l_t(k_0,\eta))\big) + \int \tau_l'(\eta w_t l_t(k_0,\eta))\eta l_t(k_0,\eta)d\Gamma\Big] + \tau_{k,t}r_t\frac{dK_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w_t l_t(k_0,\eta))\eta w_t\frac{dl_t(k_0,\eta)}{d(1-\tau_k^r)}d\Gamma\Big\}d\Gamma = 0.$$

Assumption 3 implies that we can evaluate this condition at the initial steady state.[A.1] Hence,

$$\frac{dW}{d(1-\tau_k^r)} = \sum_{t=0}^{\infty}\beta^t \int \omega(k_0,\eta)u_c(k_0,\eta)\Big\{r(k_0-K)1_{t\geq t^a} + \frac{dw_t}{d(1-\tau_k^r)}L\Big[-(1-\tau_k)\frac{k_0}{K} - \tau_k + \frac{\eta l(k_0,\eta)}{L}\big(1-\tau_l'(\eta w l(k_0,\eta))\big) + \int \tau_l'(\eta w l(k_0,\eta))\frac{\eta l(k_0,\eta)}{L}d\Gamma\Big] + \tau_k r\frac{dK_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w l(k_0,\eta))\eta w\frac{dl_t(k_0,\eta)}{d(1-\tau_k^r)}d\Gamma\Big\}d\Gamma = 0,$$

which using that $\omega(k_0,\eta)u_c(k_0,\eta) = g(k_0,\eta)$ and the definition for the income weighted average marginal labor tax rate

$$\bar{\tau}_l' = \frac{\int \tau_l'(k_0,\eta)\eta w l(k_0,\eta)d\Gamma}{wL},$$

is equivalent to

$$\frac{dW}{d(1-\tau_k^r)} = \sum_{t=0}^{\infty}\beta^t \int g(k_0,\eta)\Big\{r(k_0-K)1_{t\geq t^a} + \tau_k r\frac{dK_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w l(k_0,\eta))\eta w\frac{dl_t(k_0,\eta)}{d(1-\tau_k^r)}d\Gamma + \frac{dw_t}{d(1-\tau_k^r)}L\Big[(1-\tau_k)\Big(1-\frac{rk_0}{rK}\Big) - (1-\bar{\tau}_l')\Big(1-\frac{(1-\tau_l'(\eta w l(k_0,\eta)))\eta w l(k_0,\eta)}{wL}\Big)\Big]\Big\}d\Gamma = 0.$$

Using the normalization $\int g(k_0,\eta)d\Gamma = 1$ and multiplying by $\frac{1-\beta}{rK}$ gives

$$\frac{dW}{d(1-\tau_k^r)}Y^k = \beta^{t^a}(\bar{g}^k - 1) + \frac{\tau_k}{1-\tau_k}(1-\beta)\sum_{t=0}^{\infty}\beta^t\epsilon_{K_t,1-\tau_k} + \frac{1}{1-\tau_k}\frac{wL}{rK}(1-\beta)\sum_{t=0}^{\infty}\beta^t\int \tau_l'(\eta w l(k_0,\eta))\frac{\eta l(k_0,\eta)}{L}\epsilon_{l_t(k_0,\eta),1-\tau_k}d\Gamma + \Big[(1-\bar{g}^k) - \frac{1-\bar{\tau}_l'}{1-\tau_k}(1-\bar{g}^l)\Big](1-\beta)\sum_{t=0}^{\infty}\beta^t\epsilon_{w_t,1-\tau_k}\frac{wL}{rK} = 0.$$

[A.1] This implies that we can drop the time indices of equilibrium variables but not of their derivatives. Transitional dynamics induced by changing capital taxes are hence accounted for.

Now note that

$$(1-\beta)\sum_{t=0}^{\infty}\beta^t\int\tau_l'(\eta wl(k_0,\eta))\frac{\eta l(k_0,\eta)}{L}\epsilon_{l_t(k_0,\eta),1-\tau_k}d\Gamma =$$
$$\mathrm{Cov}_\Gamma\left(\tau_l'(k_0,\eta),\frac{y^l(k_0,\eta)}{Y^l}\bar{\epsilon}_{l(k_0,\eta),1-\tau_k}\right)+\mathrm{E}_\Gamma[\tau_l']\bar{\epsilon}_{L,1-\tau_k}.$$

Hence we obtain

$$\frac{dW}{d(1-\tau_k^r)}Y^k=\beta^{t^a}(\bar{g}^k-1)+\frac{\tau_k}{1-\tau_k}\bar{\epsilon}_{K,1-\tau_k}+\frac{1}{1-\tau_k}\frac{wL}{rK}\left[\mathrm{Cov}_\Gamma\left(\tau_l',\frac{y^l}{Y^l}\bar{\epsilon}_{l,1-\tau_k}\right)+\mathrm{E}_\Gamma[\tau_l']\bar{\epsilon}_{L,1-\tau_k}\right]$$
$$+\bar{\epsilon}_{w,1-\tau_k}\frac{wL}{rK}\left[(1-\bar{g}^k)-\frac{1-\bar{\tau}_l'}{1-\tau_k}(1-\bar{g}^l)\right]=0.$$

Using the definitions

$$\begin{aligned}
EQ_M =&\beta^{t^a}(1-\bar{g}^k)\\
MEB_K =&\frac{\tau_k}{1-\tau_k}\bar{\epsilon}_{K,1-\tau_k}=\tau_k\bar{\varepsilon}_{K,1-\tau_k}\\
MEB_L =&\frac{1}{1-\tau_k}\frac{wL}{rK}\left[\mathrm{E}_\Gamma[\tau_l']+\mathrm{Cov}_\Gamma\left(\tau_l',\frac{y^l}{Y^l}\bar{\epsilon}_{l,1-\tau_k}\right)\right]=\frac{\alpha^l}{\alpha^k}\bar{\varepsilon}_{L,1-\tau_k}\left[\mathrm{E}_\Gamma[\tau_l']+\mathrm{Cov}_\Gamma\left(\tau_l',\frac{y^l}{Y^l}\frac{\bar{\varepsilon}_{l,1-\tau_k}}{\bar{\varepsilon}_{L,1-\tau_k}}\right)\right]\\
P =&-\bar{\epsilon}_{w,1-\tau_k}\frac{wL}{rK}\left[(1-\bar{g}^k)-\frac{1-\bar{\tau}_l'}{1-\tau_k}(1-\bar{g}^l)\right]=\bar{\varepsilon}_{w,1-\tau_k}\frac{\alpha^l}{\alpha^k}\left[(1-\tau_k)\bar{g}^k-(1-\bar{\tau}_l')\bar{g}^l-(\bar{\tau}_l'-\tau_k)\right]\\
EQ_P =&\bar{\varepsilon}_{w,1-\tau_k}\frac{\alpha^l}{\alpha^k}\left[(1-\tau_k)\bar{g}^k-(1-\bar{\tau}_l')\bar{g}^l\right]\\
MEB_P =&=\bar{\varepsilon}_{w,1-\tau_k}\frac{\alpha^l}{\alpha^k}\left[\bar{\tau}_l'-\tau_k\right]
\end{aligned}$$

this is equivalent to

$$\frac{dW}{d\tau_k^r}Y^k=EQ_M+EQ_P-\left[MEB_K+MEB_L+MEB_P\right]=0,$$

proving both the decomposition (4) of the welfare change as well as the optimality condition (5) for the optimal tax rate,

$$\begin{aligned}
\tau_k =&\frac{\beta^{t^a}(1-\bar{g}^k)-MEB_L+P}{\bar{\varepsilon}_{K,1-\tau_k}}\\
=&\frac{\beta^{t^a}(1-\bar{g}^k)}{\bar{\varepsilon}_{K,1-\tau_k}}-\frac{\bar{\varepsilon}_{L,1-\tau_k}}{\bar{\varepsilon}_{K,1-\tau_k}}\frac{\alpha^l}{\alpha^k}\left[\mathrm{E}_\Gamma[\tau_l']+\frac{1}{Y^l\bar{\varepsilon}_{L,1-\tau_k}}\mathrm{Cov}_\Gamma\left(\tau_l',y^l\bar{\varepsilon}_{l,1-\tau_k}\right)\right]\\
&+\frac{\bar{\varepsilon}_{w,1-\tau_k}}{\bar{\varepsilon}_{K,1-\tau_k}}\frac{\alpha^l}{\alpha^k}\left[(1-\tau_k)\bar{g}^k-(1-\bar{\tau}_l')\bar{g}^l+\tau_k-\bar{\tau}_l'\right].
\end{aligned}$$

Finally, $\tau_k<1$ ensures that the solution is interior. Hence equation (5) is indeed a necessary condition for τ_k to be optimal. □

A.2 Proofs of Theoretical Results in the Appendix

A.2.1 Proof of Lemma B.1

Proof. The entrepreneur's profit is $\pi = \max_l\{F(k,l^d) - \delta k - wl^d\}$. The first order condition is $F_l(k,l^d) = w$. Linear homogeneity of F implies that $F_l\left(\frac{k}{l^d},1\right) = F_l\left(\frac{K^c}{L^c},1\right) = w$, i.e. all firms employ the same ratio of effective capital to effective labor, proving part (a).

Furthermore, linear homogeneity of F implies

$$F(k,l) = kF_k(k,l^d) + l^d F_l(k,l^d) = kF_k(k,l^d) + l^d w$$

by Euler's Theorem. Plugging into the profit function above gives

$$\pi = kF_k(k,l^d) - k\delta = kF_k(K^c,L^c) - k\delta = kr,$$

which proves part (b). □

A.2.2 Proof of Proposition B.1

Proof. For convenience, I use the shorthand notation

$$u_{z,t}(\mathbf{x}^t) \equiv u_{z,t}\big(c_t(\mathbf{x}^t;\Gamma_t), l_t(\mathbf{x}^t;\Gamma_t)\big)$$

for $z \in \{c, l, cc, ll, cl\}$ throughout this proof.

I will also make use of the households' optimality conditions. The intra-temporal labor supply condition is the same as in the framework of the main text, that is

$$\eta w_t\big(1 - \tau_l'(\eta w_t l_t(\mathbf{x}^t))\big)u_{c,t}(\mathbf{x}^t) = -u_{l,t}(\mathbf{x}^t).$$

However, now there are two intertemporal Euler equations that characterize optimal bond and private equity holdings, respectively. The Euler equation for bond holdings is given by

$$u_{c,t}(\mathbf{x}^t) = \beta \int p(\theta',\eta'|\theta,\eta)[1 + (1-\tau_{k,t+1})r]u_{c,t+1}(\mathbf{x}^t,\theta',\eta')d(\theta',\eta') + \mu_t^b(\mathbf{x}^t),$$

where $\mu_t^b(\mathbf{x}^t)$ is the Lagrange multiplier on the borrowing constraint $b_{t+1}(\mathbf{x}^t) \geq -\underline{b}(e_{t+1}(\mathbf{x}^t))$. The Euler equation for private equity holdings is given by

$$u_{c,t}(\mathbf{x}^t) = \beta \int p(\theta',\eta'|\theta,\eta)[1 + (1-\tau_{k,t+1})\theta' r]u_{c,t+1}(\mathbf{x}^t,\theta',\eta')d(\theta',\eta') + \mu_t^e(\mathbf{x}^t),$$

where $\mu_t^e(\mathbf{x}^t)$ is the Lagrange multiplier on the non-negativity constraint for private equity, $e_{t+1}(\mathbf{x}^t) \geq 0$.

The final two optimality conditions are the complementary slackness conditions

$$\mu_t^b(\mathbf{x}^t)\big(b_{t+1}(\mathbf{x}^t) + \underline{b}(e_{t+1}(\mathbf{x}^t))\big) = 0$$

and

$$\mu_t^e(\mathbf{x}^t)e_{t+1}(\mathbf{x}^t) = 0.$$

The social planner's problem $(\tilde{P})$ is equivalent to

$$\begin{aligned}&\max_{\tau_k\leq 1}\sum_{t=0}^{\infty}\beta^t\int\int \omega_t(\mathbf{x}^t)p(\mathbf{x}^t|\mathbf{x}_0)u\big(c_t(\mathbf{x}^t),l_t(\mathbf{x}^t)\big)d\mathbf{x}^t d\Gamma_0\\ &s.t. \quad T\geq \underline{T}.\end{aligned}$$

The first order condition with respect to $1-\tau_k^r$ is

$$\sum_{t=0}^{\infty}\beta^t\int\int \omega_t(\mathbf{x}^t)p(\mathbf{x}^t|\mathbf{x}_0)\left[u_c\big(c_t(\mathbf{x}^t)\big)\frac{dc_t(\mathbf{x}^t)}{d(1-\tau_k)} + u_c\big(l_t(\mathbf{x}^t)\big)\frac{dl_t(\mathbf{x}^t)}{d(1-\tau_k)}\right]d\mathbf{x}^t d\Gamma_0 = 0.$$

In each period t, the households' budget constraint is given by

$$\begin{aligned}c_t(\mathbf{x}^t) + e_{t+1}(\mathbf{x}^t) + b_{t+1}(\mathbf{x}^t) =& (1-\tau_{k,t})r_t\big[\theta_t e_t(\mathbf{x}^{t-1}) + b_t(\mathbf{x}^{t-1})\big] + e_t(\mathbf{x}^{t-1}) + b_t(\mathbf{x}^{t-1})\\ &+ \eta_t w_t l_t(\mathbf{x}^t) - \tau_l(\eta_t w_t l_t(\mathbf{x}^t)) + T_t.\end{aligned}$$

Deriving with respect to $(1-\tau_k^r)$ gives

$$\begin{aligned}\frac{dc_t(\mathbf{x}^t)}{d(1-\tau_k^r)} + \frac{de_{t+1}(\mathbf{x}^t)}{d(1-\tau_k^r)} + \frac{db_{t+1}(\mathbf{x}^t)}{d(1-\tau_k^r)} &= 1_{t\geq t^a}r_t k_t(\mathbf{x}^t) + (1-\tau_{k,t})r_t\left[\theta_t\frac{de_t(\mathbf{x}^{t-1})}{d(1-\tau_k^r)} + \frac{db_t(\mathbf{x}^{t-1})}{d(1-\tau_k^r)}\right]\\ +\frac{de_t(\mathbf{x}^{t-1})}{d(1-\tau_k^r)} + \frac{db_t(\mathbf{x}^{t-1})}{d(1-\tau_k^r)} &+ (1-\tau_{k,t})\frac{dr_t}{d(1-\tau_k^r)}k_t(\mathbf{x}^t) + \eta_t w_t\big(1-\tau_l'(\eta_t w_t l_t(k_0,\eta))\big)\frac{dl_t(\mathbf{x}^t)}{d(1-\tau_k^r)}\\ &+\eta_t l_t(\mathbf{x}^t)\big(1-\tau_l'(\eta_t w_t l_t(\mathbf{x}^t))\big)\frac{dw_t}{d(1-\tau_k^r)} + \frac{dT_t}{d(1-\tau_k^r)},\end{aligned}$$

where $1_{t\geq t^a}$ denotes an indicator function that takes the value one if $t\geq t^a$ and the value zero otherwise and

$$k_t(\mathbf{x}^t) \equiv \theta_t e_t(\mathbf{x}^{t-1}) + b_t(\mathbf{x}^{t-1}).$$

The government transfer in period t is given by

$$T_t + G = \tau_{k,t}r_t K_t + \int \tau_l(\eta w_t l_t(\mathbf{x}_t(\mathbf{x}^t)))d\Gamma_t,$$

where $\mathbf{x}_t(\mathbf{x}^t)$ denotes the state (a_t, θ_t, η_t) in period t that corresponds to history $\mathbf{x}^t$. Deriving it with respect to $(1-\tau_k^r)$ gives

$$\frac{dT_t}{d(1-\tau_k^r)} = -r_t K_t 1_{t \geq t^a} + \tau_{k,t} r_t \frac{dK_t}{d(1-\tau_k^r)} + \tau_{k,t} K_t \frac{dr_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w_t l_t(\mathbf{x}_t(\mathbf{x}^t))) \left[\eta w_t \frac{dl_t(\mathbf{x}_t(\mathbf{x}^t))}{d(1-\tau_k^r)} + \eta l_t(\mathbf{x}_t(\mathbf{x}^t)) \frac{dw_t}{d(1-\tau_k^r)}\right] d\Gamma_t.$$

Plugging the expression for $\frac{dc_t(\mathbf{x}^t)}{d(1-\tau_k^r)}$ into the first order condition of the planner then gives

$$\sum_{t=0}^{\infty} \beta^t \int\int \omega_t(\mathbf{x}^t) p(\mathbf{x}^t|\mathbf{x}_0) u_{c,t}(\mathbf{x}^t) \Big\{ r_t\big(k_t(\mathbf{x}^t) - K_t\big) 1_{t\geq t^a} + (1-\tau_{k,t}) \frac{dr_t}{d(1-\tau_k)} k_t(\mathbf{x}^t) + \tau_{k,t} K_t \frac{dr_t}{d(1-\tau_k)} + \eta l_t(\mathbf{x}^t)\big(1 - \tau_l'(\eta w_t l_t(\mathbf{x}^t))\big) \frac{dw_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w_t l_t(\mathbf{x}^t)) \eta l_t(\mathbf{x}^t) \frac{dw_t}{d(1-\tau_k^r)} d\Gamma + \tau_{k,t} r_t \frac{dK_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w_t l_t(\mathbf{x}_t(\mathbf{x}^t))) \eta w_t \frac{dl_t(\mathbf{x}_t(\mathbf{x}^t))}{d(1-\tau_k^r)} d\Gamma_t \Big\} d\mathbf{x}^t d\Gamma_0 = 0.$$

Note that the derivatives $\frac{dl_t(\mathbf{x}^t)}{d(1-\tau_k^r)}$, $\frac{de_{t+1}(\mathbf{x}^t)}{d(1-\tau_k^r)}$ and $\frac{db_{t+1}(\mathbf{x}^t)}{d(1-\tau_k^r)}$ all drop out because of households' optimization behavior (envelope conditions). Specifically, note that for histories $\mathbf{x}^t$, in which the Euler equation for bonds, respectively private equity, holds with strict inequality, we have $\frac{db_{t+1}(\mathbf{x}^t)}{d(1-\tau_k^r)} = 0$, respectively $\frac{de_{t+1}(\mathbf{x}^t)}{d(1-\tau_k^r)} = 0$.

Furthermore, linear homogeneity of F implies that $\frac{dr_t}{d(1-\tau_k)} = -\frac{L_t}{K_t} \frac{dw_t}{d(1-\tau_k)}$ (see Lemma 1). Hence, the first order condition can be further simplified to

$$\sum_{t=0}^{\infty} \beta^t \int\int \omega_t(\mathbf{x}^t) p(\mathbf{x}^t|\mathbf{x}_0) u_{c,t}(\mathbf{x}^t) \Big\{ r_t\big(k_t(\mathbf{x}^t) - K_t\big) 1_{t\geq t^a} + \frac{dw_t}{d(1-\tau_k^r)} \Big[-L_t(1-\tau_{k,t}) \frac{k_t(\mathbf{x}^t)}{K_t} - \tau_{k,t} L_t + \eta l_t(\mathbf{x}^t)\big(1 - \tau_l'(\eta w_t l_t(\mathbf{x}^t))\big) + \int \tau_l'(\eta w_t l_t(\mathbf{x}_t(\mathbf{x}^t))) \eta l_t(\mathbf{x}_t(\mathbf{x}^t)) d\Gamma_t \Big] + \tau_{k,t} r_t \frac{dK_t}{d(1-\tau_k^r)} + \int \tau_l'(\eta w_t l_t(\mathbf{x}_t(\mathbf{x}^t))) \eta w_t \frac{dl_t(\mathbf{x}_t(\mathbf{x}^t))}{d(1-\tau_k^r)} d\Gamma_t \Big\} d\mathbf{x}^t d\Gamma_0 = 0.$$

Using that if τ_k solves the planners problem we have that $\tau_{k,t} = \tau_k$ for all t and employing Assumption 3, this is equivalent to

$$\sum_{t=0}^{\infty} \beta^t \int\int \omega_t(\mathbf{x}^t) p(\mathbf{x}^t|\mathbf{x}_0) u_{c,t}(\mathbf{x}^t) \Big\{ \big(y_t^k(\mathbf{x}^t) - Y^k\big) 1_{t\geq t^a} + \frac{\tau_k}{1-\tau_k} Y^k \epsilon_{K_t, 1-\tau_k} + \epsilon_{w_t, 1-\tau_k} Y^l \Big[-\frac{y_t^k(\mathbf{x}^t)}{Y^k} - \frac{\tau_k}{1-\tau_k} + \frac{y_t^l(\mathbf{x}^t)}{Y^l} \frac{1 - \tau_l'(y_t^l(\mathbf{x}^t))}{1-\tau_k} + \frac{\bar{\tau}_l'}{1-\tau_k} \Big] + \frac{1}{1-\tau_k} \int \tau_l'(y_t^l(\mathbf{x})) y_t^l(\mathbf{x}) \epsilon_{l_t(\mathbf{x}), 1-\tau_k} d\Gamma \Big\} d\mathbf{x}^t d\Gamma = 0,$$

where

$$\bar{\tau}'_{l,t} = \frac{\int \tau_l'(\eta w_t l_t(\mathbf{x}_t(\mathbf{x}^t)))\eta w_t l_t(\mathbf{x}_t(\mathbf{x}^t))d\Gamma_t}{w_t L_t},$$

and because of Assumption 3 we have $\bar{\tau}'_{l,t} = \bar{\tau}'_l$.

Now use that $g_t(\mathbf{x}^t) = \omega_t(\mathbf{x}^t)u_{c,t}(\mathbf{x}^t)$ and that by normalization (B.1)

$$(1-\beta)\sum_{t=0}^{\infty}\beta^t \int\int \omega_t(\mathbf{x}^t)p(\mathbf{x}^t|\mathbf{x}_0)u_{c,t}(\mathbf{x}^t)d\mathbf{x}^t d\Gamma = 1$$

as well as

$$(1-\beta)\sum_{t=0}^{\infty}\beta^t \int \tau_l'(y_t^l(\mathbf{x}))\frac{y_t^l(\mathbf{x})}{Y^l}\epsilon_{l_t(\mathbf{x}),1-\tau_k}d\Gamma = \mathrm{E}_\Gamma[\tau_l']\bar{\epsilon}_{L,1-\tau_k} + \mathrm{Cov}_\Gamma\Big(\tau_l', \frac{y^l}{Y^l}\bar{\epsilon}_{l,1-\tau_k}\Big).$$

Multiplying the first order condition by $(1-\beta)$ then gives

$$\begin{aligned}(1-\beta)\sum_{t=0}^{\infty}\beta^t \int g_t(\mathbf{x}^t)\big(y_t^k(\mathbf{x}^t) - Y^k\big)1_{t\geq t^a}dP_t + \frac{\tau_k}{1-\tau_k}Y^k(1-\beta)\sum_{t=0}^{\infty}\beta^t\epsilon_{K_t,1-\tau_k}\bar{g}_t \\ +(1-\beta)\sum_{t=0}^{\infty}\beta^t\epsilon_{w_t,1-\tau_k}\int g_t(\mathbf{x}^t)Y^l\Big[-\frac{y_t^k(\mathbf{x}^t)}{Y^k} - \frac{\tau_k}{1-\tau_k} + \frac{y_t^l(\mathbf{x}^t)}{Y^l}\frac{1-\tau_l'(y_t^l(\mathbf{x}^t))}{1-\tau_k} + \frac{\bar{\tau}_l'}{1-\tau_k}\Big]dP_t \\ +\frac{Y^l}{1-\tau_k}\Big[\mathrm{E}_\Gamma[\tau_l']\bar{\epsilon}_{L,1-\tau_k} + \mathrm{Cov}_\Gamma\Big(\tau_l', \frac{y^l}{Y^l}\bar{\epsilon}_{l,1-\tau_k}\Big)\Big] = 0.\end{aligned}$$

Dividing by Y^k and rearranging terms gives

$$\begin{aligned}(1-\beta)\sum_{t=t^a}^{\infty}\beta^t \int g_t(\mathbf{x}^t)\Big(\frac{y_t^k(\mathbf{x}^t)}{Y^k} - 1\Big)dP_t + \frac{\tau_k}{1-\tau_k}(1-\beta)\sum_{t=0}^{\infty}\beta^t\epsilon_{K_t,1-\tau_k}\bar{g}_t \\ -\frac{\frac{Y^l}{Y^k}}{1-\tau_k}(1-\beta)\times \\ \sum_{t=0}^{\infty}\beta^t\epsilon_{w_t,1-\tau_k}\int g_t(\mathbf{x}^t)\Big[(1-\tau_k)\frac{y_t^k(\mathbf{x}^t)}{Y^k} - (1-\bar{\tau}_l')\frac{(1-\tau_l'(y_t^l(\mathbf{x}^t)))y_t^l(\mathbf{x}^t)}{(1-\bar{\tau}_l')Y^l} + \tau_k - \bar{\tau}_l'\Big]dP_t \\ +\frac{\frac{Y^l}{Y^k}}{1-\tau_k}\Big[\mathrm{E}_\Gamma[\tau_l']\bar{\epsilon}_{L,1-\tau_k} + \mathrm{Cov}_\Gamma\Big(\tau_l', \frac{y^l}{Y^l}\bar{\epsilon}_{l,1-\tau_k}\Big)\Big] = 0.\end{aligned}$$

Using the definitions of $\hat{\epsilon}_{K,1-\tau_k}$, $\bar{G}^k$ $\bar{\mathcal{G}}^k$ and $\tilde{\mathcal{G}}^l$ this is equivalent to

$$\begin{aligned}\beta^{t^a}(\bar{G}^k - 1) + \frac{\tau_k}{1-\tau_k}\hat{\epsilon}_{K,1-\tau_k} - \bar{\epsilon}_{w,1-\tau_k}\frac{\frac{Y^l}{Y^k}}{1-\tau_k}\Big[(1-\tau_k)\bar{\mathcal{G}}^k - (1-\bar{\tau}_l')\tilde{\mathcal{G}}^l + \tau_k - \bar{\tau}_l'\Big] \\ +\frac{\frac{Y^l}{Y^k}}{1-\tau_k}\Big[\mathrm{E}_\Gamma[\tau_l']\bar{\epsilon}_{L,1-\tau_k} + \mathrm{Cov}_\Gamma\Big(\tau_l', \frac{y^l}{Y^l}\bar{\epsilon}_{l,1-\tau_k}\Big)\Big] = 0.\end{aligned}$$

Rearranging terms then gives the desired expression.

Finally, $\tau_k < 1$ ensure that the solution is interior. Hence equation (B.2) is indeed a necessary condition for τ_k to be optimal. □

A.2.3 Proof of Lemma C.1

Proof. The intra-temporal optimality condition is given by

$$\left(1 - \tau_l'(\eta w_t l_t(k_0, \eta))\right)\eta w_t u_{c,t} = -u_{l,t},$$

where I denote

$$u_{c,t} \equiv u_c(c_t(k_0, \eta), l_t(k_0, \eta)) \quad \text{and} \quad u_{l,t} \equiv u_l(c_t(k_0, \eta), l_t(k_0, \eta)).$$

Deriving with respect to $x \in \{1 - \tau_k, \bar{r}_s, T\}$ or with respect to $x \in \{w_s\}_{s \neq t}$ and evaluating at the steady state gives

$$-\tau_l''(\eta w l(k_0, \eta))(\eta w)^2 \frac{\partial l_t(k_0, \eta)}{\partial x} u_c + \left(1 - \tau_l'(\eta w l(k_0, \eta))\right)\eta w \left[u_{cc} \frac{\partial c_t(k_0, \eta)}{\partial x} + u_{cl} \frac{\partial l_t(k_0, \eta)}{\partial x}\right] = -u_{cl} \frac{\partial c_t(k_0, \eta)}{\partial x} - u_{ll} \frac{\partial l_t(k_0, \eta)}{\partial x},$$

which is equivalent to

$$\frac{\partial c_t(k_0, \eta)}{\partial x}\left[\left(1 - \tau_l'(\eta w l(k_0, \eta))\right)\eta w u_{cc} + u_{cl}\right] = \frac{\partial l_t(k_0, \eta)}{\partial x}\left[\tau_l''(\eta w l(k_0, \eta))(\eta w)^2 u_c - \left(1 - \tau_l'(\eta w l(k_0, \eta))\right)\eta w u_{cl} - u_{ll}\right]$$

Dividing by $\left(1 - \tau_l'(\eta w l(k_0, \eta))\right)\eta w$ gives

$$\frac{\partial c_t(k_0, \eta)}{\partial x}\left[u_{cc} + \frac{u_{cl}}{\left(1 - \tau_l'(\eta w l(k_0, \eta))\right)\eta w}\right] = \frac{\partial l_t(k_0, \eta)}{\partial x}\left[\frac{\tau_l''(\eta w l(k_0, \eta))\eta w u_c}{1 - \tau_l'(\eta w l(k_0, \eta))} - u_{cl} - \frac{u_{ll}}{\left(1 - \tau_l'(\eta w l(k_0, \eta))\right)\eta w}\right].$$

Plugging in the intra-temporal first order condition and the definition of the local rate of tax progressivity gives

$$\frac{\partial c_t(k_0, \eta)}{\partial x}\left[u_{cc} - \frac{u_{cl} u_c}{u_l}\right] = \frac{\partial l_t(k_0, \eta)}{\partial x}\left[\frac{p(\eta w l(k_0, \eta)) u_c}{l(k_0, \eta)} - u_{cl} + \frac{u_{ll} u_c}{u_l}\right].$$

Dividing by u_c then gives

$$\frac{\partial c_t(k_0, \eta)}{\partial x}\left[\frac{u_{cc}}{u_c} - \frac{u_{cl}}{u_l}\right] = \frac{\partial l_t(k_0, \eta)}{\partial x}\left[\frac{p(\eta w l(k_0, \eta))}{l(k_0, \eta)} - \frac{u_{cl}}{u_c} + \frac{u_{ll}}{u_l}\right].$$

Now observe that Assumption 4 implies

$$\frac{u_{cc}}{u_c} - \frac{u_{cl}}{u_l} = 0$$

and therefore

$$\frac{\partial l_t(k_0, \eta)}{\partial x} = 0,$$

which proves the Lemma for $x \in \{1 - \tau_k, \bar{r}_s, T_s\}$ and for $x \in \{w_s\}_{s \neq t}$.

Deriving the intra-temporal optimality condition with respect to w_t and evaluating at the steady state

$$\begin{aligned}
&(1 - \tau_l'(\eta w l(k_0, \eta)))\eta \left[u_c + w\left(u_{cc}\frac{\partial c_t(k_0, \eta)}{\partial w_t} + u_{cl}\frac{\partial l_t(k_0, \eta)}{\partial w_t}\right)\right] \\
&-\tau_l''(\eta w l(k_0, \eta))\eta w u_c \eta \left[l(k_0, \eta) + w\frac{\partial l_t(k_0, \eta)}{\partial w_t}\right] = -u_{cl}\frac{\partial c_t(k_0, \eta)}{\partial w_t} - u_{ll}\frac{\partial l_t(k_0, \eta)}{\partial w_t},
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\frac{\partial c_t(k_0, \eta)}{\partial w_t}\left[(1 - \tau_l'(\eta w l(k_0, \eta)))\eta w u_{cc} + u_{cl}\right] + (1 - \tau_l'(\eta w l(k_0, \eta)))\eta u_c \\
&\qquad -\tau_l''(\eta w l(k_0, \eta))\eta w u_c \eta l(k_0, \eta) = \\
&\frac{\partial l_t(k_0, \eta)}{\partial w_t}\left[-(1 - \tau_l'(\eta w l(k_0, \eta)))\eta w u_{cl} + \tau_l''(\eta w l(k_0, \eta))(\eta w)^2 u_c - u_{ll}\right]
\end{aligned}$$

Dividing by $(1 - \tau_l'(\eta w l(k_0, \eta)))\eta w$ gives

$$\begin{aligned}
&\frac{\partial c_t(k_0, \eta)}{\partial w_t}\left[u_{cc} + \frac{u_{cl}}{(1 - \tau_l'(\eta w l(k_0, \eta)))\eta w}\right] + \frac{u_c}{w} - \frac{\tau_l''(\eta w l(k_0, \eta))\eta w u_c \eta l(k_0, \eta)}{(1 - \tau_l'(\eta w l(k_0, \eta)))\eta w} = \\
&\frac{\partial l_t(k_0, \eta)}{\partial w_t}\left[-u_{cl} + \frac{\tau_l''(\eta w l(k_0, \eta))(\eta w)^2 u_c}{(1 - \tau_l'(\eta w l(k_0, \eta)))\eta w} - \frac{u_{ll}}{(1 - \tau_l'(\eta w l(k_0, \eta)))\eta w}\right]
\end{aligned}$$

Plugging in the intra-temporal first order condition and the definition of the local rate of tax progressivity gives

$$\frac{\partial c_t(k_0, \eta)}{\partial w_t}\left[u_{cc} - \frac{u_{cl}u_c}{u_l}\right] + \frac{u_c}{w}(1 - p(\eta w l(k_0, \eta))) = \frac{\partial l_t(k_0, \eta)}{\partial w_t}\left[-u_{cl} + \frac{p(\eta w l(k_0, \eta))u_c}{l(k_0, \eta)} + \frac{u_{ll}u_c}{u_l}\right].$$

Dividing by u_c then gives

$$\frac{\partial c_t(k_0, \eta)}{\partial w_t}\left[\frac{u_{cc}}{u_c} - \frac{u_{cl}}{u_l}\right] + \frac{(1 - p(\eta w l(k_0, \eta)))}{w} = \frac{\partial l_t(k_0, \eta)}{\partial w_t}\left[-\frac{u_{cl}}{u_c} + \frac{p(\eta w l(k_0, \eta))}{l(k_0, \eta)} + \frac{u_{ll}}{u_l}\right].$$

Again using the fact that Assumption 4 implies

$$\frac{u_{cc}}{u_c} - \frac{u_{cl}}{u_l} = 0$$

gives

$$\frac{(1 - p(\eta w l(k_0, \eta)))}{w} = \frac{\partial l_t(k_0, \eta)}{\partial w_t}\left[\frac{p(\eta w l(k_0, \eta))}{l(k_0, \eta)} + \frac{u_{ll}}{u_l} - \frac{u_{cl}}{u_c}\right].$$

Furthermore, one can check that Assumption 4 also implies

$$\frac{u_{ll}}{u_l} - \frac{u_{cl}}{u_c} = \frac{v''(l(k_0, \eta))}{v'(l(k_0, \eta))}.$$

Hence,

$$\frac{\partial l_t(k_0, \eta)}{\partial w_t} = \frac{l(k_0, \eta)(1 - p(\eta w l(k_0, \eta)))}{w\left[p(\eta w l(k_0, \eta)) + l(k_0, \eta)\frac{v''(l(k_0, \eta))}{v'(l(k_0, \eta))}\right]}$$

which is equivalent to

$$\frac{\partial l_t(k_0, \eta)}{\partial w_t} = \frac{l(k_0, \eta)(1 - p(\eta w l(k_0, \eta)))}{w\left[p(\eta w l(k_0, \eta)) + \frac{1}{\gamma_l(k_0, \eta)}\right]}.$$

In terms of elasticities this is the same as

$$\tilde{\epsilon}_{l_t(k_0, \eta), w_t} = \frac{\gamma_l(k_0, \eta)(1 - p(\eta w l(k_0, \eta)))}{1 + \gamma_l(k_0, \eta) p(\eta w l(k_0, \eta))}.$$

which proofs the final part of the lemma. □

A.2.4 Proof of Lemma C.2

Proof. The transfer in period s is given by

$$T_s = 1_{t \geq t^a} \tau_k r_s K_s + \int \tau_l(\eta w_s l_s(k_0, \eta)) d\Gamma - G$$

Deriving with respect to $1 - \tau_k^r$ gives

$$\begin{aligned}\frac{dT_s}{d(1 - \tau_k^r)} =& - 1_{t \geq t^a} r_s K_s + \tau_k \left[r_s \frac{dK_s}{d(1 - \tau_k^r)} + \frac{dr_s}{d(1 - \tau_k)} K_s\right] \\ &+ \int \tau_l'(y^l(k_0, \eta))\eta \left[w_s \frac{dl_s(k_0, \eta)}{d(1 - \tau_k)} + \frac{dw_s}{d(1 - \tau_k)} l_s(k_0, \eta)\right] d\Gamma,\end{aligned}$$

which can be written in terms of elasticities as

$$\begin{aligned}\epsilon_{T_s,1-\tau_k} =& -1_{t\geq t^a}(1-\tau_k)\frac{Y_s^k}{T_s} + \tau_k\left[\frac{Y_s^k}{T_s}\epsilon_{K_s,1-\tau_k} + \frac{Y_s^k}{T_s}\epsilon_{r_s,1-\tau_k}\right] \\ &+ \frac{1}{T_s}\int \tau_l'(y^l(k_0,\eta))\eta\left[w_s l_s(k_0,\eta)\epsilon_{l_s(k_0,\eta),1-\tau_k} + \epsilon_{w_s,1-\tau_k} w_s l_s(k_0,\eta)\right]d\Gamma\end{aligned}$$

Note that by the same arguments as in the proof of Lemma C.1 we have that $\epsilon_{l_s(k_0,\eta),1-\tau_k} = \tilde{\epsilon}_{l_s(k_0,\eta),w_s}\epsilon_{w_s,1-\tau_k}$. Plugging in this result gives

$$\begin{aligned}\epsilon_{T_s,1-\tau_k} =& -1_{t\geq t^a}(1-\tau_k)\frac{Y_s^k}{T_s} + \tau_k\left[\frac{Y_s^k}{T_s}\epsilon_{K_s,1-\tau_k} + \frac{Y_s^k}{T_s}\epsilon_{r_s,1-\tau_k}\right] \\ &+ \frac{Y_s^l}{T_s}\epsilon_{w_s,1-\tau_k}\left[\int \tau_l'(y^l(k_0,\eta))\frac{y_s^l(k_0,\eta)}{Y_s^l}\tilde{\epsilon}_{l_s(k_0,\eta),w_s}d\Gamma + \bar{\tau}_{l,s}'\right].\end{aligned}$$

Now note that

$$\int \tau_l'(y^l(k_0,\eta))\frac{y_s^l(k_0,\eta)}{Y_s^l}\tilde{\epsilon}_{l_s(k_0,\eta),w_s}d\Gamma = \mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}\right).$$

Using this and plugging in the expressions for $\epsilon_{r_s,1-\tau_k}$ and $\epsilon_{w_s,1-\tau_k}$ from Lemma 1 gives

$$\begin{aligned}\epsilon_{T_s,1-\tau_k} =& -1_{t\geq t^a}(1-\tau_k)\frac{Y^k}{T} + \tau_k\left[\frac{Y^k}{T}\epsilon_{K_s,1-\tau_k} + \frac{Y^k}{T}\frac{\tilde{\alpha}^l}{\sigma}\frac{r+\delta}{r}\left[\epsilon_{L_s,1-\tau_k} - \epsilon_{K_s,1-\tau_k}\right]\right] \\ &+ \frac{Y^l}{T}\frac{\tilde{\alpha}^k}{\sigma}\left[\epsilon_{K_s,1-\tau_k} - \epsilon_{L_s,1-\tau_k}\right]\left[\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}\right) + \bar{\tau}_l'\right],\end{aligned}$$

where Assumption 3 allowed me to drop some time indices.

Finally, plugging in the expression for $\epsilon_{L_t,1-\tau_k}$ from Corollary C.1 gives

$$\begin{aligned}\epsilon_{T_s,1-\tau_k} =& -1_{t\geq t^a}(1-\tau_k)\frac{Y^k}{T} + \tau_k\left[\frac{Y^k}{T}\epsilon_{K_s,1-\tau_k} - \frac{Y^k}{T}\frac{\tilde{\alpha}^l}{\sigma}\frac{r+\delta}{r}\epsilon_{K_s,1-\tau_k}\frac{\sigma}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\right] \\ &+ \frac{Y^l}{T}\frac{\tilde{\alpha}^k}{\sigma}\epsilon_{K_s,1-\tau_k}\frac{\sigma}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\left[\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}\right) + \bar{\tau}_l'\right],\end{aligned}$$

which is equivalent to

$$\begin{aligned}\epsilon_{T_s,1-\tau_k} =& \frac{Y^k}{T}\left[-1_{t\geq t^a}(1-\tau_k) + \tau_k\epsilon_{K_s,1-\tau_k}\right] \\ &+ \frac{Y}{T}\frac{\epsilon_{K_s,1-\tau_k}}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\left[\alpha^l\tilde{\alpha}^k\left(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}\right) + \bar{\tau}_l'\right) - \alpha^k\tilde{\alpha}^l\frac{r+\delta}{r}\tau_k\right].\end{aligned}$$

Finally, observe that

$$\alpha^k\tilde{\alpha}^l\frac{r+\delta}{r} = \alpha^k\tilde{\alpha}^l\frac{\frac{\tilde{\alpha}^k\tilde{Y}}{K}}{\frac{\alpha^k Y}{K}} = \alpha^k\tilde{\alpha}^l\frac{\tilde{\alpha}^k\tilde{Y}}{\alpha^k Y} = \tilde{\alpha}^k\frac{\tilde{\alpha}^l\tilde{Y}}{Y} = \tilde{\alpha}^k\frac{wL}{Y} = \tilde{\alpha}^k\alpha^l.$$

Plugging in above hence gives

$$\begin{aligned}\epsilon_{T_s,1-\tau_k} =& \frac{Y^k}{T}\Big[-1_{t\geq t^a}(1-\tau_k)+\tau_k\epsilon_{K_s,1-\tau_k}\Big]\\ &+\frac{Y}{T}\frac{\epsilon_{K_s,1-\tau_k}}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\alpha^l\bigg[\Big(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\big(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\big)+\bar{\tau}_l'\Big)-\tau_k\bigg].\end{aligned}$$

□

A.2.5 Proof of Lemma C.3

Proof. For simplicity, I denote the net return to capital in period t by

$$\bar{r}_t = (1-\tau_k)r_t.$$

In the proof, I extensively make use of the fact that at the steady state, we have $\beta(1+\bar{r}_t)=1$ for all t.

The Euler equation is given by

$$u_{c,t}(k_0,\eta) = \beta(1+\bar{r}_t)u_{c,t+1}(k_0,\eta).$$

Hence, we can write the Euler equation as

$$u_{c,0}(k_0,\eta) = \big[\beta(1+\bar{r})\big]^t u_{c,t}(k_0,\eta).$$

Partially deriving with respect to $(1-\tau_k)$ gives

$$\begin{aligned} u_{cc,0}(k_0,\eta)\frac{\partial c_0(k_0,\eta)}{\partial(1-\tau_k)}+u_{cl,0}(k_0,\eta)\frac{\partial l_0(k_0,\eta)}{\partial(1-\tau_k)} =&\\ \big[\beta(1+\bar{r})\big]^t\bigg[u_{cc,t}(k_0,\eta)\frac{\partial c_t(k_0,\eta)}{\partial(1-\tau_k)}+u_{cl,t}(k_0,\eta)\frac{\partial l_t(k_0,\eta)}{\partial(1-\tau_k)}\bigg]&+t\big[\beta(1+\bar{r})\big]^{t-1}\beta r u_{c,t}(k_0,\eta),\end{aligned}$$

which, evaluated at the initial steady state, is equivalent to

$$\begin{aligned} u_{cc}(k_0,\eta)\frac{\partial c_0(k_0,\eta)}{\partial(1-\tau_k)}+u_{cl}(k_0,\eta)\frac{\partial l_0(k_0,\eta)}{\partial(1-\tau_k)} =&\\ \bigg[u_{cc}(k_0,\eta)\frac{\partial c_t(k_0,\eta)}{\partial(1-\tau_k)}+u_{cl}(k_0,\eta)\frac{\partial l_t(k_0,\eta)}{\partial(1-\tau_k)}\bigg]&+t\beta r u_c(k_0,\eta),\end{aligned}$$

which is the same as

$$u_{cc}(k_0,\eta)\left[\frac{\partial c_t(k_0,\eta)}{\partial(1-\tau_k)}-\frac{\partial c_0(k_0,\eta)}{\partial(1-\tau_k)}\right]+u_{cl}(k_0,\eta)\left[\frac{\partial l_t(k_0,\eta)}{\partial(1-\tau_k)}-\frac{\partial l_0(k_0,\eta)}{\partial(1-\tau_k)}\right]=-t\beta r u_c(k_0,\eta).$$

Using Lemma C.1 this is equivalent to

$$\frac{\partial c_t(k_0,\eta)}{\partial(1-\tau_k)}=\frac{\partial c_0(k_0,\eta)}{\partial(1-\tau_k)}-t\beta r\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}.$$

This can be rewritten in terms of elasticities as

$$\tilde{\epsilon}_{c_t(k_0,\eta),1-\tau_k}=\tilde{\epsilon}_{c_0(k_0,\eta),1-\tau_k}+t(1-\beta)\frac{u_c(k_0,\eta)}{c(k_0,\eta)u_{cc}(k_0,\eta)}. \tag{A.1}$$

Deriving the intertemporal budget constraint

$$\sum_{t=0}^{\infty}\frac{c_t(k_0,\eta)}{(1+\bar{r})^t}=(1+\bar{r})k_0+\sum_{t=0}^{\infty}\frac{\eta w l_t(k_0,\eta)-\tau_l(\eta w l_t(k_0,\eta))+T}{(1+\bar{r})^t}$$

with respect to $(1-\tau_k)$ gives

$$\sum_{t=0}^{\infty}\left[\frac{\partial c_t(k_0,\eta)}{\partial(1-\tau_k)}\frac{1}{(1+\bar{r})^t}-t(1+\bar{r})^{-t-1}rc(k_0,\eta)\right]=rk_0$$
$$+\sum_{t=0}^{\infty}\frac{\partial l_t(k_0,\eta)}{\partial(1-\tau_k)}\frac{\eta w\big(1-\tau_l'(\eta w l(k_0,\eta))\big)}{(1+\bar{r})^t}-[\eta w l(k_0,\eta)-\tau_l(\eta w l(k_0,\eta))+T]\sum_{t=0}^{\infty}t(1+\bar{r})^{-t-1}r,$$

which evaluated at the steady state equals

$$\sum_{t=0}^{\infty}\frac{\partial c_t(k_0,\eta)}{\partial(1-\tau_k)}\frac{1}{(1+\bar{r})^t}=rk_0+\sum_{t=0}^{\infty}\frac{\partial l_t(k_0,\eta)}{\partial(1-\tau_k)}\frac{\eta w\big(1-\tau_l'(\eta w l(k_0,\eta))\big)}{(1+\bar{r})^t}+\bar{r}k_0\sum_{t=0}^{\infty}t(1+\bar{r})^{-t-1}r.$$

Using the formulas for geometric and arithmetico-geometric sequences, and plugging in the result above gives

$$\frac{\partial c_0(k_0,\eta)}{\partial(1-\tau_k)}=rk_0+\frac{\beta^2 r}{1-\beta}\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}$$

The budget constraint is given by

$$c_t(k_0,\eta)=(1+\bar{r})k_t(k_0,\eta)-k_{t+1}(k_0,\eta)+\big(1-\tau_l(\eta w l_t(k_0,\eta))\big)\eta w l_t(k_0,\eta)+T$$

Deriving this constraint for $t=0$ gives

$$\frac{\partial k_1(k_0,\eta)}{\partial(1-\tau_k)}=rk_0-\frac{\partial c_0(k_0,\eta)}{\partial(1-\tau_k)}+\eta w\big(1-\tau_l'(\eta w l(k_0,\eta))\big)\frac{\partial l_0(k_0,\eta)}{\partial(1-\tau_k)}$$

$$
\begin{aligned}
&= rk_0 - \frac{\partial c_0(k_0,\eta)}{\partial(1-\tau_k)} \\
&= -\frac{\beta^2 r}{1-\beta}\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)},
\end{aligned}
$$

where the second equality follows from Lemma C.1.

Expressed in elasticities this is equivalent to

$$
\tilde{\epsilon}_{k_1(k_0,\eta),1-\tau_k} = -\frac{\beta}{k_0}\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}.
$$

This proves the Lemma for $t = 1$.

Consider now an arbitrary $t \geq 1$. Assume the condition holds for an arbitrary $t \geq 1$. Deriving the budget constraint for any $t \geq 1$ and evaluating at the steady state gives

$$
\begin{aligned}
\frac{\partial k_{t+1}(k_0,\eta)}{\partial(1-\tau_k)} &= rk_0 + (1+\bar{r})\frac{\partial k_t(k_0,\eta)}{\partial(1-\tau_k)} - \frac{\partial c_t(k_0,\eta)}{\partial(1-\tau_k)} + \eta w\big(1-\tau_l'(\eta w l(k_0,\eta))\big)\frac{\partial l_t(k_0,\eta)}{\partial(1-\tau_k)} \\
&= rk_0 - (1+\bar{r})t\frac{\beta^2 r}{1-\beta}\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} - \frac{\partial c_t(k_0,\eta)}{\partial(1-\tau_k)} \\
&= rk_0 - (1+\bar{r})t\frac{\beta^2 r}{1-\beta}\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} - \left[\frac{\partial c_0(k_0,\eta)}{\partial(1-\tau_k)} - t\beta r\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}\right] \\
&= -t\frac{\beta}{1-\beta}r\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} - \left[\frac{\beta^2 r}{1-\beta} - t\beta r\right]\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} \\
&= -\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}\frac{t\beta r + \beta^2 r - t\beta r + t\beta^2 r}{1-\beta} \\
&= -(t+1)\frac{\beta^2 r}{1-\beta}\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}.
\end{aligned}
$$

Expressed in terms of elasticities this is the same as

$$
\tilde{\epsilon}_{k_{t+1}(k_0,\eta),1-\tau_k} = -(t+1)\frac{\beta}{k_0}\frac{u_c(k_0,\eta)}{u_{cc}(k_0,\eta)},
$$

which completes the proof. □

A.2.6 Proof of Lemma C.4

Proof. Assuming that $F_{kl} = 0$, the policy elasticity is given by

$$
\epsilon_{K_t,1-\tau_k} = \tilde{\epsilon}_{K_t,1-\tau_k} + \sum_{s=0}^{\infty}\tilde{\epsilon}_{K_t,T_s}\epsilon^{ex}_{T_s,1-\tau_k},
$$

where

$$
\epsilon^{ex}_{T_s,1-\tau_k} = \frac{Y^k}{T}\Big[-(1-\tau_k) + \tau_k\epsilon_{K_s,1-\tau_k}\Big]
$$

by Lemma C.2. Using Lemma A.3 gives

$$\begin{aligned}
\epsilon_{K_t,1-\tau_k} =& \tilde{\epsilon}_{K_t,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s} \frac{Y^k}{T} \Big[-(1-\tau_k) + \tau_k \epsilon_{K_s,1-\tau_k} \Big] \\
=& \tilde{\epsilon}_{K_t,1-\tau_k} + \sum_{s=0}^{t-1} \frac{T}{K} \beta^{s+1} \frac{Y^k}{T} \Big[-(1-\tau_k) + \tau_k \epsilon_{K_s,1-\tau_k} \Big] \\
& - \sum_{s=t}^{\infty} \frac{T}{K} \beta^{s+1} (\beta^{-t} - 1) \frac{Y^k}{T} \Big[-(1-\tau_k) + \tau_k \epsilon_{K_s,1-\tau_k} \Big].
\end{aligned}$$

Therefore,

$$\epsilon_{K_t,1-\tau_k} = \tilde{\epsilon}_{K_t,1-\tau_k} + \tau_k r \Big[\sum_{s=0}^{\infty} \beta^{s+1} \epsilon_{K_s,1-\tau_k} - \beta^{-t} \sum_{s=t}^{\infty} \beta^{s+1} \epsilon_{K_s,1-\tau_k} \Big].$$

I will now find a sequence $\{\epsilon_{K_t,1-\tau_k}\}_{t=1}^{\infty}$ that satisfies the above equation for all t using the guess

$$\epsilon_{K_t,1-\tau} = \lambda \tilde{\epsilon}_{K_t,1-\tau_k}$$

for some $\lambda > 0$. We know from Lemma C.3 that $\tilde{\epsilon}_{K_t,1-\tau_k} = t \tilde{\epsilon}_{K_1,1-\tau_k}$. Hence, this guess implies

$$\epsilon_{K_t,1-\tau} = t \lambda \tilde{\epsilon}_{K_1,1-\tau_k}.$$

Plugging in the above equation gives

$$\begin{aligned}
t \lambda \tilde{\epsilon}_{K_1,1-\tau_k} =& t \tilde{\epsilon}_{K_1,1-\tau_k} + \tau_k r \lambda \tilde{\epsilon}_{K_1,1-\tau_k} \Big[\sum_{s=0}^{\infty} s \beta^{s+1} - \beta^{-t} \sum_{s=t}^{\infty} s \beta^{s+1} \Big] \\
=& t \tilde{\epsilon}_{K_1,1-\tau_k} + \tau_k r \beta \lambda \tilde{\epsilon}_{K_1,1-\tau_k} \Big[\sum_{s=0}^{\infty} s \beta^{s} - \sum_{s=0}^{\infty} s \beta^{s} - t \sum_{s=0}^{\infty} \beta^{s} \Big] \\
=& t \tilde{\epsilon}_{K_1,1-\tau_k} - t \tau_k r \lambda \tilde{\epsilon}_{K_1,1-\tau_k} \frac{\beta}{1-\beta} \\
=& t \left(1 - \tau_k r \lambda \frac{\beta}{1-\beta} \right) \tilde{\epsilon}_{K_1,1-\tau_k}.
\end{aligned}$$

This equation is satisfied for all t if and only if

$$\lambda = 1 - \tau_k r \lambda \frac{\beta}{1-\beta},$$

implying

$$\lambda = \frac{1-\beta}{1-\beta(1-\tau_k r)} < 1.$$

Hence, the policy elasticity with constant prices is given by

$$\epsilon_{K_t,1-\tau_k} = t\frac{1-\beta}{1-\beta(1-\tau_k r)}\tilde{\epsilon}_{K_1,1-\tau_k}.$$

□

A.2.7 Proof of Proposition C.1

Proof. Consider an agent with initial wealth k_0 and labor productivity η. The total derivative of her capital in period t with respect to $1-\tau_k$ can be written as

$$\begin{aligned}\frac{dk_t(k_0,\eta)}{d(1-\tau_k)} =& \frac{\partial k_t(k_0,\eta)}{\partial(1-\tau_k)} + \sum_{s=0}^{\infty}\frac{\partial k_t(k_0,\eta)}{\partial r_s}\frac{dr_s}{d(1-\tau_k)} \\ &+ \sum_{s=0}^{\infty}\frac{\partial k_t(k_0,\eta)}{\partial T_s}\frac{dT_s}{d(1-\tau_k)} + \sum_{s=0}^{\infty}\frac{\partial k_t(k_0,\eta)}{\partial w_s}\frac{dw_s}{d(1-\tau_k)}.\end{aligned}$$

Defining the net return to capital in period t as

$$\bar{r}_t = (1-\tau_k)r_t$$

the above equation is equivalent to

$$\begin{aligned}\frac{dk_t(k_0,\eta)}{d(1-\tau_k)} =& \sum_{s=0}^{\infty}\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_s}\left[r + (1-\tau_k)\frac{dr_s}{d(1-\tau_k)}\right] \\ &+ \sum_{s=0}^{\infty}\frac{\partial k_t(k_0,\eta)}{\partial T_s}\left[\frac{dT_s}{d(1-\tau_k)} + \eta\left(1-\tau_l'(\eta w l(k_0,\eta))\right)l(k_0,\eta)\frac{dw_s}{d(1-\tau_k)}\right],\end{aligned}$$

that is one can express the capital change solely in terms of a price effects $\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_s}$ and an income effects $\frac{\partial k_t(k_0,\eta)}{\partial T_s}$.

In terms of elasticities, this can be rewritten as

$$\begin{aligned}\epsilon_{k_t(k_0,\eta),1-\tau_k} =& \sum_{s=0}^{\infty}\tilde{\epsilon}_{k_t(k_0,\eta),\bar{r}_s}\left[1+\epsilon_{r_s,1-\tau_k}\right] \\ &+ \sum_{s=0}^{\infty}\tilde{\epsilon}_{k_t(k_0,\eta),T_s}\left[\epsilon_{T_s,1-\tau_k} + \frac{\left(1-\tau_l'(y^l(k_0,\eta))\right)y^l(k_0,\eta)}{T}\epsilon_{w_s,1-\tau_k}\right]\end{aligned}$$

Lemma A.3 in Appendix A.3 shows that

$$\tilde{\epsilon}_{k_t(k_0,\eta),T_s} = \frac{K}{k_0}\tilde{\epsilon}_{K_t,T_s},$$

that is the individual savings elasticities with respect to unearned income are inversely proportional to wealth.

Lemma A.3 in Appendix A.3 shows that

$$\tilde{\epsilon}_{k_t(k_0,\eta),T_s} = \frac{K}{k_0}\tilde{\epsilon}_{K_t,T_s},$$

that is the individual savings elasticities with respect to unearned income are inversely proportional to wealth. The aggregate elasticity

$$\epsilon_{K_t,1-\tau_k} = \int \frac{k_0}{K}\epsilon_{k_t(k_0,\eta),1-\tau_k} d\Gamma$$

is therefore given by

$$\epsilon_{K_t,1-\tau_k} = \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,\bar{r}_s}\left[1+\epsilon_{r_s,1-\tau_k}\right] + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s}\left[\epsilon_{T_s,1-\tau_k} + \frac{1}{T}(1-\bar{\tau}_l')Y^l\epsilon_{w_s,1-\tau_k}\right]. \tag{A.2}$$

Also given by Lemma A.3 is the elasticity of aggregate capital supply in period t with respect to unearned income in period s,

$$\tilde{\epsilon}_{K_t,T_s} = \frac{T}{K}\frac{\partial K_t}{\partial T_s} = \begin{cases} -\frac{T}{K}\beta^{s+1}(\beta^{-t}-1) & \text{if } 1 \le t \le s \\ \frac{T}{K}\beta^{s+1} & \text{if } t > s \ge 0. \end{cases}$$

Lemma A.4 shows that the elasticity of aggregate capital supply in period t with respect to the interest rate in period s is given by

$$\tilde{\epsilon}_{K_t,\bar{r}_s} = \begin{cases} \beta^s(\beta^{-t}-1)\left[\tilde{\epsilon}_{K_1,1-\tau_k} - (1-\beta)\right] & \text{if } 1 \le t \le s \\ \beta^s(1-\beta) + (1-\beta^s)\tilde{\epsilon}_{K_1,1-\tau_k} & \text{if } t > s \ge 0. \end{cases}$$

Observe that this is the same as

$$\tilde{\epsilon}_{K_t,\bar{r}_s} = \begin{cases} \frac{(1-\tau_k)Y^k - \frac{K\tilde{\epsilon}_{K_1,1-\tau_k}}{\beta}}{T}\tilde{\epsilon}_{K_t,T_s} & \text{if } 1 \le t \le s \\ \frac{(1-\tau_k)Y^k - \frac{K\tilde{\epsilon}_{K_1,1-\tau_k}}{\beta}}{T}\tilde{\epsilon}_{K_t,T_s} + \tilde{\epsilon}_{K_1,1-\tau_k} & \text{if } t > s \ge 0. \end{cases}$$

Hence we can write (A.2) as

$$\begin{aligned} \epsilon_{K_t,1-\tau_k} =& \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,\bar{r}_s}\left[1+\epsilon_{r_s,1-\tau_k}\right] + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s}\left[\epsilon_{T_s,1-\tau_k} + \frac{1}{T}(1-\bar{\tau}_l')Y^l\epsilon_{w_s,1-\tau_k}\right] \\ =& \tilde{\epsilon}_{K_1,1-\tau_k}\sum_{s=0}^{t-1}\left[1+\epsilon_{r_s,1-\tau_k}\right] \\ &+ \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s}\left[\epsilon_{T_s,1-\tau_k} + \frac{(1-\tau_k)Y^k - \frac{K\tilde{\epsilon}_{K_1,1-\tau_k}}{\beta}}{T}\left[1+\epsilon_{r_s,1-\tau_k}\right] + \frac{(1-\bar{\tau}_l')Y^l}{T}\epsilon_{w_s,1-\tau_k}\right] \end{aligned}$$

Furthermore, from Lemma 1 we know that

$$\epsilon_{r_s,1-\tau_k} = -\frac{\tilde{\alpha}^k}{\sigma}\frac{\alpha^l}{\alpha^k}\left[\epsilon_{K_s,1-\tau_k} - \epsilon_{L_s,1-\tau_k}\right] \quad \text{and} \quad \epsilon_{w_s,1-\tau_k} = \frac{\tilde{\alpha}^k}{\sigma}\left[\epsilon_{K_s,1-\tau_k} - \epsilon_{L_s,1-\tau_k}\right],$$

which by Corollary C.1 is equivalent to

$$\epsilon_{r_s,1-\tau_k} = -\frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k} \quad \text{and} \quad \epsilon_{w_s,1-\tau_k} = \frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\epsilon_{K_s,1-\tau_k}.$$

Finally, from Lemma C.2 we know

$$\begin{aligned}
\epsilon_{T_s,1-\tau_k} =& \frac{Y^k}{T}\left[-(1-\tau_k) + \tau_k\epsilon_{K_s,1-\tau_k}\right] \\
&+ \frac{Y}{T}\frac{\epsilon_{K_s,1-\tau_k}}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\alpha^l\left[\left(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}\right) + \bar{\tau}_l'\right) - \tau_k\right].
\end{aligned}$$

Plugging all these results into the equation above gives

$$\begin{aligned}
\epsilon_{K_t,1-\tau_k} =& \tilde{\epsilon}_{K_1,1-\tau_k}\sum_{s=0}^{t-1}\left(1 - \frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\right) \\
&+ \sum_{s=0}^{t-1}\frac{T}{K}\beta^{s+1}\Bigg[-\frac{\bar{r}K}{T} + \frac{Y}{T}\alpha^k\tau_k\epsilon_{K_s,1-\tau_k} \\
&+ \frac{Y}{T}\frac{\epsilon_{K_s,1-\tau_k}}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\alpha^l\left[\left(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}\right) + \bar{\tau}_l'\right) - \tau_k\right] \\
&+ \frac{(1-\tau_k)Y^k - \frac{K\tilde{\epsilon}_{K_1,1-\tau_k}}{\beta}}{T}\left(1 - \frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\right) + \frac{(1-\bar{\tau}_l')Y^l}{T}\frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\epsilon_{K_s,1-\tau_k}\Bigg] \\
&+ \sum_{s=t}^{\infty}\left(-\frac{T}{K}\beta^{s+1}(\beta^{-t}-1)\right)\Bigg[-\frac{\bar{r}K}{T} + \frac{Y}{T}\alpha^k\tau_k\epsilon_{K_s,1-\tau_k} \\
&+ \frac{Y}{T}\frac{\epsilon_{K_s,1-\tau_k}}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\alpha^l\left[\left(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}\right) + \bar{\tau}_l'\right) - \tau_k\right] \\
&+ \frac{(1-\tau_k)Y^k - \frac{K\tilde{\epsilon}_{K_1,1-\tau_k}}{\beta}}{T}\left(1 - \frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\right) + \frac{(1-\bar{\tau}_l')Y^l}{T}\frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\epsilon_{K_s,1-\tau_k}\Bigg],
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
\epsilon_{K_t,1-\tau_k} =& \tilde{\epsilon}_{K_1,1-\tau_k}\Bigg[\sum_{s=0}^{t-1}\left(1 - \frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\right) - \sum_{s=0}^{\infty}\beta^s\left(1 - \frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\right) \\
&+ \beta^{-t}\sum_{s=t}^{\infty}\beta^s\left(1 - \frac{\tilde{\alpha}^k}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\right)\Bigg] \\
&+ \sum_{s=0}^{t-1}\frac{T}{K}\beta^{s+1}\Bigg[\frac{Y}{T}\alpha^k\tau_k\epsilon_{K_s,1-\tau_k}
\end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{Y}{T}\frac{\epsilon_{K_s,1-\tau_k}}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\alpha^l\Big(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\big(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\big)\Big)\Big] \\
&+\sum_{s=t}^{\infty}\Big(-\frac{T}{K}\beta^{s+1}(\beta^{-t}-1)\Big)\Big[\frac{Y}{T}\alpha^k\tau_k\epsilon_{K_s,1-\tau_k} \\
&\quad + \frac{Y}{T}\frac{\epsilon_{K_s,1-\tau_k}}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\alpha^l\Big(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\big(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\big)\Big)\Big],
\end{aligned}
$$

which in turn is equivalent to

$$
\begin{aligned}
\epsilon_{K_t,1-\tau_k} =& -\tilde{\epsilon}_{K_1,1-\tau_k}\sum_{s=0}^{\infty}\beta^s\Big(1-\frac{\tilde{\alpha}^k}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\Big) \\
&+\tilde{\epsilon}_{K_1,1-\tau_k}\sum_{s=0}^{t-1}\Big(1-\frac{\tilde{\alpha}^k}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\Big) \\
&+\tilde{\epsilon}_{K_1,1-\tau_k}\beta^{-t}\sum_{s=t}^{\infty}\beta^s\Big(1-\frac{\tilde{\alpha}^k}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\frac{\alpha^l}{\alpha^k}\epsilon_{K_s,1-\tau_k}\Big) \\
&+\sum_{s=0}^{t-1}\beta^{s+1}\Big[r\tau_k\epsilon_{K_s,1-\tau_k}+r\frac{\epsilon_{K_s,1-\tau_k}}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\frac{\alpha^l}{\alpha^k}\Big(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\big(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\big)\Big)\Big] \\
&-(\beta^{-t}-1)\sum_{s=t}^{\infty}\beta^{s+1}\Big[r\tau_k\epsilon_{K_s,1-\tau_k}+r\frac{\epsilon_{K_s,1-\tau_k}}{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\frac{\alpha^l}{\alpha^k}\Big(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\big(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\big)\Big)\Big]
\end{aligned}
$$

I will now guess – and verify – that

$$
\epsilon_{K_s,1-\tau_k} = (1-\lambda^s)\epsilon_{K_\infty,1-\tau_k} \tag{A.3}
$$

with

$$
\epsilon_{K_\infty,1-\tau_k} = \frac{\sigma+\tilde{\alpha}^k\tilde{\epsilon}_{L_1,w_1}}{\tilde{\alpha}^k}\frac{\alpha^k}{\alpha^l}
$$

and $\lambda \in (0,1)$.

Plugging this guess into the equation above gives

$$
\begin{aligned}
(1-\lambda^t)\epsilon_{K_\infty,1-\tau_k} =& \tilde{\epsilon}_{K_1,1-\tau_k}\Big[-\sum_{s=0}^{\infty}(\beta\lambda)^s+\sum_{s=0}^{t-1}\lambda^s+\beta^{-t}\sum_{s=t}^{\infty}(\beta\lambda)^s\Big] \\
&+r\Big[\tau_k\epsilon_{K_\infty,1-\tau_k}+\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\big(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\big)\Big]\times \\
&\times\Big[\sum_{s=0}^{t-1}\beta^{s+1}(1-\lambda^s)-(\beta^{-t}-1)\sum_{s=t}^{\infty}\beta^{s+1}(1-\lambda^s)\Big].
\end{aligned}
$$

Using the limits of the geometric series this is the same as

$$
(1-\lambda^t)\epsilon_{K_\infty,1-\tau_k} = \tilde{\epsilon}_{K_1,1-\tau_k}\Big[-\frac{1}{1-\beta\lambda}+\frac{1-\lambda^t}{1-\lambda}+\frac{\lambda^t}{1-\beta\lambda}\Big]
$$

$$+\beta\left[\left[\frac{1-\beta^t}{1-\beta}-\frac{1-(\beta\lambda)^t}{1-\beta\lambda}\right]-(\beta^{-t}-1)\left[\frac{\beta^t}{1-\beta}-\frac{(\beta\lambda)^t}{1-\beta\lambda}\right]\right]\times$$
$$r\left[\tau_k\epsilon_{K_\infty,1-\tau_k}+\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\left(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\right)\right],$$

which in turn is equivalent to

$$\begin{aligned}(1-\lambda^t)\epsilon_{K_\infty,1-\tau_k} =&\tilde{\epsilon}_{K_1,1-\tau_k}\left[\frac{\lambda^t-1}{1-\beta\lambda}+\frac{1-\lambda^t}{1-\lambda}\right]\\ &-\beta\frac{1-\lambda^t}{1-\beta\lambda}\times r\left[\tau_k\epsilon_{K_\infty,1-\tau_k}+\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\left(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\right)\right]\end{aligned}$$

Setting $t=1$ gives

$$(1-\lambda)\epsilon_{K_\infty,1-\tau_k} =\tilde{\epsilon}_{K_1,1-\tau_k}\frac{\lambda(1-\beta)}{1-\beta\lambda}-\beta\frac{1-\lambda}{1-\beta\lambda}\times r\left[\tau_k\epsilon_{K_\infty,1-\tau_k}+\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\left(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\right)\right]. \tag{A.4}$$

Multiplying by $1-\beta\lambda$ gives

$$\begin{aligned}(1-\lambda)(1-\beta\lambda)\epsilon_{K_\infty,1-\tau_k} =&\tilde{\epsilon}_{K_1,1-\tau_k}\lambda(1-\beta)\\ &-\beta(1-\lambda)r\left[\tau_k\epsilon_{K_\infty,1-\tau_k}+\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\left(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\right)\right],\end{aligned}$$

which is equivalent to the quadratic equation

$$\begin{aligned}&\lambda^2\underbrace{\beta\epsilon_{K_\infty,1-\tau_k}}_{\equiv a}\\ -\lambda&\underbrace{\left\{(1+\beta(1+\tau_k r))\epsilon_{K_\infty,1-\tau_k}+(1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k}+r\beta\left[\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\left(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\right)\right]\right\}}_{\equiv -b}\\ &+\underbrace{\epsilon_{K_\infty,1-\tau_k}(1+\beta\tau_k r)+r\beta\left[\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s}+\mathrm{Cov}_\Gamma\left(\tau_l',y^l\bar{\epsilon}_{l,1-\tau_k}\right)\right]}_{\equiv c}=0.\end{aligned}$$

There are two solutions to this quadratic equations. I will show that the root

$$\lambda=\frac{-b-\sqrt{b^2-4ac}}{2a}$$

has a value in $(0,1)$ and is decreasing in $\tilde{\epsilon}_{K_1,1-\tau_k}$.

First note that $\sqrt{b^2-4ac}$ is real since

$$b=-\left(a+c+(1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k}\right)$$

and hence

$$
\begin{aligned}
b^2 - 4ac =& (a+c)^2 + 2(a+c)(1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k} + (1-\beta)^2\tilde{\epsilon}^2_{K_1,1-\tau_k} - 4ac \\
=& (a-c)^2 + 2(a+c)(1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k} + (1-\beta)^2\tilde{\epsilon}^2_{K_1,1-\tau_k} > 0.
\end{aligned}
$$

Then it follows immediately from $b < 0$, $a > 0$ and $c > 0$ that λ must be positive.

To see that $\lambda < 1$ note its numerator satisfies

$$
\begin{aligned}
-b - \sqrt{b^2 - 4ac} =& a + c + (1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k} - \sqrt{(a-c)^2 + 2(a+c)(1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k} + (1-\beta)^2\tilde{\epsilon}^2_{K_1,1-\tau_k}} \\
<& a + c + (1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k} - \sqrt{(c-a)^2 + 2(c-a)(1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k} + (1-\beta)^2\tilde{\epsilon}^2_{K_1,1-\tau_k}} \\
=& a + c + (1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k} - \left(c - a + (1-\beta)\tilde{\epsilon}_{K_1,1-\tau_k}\right) \\
=& 2a.
\end{aligned}
$$

Thus, since the numerator is smaller than the denominator, we must have $\lambda < 1$.

To see that $\frac{d\lambda}{d\tilde{\epsilon}_{K_1,1-\tau_k}} < 0$ note that

$$
\frac{da}{d\tilde{\epsilon}_{K_1,1-\tau_k}} = \frac{dc}{d\tilde{\epsilon}_{K_1,1-\tau_k}} = 0
$$

and

$$
\frac{db}{d\tilde{\epsilon}_{K_1,1-\tau_k}} = -(1-\beta)
$$

Hence,

$$
\begin{aligned}
\frac{d\lambda}{d\tilde{\epsilon}_{K_1,1-\tau_k}} =& \frac{1}{2a}\left[1 - \beta + \frac{2b(1-\beta)}{2\sqrt{b^2 - 4ac}}\right] \\
=& \frac{1-\beta}{2a}\left[1 + \frac{b}{\sqrt{b^2 - 4ac}}\right] \\
<& 0,
\end{aligned}
$$

where the inequality follows from the fact that the absolute value of b is larger than the absolute value of $\sqrt{b^2 - 4ac}$ and hence the term in squared brackets is negative.

Finally, one can show by induction that condition (A.3) holds for all t. By construction, it holds for $t = 1$. What is left to show is that if it holds for an arbitrary $t \geq 1$, it must also hold for $t + 1$.

Assume that it holds for $t \geq 1$. If it holds also for $t + 1$, it must satisfy

$$
(1 - \lambda^{t+1})\epsilon_{K_\infty,1-\tau_k} = \tilde{\epsilon}_{K_1,1-\tau_k}\left[\frac{\lambda^{t+1} - 1}{1 - \beta\lambda} + \frac{1 - \lambda^{t+1}}{1 - \lambda}\right]
$$

$$- \beta \frac{1-\lambda^{t+1}}{1-\beta\lambda} \times r \left[\tau_k \epsilon_{K_\infty,1-\tau_k} + \mathrm{E}_\Gamma[\tau_l'] \tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l \bar{\epsilon}_{l,1-\tau_k}\right) \right]$$

This condition is equivalent to

$$\begin{aligned}(1-\lambda^t+\lambda^t-\lambda^{t+1})\epsilon_{K_\infty,1-\tau_k} =& \tilde{\epsilon}_{K_1,1-\tau_k} \left[\frac{\lambda^{t+1}-\lambda^t+\lambda^t-1}{1-\beta\lambda} + \frac{1-\lambda^t+\lambda^t-\lambda^{t+1}}{1-\lambda} \right] \\ & - \beta \frac{1-\lambda^t+\lambda^t-\lambda^{t+1}}{1-\beta\lambda} \times r \left[\tau_k \epsilon_{K_\infty,1-\tau_k} + \mathrm{E}_\Gamma[\tau_l'] \tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l \bar{\epsilon}_{l,1-\tau_k}\right) \right].\end{aligned}$$

Using that condition (A.3) hold for t gives

$$\begin{aligned}(\lambda^t-\lambda^{t+1})\epsilon_{K_\infty,1-\tau_k} =& \tilde{\epsilon}_{K_1,1-\tau_k} \left[\frac{\lambda^{t+1}-\lambda^t}{1-\beta\lambda} + \frac{\lambda^t-\lambda^{t+1}}{1-\lambda} \right] \\ & - \beta \frac{\lambda^t-\lambda^{t+1}}{1-\beta\lambda} \times r \left[\tau_k \epsilon_{K_\infty,1-\tau_k} + \mathrm{E}_\Gamma[\tau_l'] \tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l \bar{\epsilon}_{l,1-\tau_k}\right) \right].\end{aligned}$$

Dividing by λ^t and rearranging terms gives

$$(1-\lambda)\epsilon_{K_\infty,1-\tau_k} = \tilde{\epsilon}_{K_1,1-\tau_k} \frac{\lambda(1-\beta)}{1-\beta\lambda} - \beta \frac{1-\lambda}{1-\beta\lambda} \times r \left[\tau_k \epsilon_{K_\infty,1-\tau_k} + \mathrm{E}_\Gamma[\tau_l'] \tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\left(\tau_l', y^l \bar{\epsilon}_{l,1-\tau_k}\right) \right],$$

which is identical to condition (A.4). This quadratic equation has therefore the same solutions. This completes the proof. □

A.3 Auxiliary Lemmas and their Proofs

In this section I state and prove some auxiliary Lemmas, which I use in the proofs above.

A.3.1 Inter-temporal Consumption Response to Interest Rate Changes

Lemma A.1. *Let Assumptions 1-3 be satisfied. Then*

$$\frac{\partial c_t(k_0,\eta)}{\partial \bar{r}_0} = \frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_0}$$

for all $t \geq 0$. Furthermore for any $s > 0$ we have that

$$\frac{\partial c_t(k_0,\eta)}{\partial \bar{r}_s} = \frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s}$$

for all $t < s$ and

$$\frac{\partial c_t(k_0,\eta)}{\partial \bar{r}_s} = \frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} - \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}$$

for all $t \geq s$.

Proof. The Euler equation in period $s-1$ is given by

$$u_{c,s-1}(k_0,\eta) = \beta(1+\bar{r}_s)u_{c,s}(k_0,\eta)$$

Partially deriving with respect to $\bar{r}_s$ gives

$$u_{cc,s-1}(k_0,\eta)\frac{\partial c_{s-1}(k_0,\eta)}{\partial \bar{r}_s} + u_{cl,s-1}(k_0,\eta)\frac{\partial l_{s-1}(k_0,\eta)}{\partial \bar{r}_s} = \beta u_{c,s}(k_0,\eta) + u_{cc,s}(k_0,\eta)\frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s} + u_{cl,s}(k_0,\eta)\frac{\partial l_s(k_0,\eta)}{\partial \bar{r}_s}.$$

Using Lemma C.1 and evaluating at the steady state gives

$$u_{cc}(k_0,\eta)\frac{\partial c_{s-1}(k_0,\eta)}{\partial \bar{r}_s} = \beta u_c(k_0,\eta) + u_{cc}(k_0,\eta)\frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s},$$

which is equivalent to

$$\frac{\partial c_{s-1}(k_0,\eta)}{\partial \bar{r}_s} = \frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s} + \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}$$

Furthermore, partially deriving the Euler equations in all other periods one obtains that

$$\frac{\partial c_t(k_0,\eta)}{\partial \bar{r}_s} = \frac{\partial c_{s-1}(k_0,\eta)}{\partial \bar{r}_s}$$

$t < s-1$ and that

$$\frac{\partial c_t(k_0,\eta)}{\partial \bar{r}_s} = \frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s}$$

for $t > s$, which proofs the Lemma. □

A.3.2 Elasticity of Capital Supply with Respect to Unearned Income

Lemma A.2. *Let Assumptions 1-4 be satisfied. Then*

$$\tilde{\epsilon}_{k_t(k_0,\eta),T_s} = \frac{T}{k_0}\frac{\partial k_t(k_0,\eta)}{\partial T_s} = \begin{cases} -\frac{T}{k_0}\beta^{s+1}(\beta^{-t}-1) & \text{if } 1 \le t \le s \\ \frac{T}{k_0}\beta^{s+1} & \text{if } t > s \ge 0. \end{cases}$$

and

$$\tilde{\epsilon}_{K_t,T_s} = \frac{T}{K}\frac{\partial K_t}{\partial T_s} = \begin{cases} -\frac{T}{K}\beta^{s+1}(\beta^{-t}-1) & \text{if } 1 \le t \le s \\ \frac{T}{K}\beta^{s+1} & \text{if } t > s \ge 0. \end{cases}$$

As a consequence,

$$\tilde{\epsilon}_{k_t(k_0,\eta),T_s} = \frac{K}{k_0}\tilde{\epsilon}_{K_t,T_s}.$$

Proof. Throughout the proof I use the shorthand notation

$$u_{c,t}(k_0,\eta) \equiv u_c(c_t(k_0,\eta), l_t(k_0,\eta)) \quad \text{and} \quad u_{l,t}(k_0,\eta) \equiv u_l(c_t(k_0,\eta), l_t(k_0,\eta))$$

The Euler equation of an agent with initial state (k_0,η) is given by

$$u_{c,t}(k_0,\eta) = \beta(1+(1-\tau_k)r_{t+1})u_{c,t+1}(k_0,\eta).$$

Deriving with respect to T_s gives

$$\begin{aligned} u_{cc,t}(k_0,\eta)\frac{\partial c_t(k_0,\eta)}{\partial T_s} + u_{cl,t}(k_0,\eta)\frac{\partial l_t(k_0,\eta)}{\partial T_s} \\ = \beta(1+(1-\tau_k)r_{t+1})\left[u_{cc,t+1}(k_0,\eta)\frac{\partial c_{t+1}(k_0,\eta)}{T_s} + u_{cl,t+1}(k_0,\eta)\frac{\partial l_{t+1}(k_0,\eta)}{\partial T_s}\right] \end{aligned}$$

Using Lemma C.1 and evaluating at the steady state, at which we have $\beta(1+(1-\tau_k)r_{t+1}) = 1$, gives

$$u_{cc}(k_0,\eta)\frac{\partial c_t(k_0,\eta)}{\partial T_s} = u_{cc}(k_0,\eta)\frac{\partial c_{t+1}(k_0,\eta)}{\partial T_s}.$$

It follows that for all $t \geq 1$

$$\frac{\partial c_t(k_0,\eta)}{\partial T_s} = \frac{\partial c_0(k_0,\eta)}{\partial T_s}$$

The inter-temporal budget constraint is given by

$$\begin{aligned} c_0(k_0,\eta) + \sum_{t=1}^{\infty}\frac{c_t(k_0,\eta)}{\prod_{u=1}^{t}(1+(1-\tau_k)r_u)} &= (1+(1-\tau_k)r_0)k_0 + \eta w_0 l_0(k_0,\eta) \\ -\tau_l(\eta w_0 l_0(k_0,\eta)) + T_0 &+ \sum_{t=1}^{\infty}\frac{\eta w_t l_t(k_0,\eta) - \tau_l(\eta w_t l_t(k_0,\eta)) + T_t}{\prod_{u=1}^{t}(1+(1-\tau_k)r_u)}. \end{aligned}$$

Partially deriving with respect to T_s, using Lemma C.1, and evaluating at the steady state gives

$$\sum_{u=0}^{\infty}\beta^u\frac{\partial c_u(k_0,\eta)}{\partial T_s} = \beta^s$$

Using that $\frac{\partial c_u(k_0,\eta)}{\partial T_s} = \frac{\partial c_0(k_0,\eta)}{\partial T_s}$ for all u we get

$$\frac{\partial c_u(k_0,\eta)}{\partial T_s} = (1-\beta)\beta^s$$

for all u.

Consider $s = 0$. The budget constraint in the initial period $t = 0$ is

$$c_0(k_0, \eta) + k_1(k_0, \eta) = (1 + (1 - \tau_k)r)k_0 + \eta w l_0(k_0, \eta) - \tau_l(\eta w l_0(k_0, \eta)) + T_0$$

Hence,

$$\frac{\partial k_1(k_0, \eta)}{\partial T_0} = 1 - \frac{\partial c_0(k_0, \eta)}{\partial T_0} + \eta w(1 - \tau_l'(\eta w l(k_0, \eta)))\frac{\partial l_0(k_0, \eta)}{\partial T_0} = 1 - (1 - \beta) = \beta,$$

Now consider an arbitrary $t \geq 1$ and assume that $\frac{\partial k_t(k_0,\eta)}{\partial T_0} = \beta$. The budget constraint in t is

$$c_t(k_0, \eta) + k_{t+1}(k_0, \eta) = (1 + (1 - \tau_k)r)k_t + \eta w l_t(k_0, \eta) - \tau_l(\eta w l_t(k_0, \eta)) + T_t$$

Partially deriving with respect to T_0 and evaluating at the steady state gives

$$\begin{aligned}
\frac{\partial k_{t+1}(k_0, \eta)}{\partial T_0} =& (1 + (1 - \tau_k)r)\frac{\partial k_t(k_0, \eta)}{\partial T_0} - \frac{\partial c_t(k_0, \eta)}{\partial T_0} + \eta w(1 - \tau_l'(\eta w l(k_0, \eta)))\frac{\partial l_t k_0, \eta)}{\partial T_0} \\
=& \frac{1}{\beta}\frac{\partial k_t(k_0, \eta)}{\partial T_0} - (1 - \beta) \\
=& \frac{1}{\beta}\beta - (1 - \beta) \\
=& \beta,
\end{aligned}$$

which implies that

$$\frac{\partial K_t}{\partial T_s} = \beta.$$

for all $t \geq 1$, proving the condition of the Lemma for $s = 0$.

Now consider $s \geq 1$. I will show the condition for $1 \leq t \leq s$ by induction. First I will show that it holds for $\tilde{\epsilon}_{K_1,T_s}$. Partially deriving the budget constraint in the initial period $t = 0$ with respect to T_s gives

$$\frac{\partial k_1(k_0, \eta)}{\partial T_s} = -\frac{\partial c_0(k_0, \eta)}{\partial T_s} + \eta w(1 - \tau_l'(\eta w l(k_0, \eta)))\frac{\partial l_0(k_0, \eta)}{\partial T_s} = -(1 - \beta)\beta^s,$$

which implies that

$$\frac{\partial K_1}{\partial T_s} = -(1 - \beta)\beta^s = -\beta^{s+1}(\beta^{-1} - 1).$$

Hence the condition of the Lemma is satisfied for any $s \geq 1$ and $t = 1$.

Now assume that the condition is satisfied for an arbitrary $t-1 < s$. That is, let $t-1 < s$ with

$$\frac{\partial k_{t-1}(k_0,\eta)}{\partial T_s} = -\beta^{s+1}(\beta^{-(t-1)} - 1).$$

Implicitly deriving the budget constraint in period $t-1$ with respect to T_s gives

$$\begin{aligned}
\frac{\partial k_t(k_0,\eta)}{\partial T_s} =& \frac{1}{\beta}\frac{\partial k_{t-1}(k_0,\eta)}{\partial T_s} - \frac{\partial c_{t-1}(k_0,\eta)}{\partial T_s} + \eta w(1-\tau_l'(\eta w l(k_0,\eta)))\frac{\partial l_{t-1}(k_0,\eta)}{\partial T_s} \\
=& -\beta^s(\beta^{-(t-1)} - 1) - (1-\beta)\beta^s \\
=& -\beta^s(\beta^{-(t-1)} - 1 + 1 - \beta) \\
=& -\beta^{s+1}(\beta^{-t} - 1)
\end{aligned}$$

Hence for all (t,s) with $s \geq 1$ and $1 \leq t \leq s$ we have

$$\frac{\partial K_t}{\partial T_s} = -\beta^{s+1}(\beta^{-t} - 1).$$

Finally, we also show that the condition of the Lemma holds for all $t > s > 0$. We again use an induction argument. First, consider the case $t = s+1$. From above we know that

$$\frac{\partial k_s(k_0,\eta)}{\partial T_s} = -\beta^{s+1}(\beta^{-s} - 1).$$

Partially deriving the budget constraint in period $s = t-1$ gives

$$\begin{aligned}
\frac{\partial k_t(k_0,\eta)}{\partial T_s} =& \frac{1}{\beta}\frac{\partial k_s(k_0,\eta)}{\partial T_s} - \frac{\partial c_s(k_0,\eta)}{\partial T_s} + \eta w(1-\tau_l'(\eta w l(k_0,\eta)))\frac{\partial l_s(k_0,\eta)}{\partial T_s} + 1 \\
=& -\beta^s(\beta^{-s} - 1) - (1-\beta)\beta^s + 1 \\
=& -1 + \beta^s - \beta^s + \beta^{s+1} + 1 \\
=& \beta^{s+1}
\end{aligned}$$

Now assume that for any arbitrary $t > s$ this condition holds. That is for any $t > s$ we have

$$\frac{\partial k_t(k_0,\eta)}{\partial T_s} = \frac{\partial K_t}{\partial T_s} = \beta^{s+1}$$

Partially deriving the budget constraint in period t with respect to T_s gives

$$\begin{aligned}
\frac{\partial k_{t+1}(k_0,\eta)}{\partial T_s} =& \frac{1}{\beta}\frac{\partial k_t(k_0,\eta)}{\partial T_s} - \frac{\partial c_t(k_0,\eta)}{\partial T_s} + \eta w(1-\tau_l'(\eta w l(k_0,\eta)))\frac{\partial l_t(k_0,\eta)}{\partial T_s} \\
=& \beta^s - (1-\beta)\beta^s \\
=& \beta^s - \beta^s + \beta^{s+1} \\
=& \beta^{s+1}
\end{aligned}$$

Hence for all $t > s > 0$ we have

$$\frac{\partial k_t(k_0,\eta)}{\partial T_s} = \frac{\partial k_{t+1}(k_0,\eta)}{\partial T_s} = \frac{\partial K_t}{\partial T_s} = \frac{\partial K_{t+1}}{\partial T_s} = \beta^{s+1}.$$

□

A.3.3 Elasticity of Capital Supply with Respect to Wages

Lemma A.3. *Let Assumptions 1-4 be satisfied. Then*

$$\tilde{\epsilon}_{k_t(k_0,\eta),w_s} = \frac{w}{k_0}\frac{\partial k_t(k_0,\eta)}{\partial w_s} = \begin{cases} -\frac{y^l(k_0,\eta)\left(1-\tau_l'(y^l(k_0,\eta))\right)}{k_0}\beta^{s+1}(\beta^{-t}-1) & \text{if } 1 \leq t \leq s \\ \frac{y^l(k_0,\eta)\left(1-\tau_l'(y^l(k_0,\eta))\right)}{k_0}\beta^{s+1} & \text{if } t > s \geq 0. \end{cases}$$

and

$$\tilde{\epsilon}_{K_t,w_s} = \frac{w}{K}\frac{\partial K_t}{\partial w_s} = \begin{cases} -\frac{(1-\bar{\tau}_l')Y^l}{K}\beta^{s+1}(\beta^{-t}-1) & \text{if } 1 \leq t \leq s \\ \frac{(1-\bar{\tau}_l')Y^l}{K}\beta^{s+1} & \text{if } t > s \geq 0. \end{cases}$$

Proof. Throughout the proof I use the shorthand notation

$$u_{c,t}(k_0,\eta) \equiv u_c(c_t(k_0,\eta), l_t(k_0,\eta)) \quad \text{and} \quad u_{l,t}(k_0,\eta) \equiv u_l(c_t(k_0,\eta), l_t(k_0,\eta))$$

The Euler equation of an agent with initial state (k_0,η) is given by

$$u_{c,t}(k_0,\eta) = \beta(1+(1-\tau_k)r_{t+1})u_{c,t+1}(k_0,\eta).$$

Deriving with respect to w_s gives

$$u_{cc,t}(k_0,\eta)\frac{\partial c_t(k_0,\eta)}{\partial w_s} + u_{cl,t}(k_0,\eta)\frac{\partial l_t(k_0,\eta)}{\partial w_s}$$
$$= \beta(1+(1-\tau_k)r_{t+1})\left[u_{cc,t+1}(k_0,\eta)\frac{\partial c_{t+1}(k_0,\eta)}{w_s} + u_{cl,t+1}(k_0,\eta)\frac{\partial l_{t+1}(k_0,\eta)}{\partial w_s}\right]$$

Using Lemma C.1 and evaluating at the steady state, at which we have $\beta(1+(1-\tau_k)r_{t+1}) = 1$, gives for all $s \notin \{t, t+1\}$

$$\frac{\partial c_t(k_0,\eta)}{\partial w_s} = \frac{\partial c_{t+1}(k_0,\eta)}{\partial w_s}.$$

However, for $s = t$ we have

$$\frac{\partial c_t(k_0,\eta)}{\partial w_t} + \frac{u_{cl}(k_0,\eta)}{u_{cc}(k_0,\eta)}\tilde{\epsilon}_{l(k_0,\eta),w}\frac{l(k_0,\eta)}{w} = \frac{\partial c_{t+1}(k_0,\eta)}{\partial w_t},$$

while for $s = t+1$ we have

$$\frac{\partial c_t(k_0,\eta)}{\partial w_{t+1}} = \frac{\partial c_{t+1}(k_0,\eta)}{\partial w_{t+1}} + \frac{u_{cl}(k_0,\eta)}{u_{cc}(k_0,\eta)}\tilde{\epsilon}_{l(k_0,\eta),w}\frac{l(k_0,\eta)}{w}.$$

Furthermore, observe that

$$\frac{u_{cl}(k_0,\eta)}{u_{cc}(k_0,\eta)}\tilde{\epsilon}_{l(k_0,\eta),w}\frac{l(k_0,\eta)}{w} = -\eta l(k_0,\eta)\big(1-\tau_l'(\eta w l(l_0,\eta))\big)\tilde{\epsilon}_{l(k_0,\eta),w}.$$

Together these results imply that the consumption change is constant for all $t \neq s$

$$\frac{\partial c_t(k_0,\eta)}{\partial w_s} = \frac{\partial c_{\neq s}(k_0,\eta)}{\partial w_s},$$

while for $t = s$ we have

$$\frac{\partial c_t(k_0,\eta)}{\partial w_s} = \frac{\partial c_{\neq s}(k_0,\eta)}{\partial w_s} + \eta l(k_0,\eta)\big(1-\tau_l'(\eta w l(l_0,\eta))\big)\tilde{\epsilon}_{l(k_0,\eta),w}.$$

The inter-temporal budget constraint is given by

$$\begin{aligned} c_0(k_0,\eta) + \sum_{t=1}^{\infty}\frac{c_t(k_0,\eta)}{\prod_{u=1}^{t}(1+(1-\tau_k)r_u)} &= (1+(1-\tau_k)r_0)k_0 + \eta w_0 l_0(k_0,\eta) \\ -\tau_l(\eta w_0 l_0(k_0,\eta)) + T_0 &+ \sum_{t=1}^{\infty}\frac{\eta w_t l_t(k_0,\eta) - \tau_l(\eta w_t l_t(k_0,\eta)) + T_t}{\prod_{u=1}^{t}(1+(1-\tau_k)r_u)}. \end{aligned}$$

Partially deriving with respect to w_s, using Lemma C.1, and evaluating at the steady state gives

$$\sum_{u=0}^{\infty}\beta^u\frac{\partial c_u(k_0,\eta)}{\partial w_s} = \beta^s\big[1-\tau_l'(\eta w l(k_0,\eta))\big]\eta l(k_0,\eta)\big[1+\tilde{\epsilon}_{l(k_0,\eta),w}\big]$$

Using the results above gives

$$\frac{\partial c_{\neq s}(k_0,\eta)}{\partial w_s} = (1-\beta)\beta^s\big[1-\tau_l'(\eta w l(k_0,\eta))\big]\eta l(k_0,\eta)$$

and

$$\frac{\partial c_s(k_0,\eta)}{\partial w_s} = \Big[(1-\beta)\beta^s + \tilde{\epsilon}_{l(k_0,\eta),w}\Big]\big[1-\tau_l'(\eta w l(k_0,\eta))\big]\eta l(k_0,\eta)$$

Consider $s = 0$. The budget constraint in the initial period $t = 0$ is

$$c_0(k_0,\eta) + k_1(k_0,\eta) = (1+(1-\tau_k)r)k_0 + \eta w_0 l_0(k_0,\eta) - \tau_l(\eta w_0 l_0(k_0,\eta)) + T_0$$

Hence,

$$\begin{aligned}
\frac{\partial k_1(k_0,\eta)}{\partial w_0} =& -\frac{\partial c_0(k_0,\eta)}{\partial w_0} + \eta l(k_0,\eta)\big(1-\tau_l'(\eta l(k_0,\eta)w)\big)\left[1+\tilde{\epsilon}_{l(k_0,\eta),w}\right] \\
=& -\left[(1-\beta)+\tilde{\epsilon}_{l(k_0,\eta),w}\right]\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) \\
&+ \eta l(k_0,\eta)\big(1-\tau_l'(\eta l(k_0,\eta)w)\big)\left[1+\tilde{\epsilon}_{l(k_0,\eta),w}\right] \\
=& \beta\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta).
\end{aligned}$$

Now consider an arbitrary $t \geq 1$ and assume that

$$\frac{\partial k_t(k_0,\eta)}{\partial w_0} = \beta\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta).$$

The budget constraint in t is

$$c_t(k_0,\eta) + k_{t+1}(k_0,\eta) = (1+(1-\tau_k)r)k_t + \eta w_t l_t(k_0,\eta) - \tau_l(\eta w l_t(k_0,\eta)) + T_t.$$

Partially deriving with respect to w_0 and using Lemma C.1 gives

$$\begin{aligned}
\frac{\partial k_{t+1}(k_0,\eta)}{\partial w_0} =& -\frac{\partial c_t(k_0,\eta)}{\partial w_0} + \frac{1}{\beta}\frac{\partial k_t(k_0,\eta)}{\partial w_0} \\
=& -(1-\beta)\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) + \left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) \\
=& \beta\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta)
\end{aligned}$$

Therefore,

$$\tilde{\epsilon}_{k_t(k_0,\eta),w_0} = \beta\frac{\left[1-\tau_l'(y^l(k_0,\eta))\right]y^l(k_0,\eta)}{k_0}$$

and

$$\tilde{\epsilon}_{K_t,w_0} = \beta\frac{(1-\bar{\tau}_l')Y^l}{K},$$

which proves the statement for $s = 0$.

Next, consider $s \geq 1$. I will show the condition for $1 \leq t \leq s$ by induction. First, I will show that it holds for $\tilde{\epsilon}_{K_1,w_s}$. Partially deriving the budget constraint in the initial period $t = 0$ with respect to w_s gives

$$\frac{\partial k_1(k_0,\eta)}{\partial w_1} = -\frac{\partial c_0(k_0,\eta)}{\partial w_1} = -(1-\beta)\beta^s\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta)$$

and therefore

$$\tilde{\epsilon}_{k_1(k_0,\eta),w_s} = -(1-\beta)\beta^s \frac{\left[1-\tau_l'(y^l(k_0,\eta))\right] y^l(k_0,\eta)}{k_0},$$

as well as

$$\tilde{\epsilon}_{K_1,w_1} = -(1-\beta)\beta \frac{(1-\bar{\tau}_l')\big] Y^l}{K_0},$$

which proves the statement for $s \geq 1$ and $t = 1$.

Now assume that the statement is satisfied for an arbitrary $t-1 < s$. That is, let $t-1 < s$ with

$$\frac{\partial k_{t-1}(k_0,\eta)}{\partial w_1} = -(\beta^{-(t-1)} - 1)\beta^{s+1}\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta).$$

Implicitly deriving the budget constraint in period $t-1$ with respect to w_s gives

$$\begin{aligned}
\frac{\partial k_t(k_0,\eta)}{\partial w_s} =& \frac{1}{\beta}\frac{\partial k_{t-1}(k_0,\eta)}{\partial w_s} - \frac{\partial c_{t-1}(k_0,\eta)}{\partial w_s} \\
=& -(\beta^{-(t-1)} - 1)\beta^s\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) - (1-\beta)\beta^s\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) \\
=& -(\beta^{-(t-1)} - \beta)\beta^s\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) \\
=& -(\beta^{-t} - 1)\beta^{s+1}\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta),
\end{aligned}$$

which proves the statement for all (t,s) where $s \geq 1$ and $1 \leq t \leq s$.

Finally, we also show that the condition of the Lemma holds for all $t > s > 0$. We again use an induction argument. First, consider the case $t = s+1$. From above we know that

$$\frac{\partial k_s(k_0,\eta)}{\partial w_s} = -(\beta^{-s} - 1)\beta^{s+1}\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta).$$

Partially deriving the budget constraint in period $s = t-1$ gives

$$\begin{aligned}
\frac{\partial k_t(k_0,\eta)}{\partial w_s} =& \frac{1}{\beta}\frac{\partial k_s(k_0,\eta)}{\partial w_s} - \frac{\partial c_s(k_0,\eta)}{\partial w_s} + \eta l(k_0,\eta)\left(1-\tau_l'(\eta w l(k_0,\eta))\right)\left[1+\tilde{\epsilon}_{l(k_0,\eta),w}\right] \\
=& -(\beta^{-s} - 1)\beta^s\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) \\
& - \left[(1-\beta)\beta^s + \tilde{\epsilon}_{l(k_0,\eta),w}\right]\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) \\
& + \eta l(k_0,\eta)\left(1-\tau_l'(\eta w l(k_0,\eta))\right)\left[1+\tilde{\epsilon}_{l(k_0,\eta),w}\right] \\
=& \beta^{s+1}\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta),
\end{aligned}$$

which proves the statement for $t = s+1$. Now assume that for any arbitrary $t > s$ this condition holds, that is for any $t > s$ we have

$$\frac{\partial k_t(k_0,\eta)}{\partial w_s} = \beta^{s+1}\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta)$$

Partially deriving the budget constraint in period t with respect to w_s gives

$$\begin{aligned}\frac{\partial k_{t+1}(k_0,\eta)}{\partial w_s} =& \frac{1}{\beta}\frac{\partial k_t(k_0,\eta)}{\partial w_s} - \frac{\partial c_t(k_0,\eta)}{\partial w_s}\\ =&\beta^s\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta) - (1-\beta)\beta^s\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta)\\ =&\beta^{s+1}\left[1-\tau_l'(\eta w l(k_0,\eta))\right]\eta l(k_0,\eta).\end{aligned}$$

Hence for all $t > s > 0$ we have

$$\tilde{\epsilon}_{k_t(k_0,\eta),w_s} = \beta^{s+1}\frac{\left[1-\tau_l'(y^l(k_0,\eta))\right]y^l(k_0,\eta)}{k_0}$$

and

$$\tilde{\epsilon}_{K_t,w_s} = \beta^{s+1}\frac{(1-\bar{\tau}_l')Y^l}{K}.$$

This completes the proof. □

A.3.4 Elasticity of Capital Supply with Respect to Net Returns

Lemma A.4. *Let Assumptions 1-4 be satisfied. Then*

$$\tilde{\epsilon}_{K_t,\bar{r}_s} = \begin{cases} 1-\beta & \text{if } s=0\\ \beta^s(\beta^{-t}-1)\left[\tilde{\epsilon}_{K_1,1-\tau_k}-(1-\beta)\right] & \text{if } s>0 \wedge t\le s\\ \beta^s(1-\beta)+(1-\beta^s)\tilde{\epsilon}_{K_1,1-\tau_k} & \text{if } s>0\wedge t>s\end{cases}$$

Proof. The inter-temporal budget constraint is given by

$$\begin{aligned}c_0(k_0,\eta)+\sum_{t=1}^{\infty}\frac{c_t(k_0,\eta)}{\prod_{u=1}^t(1+\bar{r}_u)} =& (1+\bar{r}_0)k_0+\eta w_0 l_0(k_0,\eta)-\tau_l(\eta w_0 l_0(k_0,\eta))+T_0\\ &+\sum_{t=1}^{\infty}\frac{\eta w_t l_t(k_0,\eta)-\tau_l(\eta w_t l_t(k_0,\eta))+T_t}{\prod_{u=1}^t(1+\bar{r}_u)}.\end{aligned}$$

Consider first $s=0$. Deriving the intertemporal budget constraint with respect to $\bar{r}_0$ and using Assumption 4 gives

$$\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_0}+\sum_{t=1}^{\infty}\frac{\partial c_t(k_0,\eta)}{\partial \bar{r}_0}\frac{1}{\prod_{u=1}^t(1+\bar{r}_u)} = k_0.$$

Evaluating at the steady state and using Lemma A.1 gives

$$\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_0} = (1-\beta)k_0, \tag{A.5}$$

The budget constraint in $t = 0$ is

$$k_1(k_0,\eta) + c_0(k_0,\eta) = (1+\bar{r}_0)k_0 + \eta w_0 l_0(k_0,\eta) - \tau_l(\eta w_0 l_0(k_0,\eta)) + T_0$$

Hence

$$\frac{\partial k_1(k_0,\eta)}{\partial \bar{r}_0} = k_0 - \frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_0} = \beta k_0.$$

Similarly, partially deriving the budget constraint in any period $t \geq 1$ and evaluating at the steady state gives

$$\frac{\partial k_{t+1}(k_0,\eta)}{\partial \bar{r}_0} = (1+\bar{r})\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_0} - \frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_0} = (1+\bar{r})\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_0} - (1-\beta)k_0.$$

This implies

$$\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_0} = \beta k_0$$

for all $t \geq 1$. Hence, $\epsilon_{k_t(k_0,\eta),1-\tau_k} = \beta\bar{r}$ and aggregating over all agents gives $\epsilon_{K_t,1-\tau_k} = \beta\bar{r} = 1-\beta$, i.e. the first part of the Lemma.

Next, consider an arbitrary $s > 0$. Deriving the intertemporal budget constraint with respect to $\bar{r}_s$ for $s > 0$ gives

$$\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} + \sum_{t=1}^{\infty} \frac{\partial c_t(k_0,\eta)}{\partial \bar{r}_s} \frac{1}{\prod_{u=1}^{t}(1+\bar{r}_u)} - \sum_{t=s}^{\infty} \frac{c_t(k_0,\eta)(1+\bar{r}_s)^{-1}}{\prod_{u=1}^{t}(1+\bar{r}_u)} = - \sum_{t=s}^{\infty} \frac{[\eta w_t l_t(k_0,\eta) - \tau_l(\eta w_t l_t(k_0,\eta)) + T_t](1+\bar{r}_s)^{-1}}{\prod_{u=1}^{t}(1+\bar{r}_u)}.$$

Using Lemma A.1 and evaluating at the steady state gives

$$\frac{1-\beta^s}{1-\beta}\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} + \frac{\beta^s}{1-\beta}\frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s} = \sum_{t=s}^{\infty} \beta^{t+1}\bar{r}k_0, \tag{A.6}$$

where we used that in steady state $c(k_0,\eta) - \eta w l(k_0,\eta) - \tau_l(\eta w l(k_0,\eta)) - T = \bar{r}k_0$.

Using Lemma A.1 one can plug in for $\frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s} = \frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} - \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}$ to obtain

$$\frac{1-\beta^s}{1-\beta}\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} + \frac{\beta^s}{1-\beta}\left[\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} - \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}\right] = \sum_{t=s}^{\infty} \beta^{t+1}\bar{r}k_0,$$

which, using that $\bar{r} = \frac{1-\beta}{\beta}$, is equivalent to

$$\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} = (1-\beta)\beta^s k_0 + \beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}. \tag{A.7}$$

In the following, I will show by induction that for all $1 \leq t \leq s$ we have

$$\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_s} = \left[\beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + (1-\beta)\beta^s k_0\right] \beta \frac{1-\beta^{-t}}{1-\beta}.$$

Consider first $t = 1$. Deriving the budget constraint in the initial period and evaluating at the steady state gives

$$\begin{aligned}
\frac{\partial k_1(k_0,\eta)}{\partial \bar{r}_s} &= -\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} \\
&= -\beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} - (1-\beta)\beta^s k_0,
\end{aligned}$$

which satisfies the statement. Next, assume that the condition holds for an arbitrary $1 \leq t-1 < s$. Partially deriving the budget constraint at period $t-1$ and evaluating at the steady state gives

$$\begin{aligned}
\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_s} =& (1+\bar{r})\frac{\partial k_{t-1}(k_0,\eta)}{\partial \bar{r}_s} - \frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s} \\
=& \left[\beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + (1-\beta)\beta^s k_0\right] \frac{1-\beta^{-t+1}}{1-\beta} - \beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} - (1-\beta)\beta^s k_0 \\
=& \left[\beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + (1-\beta)\beta^s k_0\right] \frac{1-\beta^{-t+1}-1+\beta}{1-\beta} \\
=& \left[\beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + (1-\beta)\beta^s k_0\right] \beta \frac{1-\beta^{-t}}{1-\beta},
\end{aligned}$$

which proves the statement.

Hence for $s > 0$ and $1 \leq t \leq s$ we have that

$$\begin{aligned}
\tilde{\epsilon}_{k_t(k_0,\eta),\bar{r}_s} =& \frac{\bar{r}}{k_0} \left[\beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + (1-\beta)\beta^s k_0\right] \beta \frac{1-\beta^{-t}}{1-\beta} \\
=& \left[\frac{1}{k_0}\beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + (1-\beta)\beta^s\right] (1-\beta^{-t}).
\end{aligned}$$

Now remember that by Lemma C.3 the pure supply elasticity of capital in the first period is

$$\tilde{\epsilon}_{k_1(k_0,\eta),1-\tau_k} = -\beta \frac{c_0(k_0,\eta)}{k_0} \frac{u_c(k_0,\eta)}{c_0(k_0,\eta)u_{cc}(k_0,\eta)}.$$

Hence,

$$\tilde{\epsilon}_{k_t(k_0,\eta),\bar{r}_s} = \left[\beta^s \tilde{\epsilon}_{k_1(k_0,\eta),1-\tau_k} - (1-\beta)\beta^s\right](\beta^{-t}-1).$$

The aggregate elasticity is then given by

$$\begin{aligned}
\tilde{\epsilon}_{K_t,\bar{r}_s} =& \int \frac{k_0}{K_0} \tilde{\epsilon}_{k_t(k_0,\eta),\bar{r}_s} d\Gamma \\
=&(\beta^{-t}-1) \int \frac{k_0}{K_0} \left[\beta^s \tilde{\epsilon}_{k_1(k_0,\eta),1-\tau_k} - (1-\beta)\beta^s\right] d\Gamma \\
=&(\beta^{-t}-1)\beta^s \int \frac{k_0}{K_0} \tilde{\epsilon}_{k_1(k_0,\eta),1-\tau_k} d\Gamma - (\beta^{-t}-1)(1-\beta)\beta^s \\
=&(\beta^{-t}-1)\beta^s \left[\tilde{\epsilon}_{K_1,1-\tau_k} - 1 + \beta\right]
\end{aligned}$$

which proves the second part of the Lemma.

Finally, I will show again by induction that for any $t > s \geq 1$ we have

$$\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_s} = -\beta \frac{1-\beta^s}{1-\beta} \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + \beta^{s+1} k_0$$

Again using equation (A.6) but this time substituting out $\frac{\partial c_0(k_0,\eta)}{\partial \bar{r}_s}$ gives

$$\frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s} = (1-\beta)\beta^s k_0 - (1-\beta^s)\frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)}.$$

Implicitly deriving the budget constraint at $t = s$ and evaluating at the steady state gives

$$\frac{\partial k_{s+1}(k_0,\eta)}{\partial \bar{r}_s} = k_0 + (1+\bar{r})\frac{\partial k_s(k_0,\eta)}{\partial \bar{r}_s} - \frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s}$$

Plugging in the results above gives

$$\begin{aligned}
\frac{\partial k_{s+1}(k_0,\eta)}{\partial \bar{r}_s} =& k_0 + (1+\bar{r})\left[\beta^s \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + (1-\beta)\beta^s k_0\right]\beta\frac{1-\beta^{-s}}{1-\beta} \\
&+ (1-\beta^s)\frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} - (1-\beta)\beta^s k_0 \\
=& \frac{\beta^s - 1 + 1 - \beta - \beta^s + \beta^{s+1}}{1-\beta} \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + \beta^{s+1} k_0 \\
=& -\beta \frac{1-\beta^s}{1-\beta} \frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + \beta^{s+1} k_0
\end{aligned}$$

which shows the statement for $t = s+1$. Next, assume that the statement holds for an arbitrary $t-1 \geq s+1$. Partially deriving the budget constraint in period $t-1$ and evaluating at the steady state gives

$$
\begin{aligned}
\frac{\partial k_t(k_0,\eta)}{\partial \bar{r}_s} =&(1+\bar{r})\frac{\partial k_{t-1}(k_0,\eta)}{\partial \bar{r}_s} - \frac{\partial c_s(k_0,\eta)}{\partial \bar{r}_s}\\
=&(1+\bar{r})\left[-\beta\frac{1-\beta^s}{1-\beta}\frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + \beta^{s+1}k_0\right] + (1-\beta^s)\frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} - (1-\beta)\beta^s k_0\\
=&\frac{-1+\beta^s+1-\beta-\beta^s+\beta^{s+1}}{1-\beta}\frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + \beta^{s+1}k_0\\
=&-\beta\frac{1-\beta^s}{1-\beta}\frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + \beta^{s+1}k_0,
\end{aligned}
$$

which proves the statement.

Hence for $t > s > 0$ we have

$$
\begin{aligned}
\tilde{\epsilon}_{k_t(k_0,\eta),\bar{r}_s} =&\frac{\bar{r}}{k_0}\left[-\beta\frac{1-\beta^s}{1-\beta}\frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + \beta^{s+1}k_0\right]\\
=&-\frac{1}{k_0}(1-\beta^s)\frac{\beta u_c(k_0,\eta)}{u_{cc}(k_0,\eta)} + (1-\beta)\beta^s.
\end{aligned}
$$

Again remember that by Lemma C.3 the pure supply elasticity of capital in the first period is

$$
\tilde{\epsilon}_{k_1(k_0,\eta),1-\tau_k} = -\beta\frac{c_0(k_0,\eta)}{k_0}\frac{u_c(k_0,\eta)}{c_0(k_0,\eta)\big(u_{cc}(k_0,\eta)\big)}.
$$

Hence,

$$
\tilde{\epsilon}_{k_t(k_0,\eta),\bar{r}_s} =(1-\beta^s)\tilde{\epsilon}_{k_1(k_0,\eta),1-\tau_k} + (1-\beta)\beta^s.
$$

The aggregate elasticity is then given by

$$
\begin{aligned}
\tilde{\epsilon}_{K_t,\bar{r}_s} =&\int \frac{k_0}{K_0}\tilde{\epsilon}_{k_t(k_0,\eta),\bar{r}_s} d\Gamma\\
=&\int \frac{k_0}{K_0}\left[(1-\beta)\beta^s + (1-\beta^s)\tilde{\epsilon}_{k_1(k_0,\eta),1-\tau_k}\right] d\Gamma\\
=&\beta^s(1-\beta) + (1-\beta^s)\tilde{\epsilon}_{K_1,1-\tau_k}
\end{aligned}
$$

which proves the third part of the Lemma. □

B The General Framework

In this section I present a more general model. It nests the model of the main text, and consequently all its special cases. Relative to the framework in the main text, I introduce uninsurable labor income- and investment risk. It thus nests, additionally, the standard incomplete markets model of Aiyagari (1994) as well as the framework of Angeletos (2007) as special cases. It has been shown analytically that such frameworks, with capital- as well as labor income risk, can replicate the observed Pareto distribution in wealth (Benhabib, Bisin, and Zhu, 2011, 2015).

B.1 Model

B.1.1 General Setup

Demographics. There is a continuum of infinitely lived agents (dynasties) of measure one. Agents differ in their managerial/investing ability $\theta \in [0, \bar{\theta}]$, and their working ability $\eta \in [\underline{\eta}, \bar{\eta}]$. The unconditional density over the ability states $p^u(\theta, \eta)$ is stationary, continuous and strictly positive over the whole set $[0, \bar{\theta}] \times [\underline{\eta}, \bar{\eta}]$. Each agent is both an investor/entrepreneur and a worker. Equivalently, one can think of an agent as a household consisting of firm owners and workers who pool their income.

Timing of Events. Time is discrete. Agents make an investment decision at the end of each period t. This decision includes a portfolio choice as agents can investment in private equity (their own firm) or in riskless bonds issued by the corporate sector. At the beginning of the following period $t+1$ agents draw their productivity $(\theta_{t+1}, \eta_{t+1}) \in [0, \bar{\theta}] \times [\underline{\eta}, \bar{\eta}]$, after which they decide how much labor to hire. Then, production takes place and agents receive their net labor- and capital income.

Idiosyncratic Shocks. Each agent draws next period's productivity from the conditional density $p(\theta', \eta' | \theta, \eta)$. Next period's productivity (θ', η') is therefore correlated with current productivity (θ, η). This is a parsimonious way to account for the fact that heterogeneous productivity is partly driven by differences in ability (the initial state) and partly by luck. I assume that for each $(\theta, \eta) \in [0, \bar{\theta}] \times [\underline{\eta}, \bar{\eta}]$ the conditional density $p(\theta', \eta' | \theta, \eta) > 0$.

Preferences. Agents maximize their expected discounted lifetime utility

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t),$$

where the utility function satisfies the same properties as in the main text, that is Assumption 1 holds.

Technology. There are two sectors of production, a public corporate sector and a private entprepreneurial sector. Agents can decide to invest their money in the public corporate sector or in their own firm. The former can be interpreted as investment in public equity or bonds and is denoted by b. The latter can be interpreted as private equity investment and is denoted by e. Investment in the corporate sector is perfectly diversified and hence riskless, while investment in private equity is risky. Both sectors have access to the same production technology. In particular, Assumption 2 from the main text holds for both sectors.

Market Structure. There are incomplete financial markets. In particular, agents cannot insure against adverse realizations of the idiosyncratic productivity shock (θ', η'). Riskless bonds, issued by the corporate sector, are the only trade-able assets in this economy. Publicly traded corporations are assumed to hire labor and rent capital under perfect competition. Privately owned firms compete with public ones for the same employees.

B.1.2 Individual Optimization

Public Corporate Sector. The problem of public corporations is standard and shall be stated first. Corporations are assumed to be perfectly competitive. They rent capital K^c and hire labor L^c in order to maximize profits

$$\max_{K^c, L^c} \{F(K^c, L^c) - (r+\delta)K^c - wL^c\}.$$

Households. Next, I describe the individual optimization problem of households in recursive form. There are four individual states: bond holdings b, private equity investment e, investment efficiency (entrepreneurial ability) θ and working ability η. The aggregate state is given by the distribution Φ over individual states. The households have rational expectations and perfect foresight implying that at each period t the current distribution Φ_t is enough for the households to perfectly foresee the path of future prices and transfers, $\{(w_s, r_s, T_s)\}_{s=t}^{\infty}$. The households' problem is given by

$$\text{(H1)} \quad V(b, e, \theta, \eta; \Phi) = \max_{c, l, l^d, b', e'} \left\{ u(c, l) + \beta \mathrm{E}\left[V(b', e', \theta', \eta'; \Phi') | \theta, \eta; \Phi \right] \right\}$$

$$\begin{aligned} \text{s.t.} \quad & e' \geq 0 \\ & b' \geq -\underline{b}(e') \\ & k = \theta e \\ & \pi = F(k, l^d) - \delta k - w l^d \\ & c + e' + b' = (1 - \tau_k)\left[rb + \pi \right] + w\eta l - \tau_l(w\eta l) + e + b + T. \end{aligned}$$

Investment in the own firm needs to be non-negative, $e' \geq 0$. As in Aiyagari (1994), the household may borrow up to some amount $\underline{b}(e') \geq 0$. I allow this ammount to potentially increase in the household's business wealth e', as this may serve as collateral. Specifically, the borrowing limit $\underline{b}(.)$ is a continuous and weakly increasing function. Investment risk is modelled as in Angeletos (2007). Specifically, after the household invested in private equity e, an idiosyncratic shock θ is drawn that determines the effective capital stock k employed in her firm.

The household's business income is denoted by π. She finances consumption c and investment $e' + b'$ with her net capital income $(1 - \tau_k)\left[rb + \pi \right]$, her net labor income $w\eta l - \tau_l(w\eta l)$, her current asset holdings $e + b$, and the lump-sum transfer T. Note that the agent chooses her own labor supply l but also labor demand l^d, that is the amount of effective labor she wants to employ in her firm. As in Angeletos (2007), hiring takes place after the idiosyncratic productivity shock θ is observed.

The households' optimization problem (H1) is quite involved with four individual states (b, e, θ, η) and five decision variables (c, l, l^d, b', e'). Luckily, we can reduce the dimensionality of the problem because we can substitute out optimal labor hiring.

Lemma B.1. ***Proportionality.*** *Let $K^c > 0$ and consider an agent with private equity e and investment efficiency θ, i.e. effective capital θe. Then:*

(a) The ratio of effective capital to effective labor is the same as in the corporate sector,

$$\frac{k}{l^d} = \frac{K^c}{L^c}.$$

(b) The agent's business income is given by

$$\pi = r\theta e.$$

Proof. See Appendix A.2.1. □

Part (a) implies that all firms, public and private, employ the same ratio of effective capital to effective labor. Part (b) implies that firm profits are linear in effective capital $k = \theta e$. Using Lemma B.1, one can reduce the number of individual states to three, where bonds b and private equity e are replaced by financial wealth a (sometimes referred to as 'cash-on-hand'). The distribution Γ is now over the three individual states (a, θ, η). The reduced problem is given by

$$\begin{aligned}
\text{(H2)} \quad V(a, \theta, \eta; \Gamma) = \max_{c,l,b',e'} & \left\{ u(c, l) + \mathrm{E}\left[V(a', \theta', \eta'; \Gamma') | \theta, \eta; \Gamma \right] \right\} \\
\text{s.t.} \quad & e' \geq 0 \\
& b' \geq -\underline{b}(e') \\
& c + e' + b' = a + \eta w l - \tau_l(\eta w l) + T \\
& a' = (1 - \tau_k')r'(\theta' e' + b') + e' + b'.
\end{aligned}$$

B.1.3 Equilibrium

To economize on notation, I from now on summarize the individual state by $\mathbf{x} = (a, \theta, \eta)$.

Definition B.1. ***General Equilibrium.*** *Denote by $\mathcal{B}(\mathbb{R}^3)$ the Borel-Sigma algebra over $\mathbb{R}^3$ and by $\mathcal{P}$ the set of all measures over the measurable space $(\mathbb{R}^3, \mathcal{B}(\mathbb{R}^3))$. Given an initial measure $\Gamma_{-1} \in \mathcal{P}$, initial individual portfolio allocations $(e_0(\boldsymbol{x}), b_0(\boldsymbol{x}))$, and a sequence of capital tax rates and transfers $\{\tau_{k,t}, T_t\}_{t=0}^{\infty}$, a general equilibrium is defined by prices $\{w_t, r_t\}_{t=0}^{\infty}$, by input factors $\{K_t^c, L_t^c\}_{t=0}^{\infty}$ of public corporations, by household decision rules for consumption $c(\boldsymbol{x}; \Gamma)$, labor supply $l(\boldsymbol{x}; \Gamma)$, private equity $e'(\boldsymbol{x}; \Gamma)$, and bond* *holdings $b'(\boldsymbol{x};\Gamma)$; by a value function $V(\boldsymbol{x};\Gamma)$, and by a sequence of measures $\{\Gamma_t(\boldsymbol{x})\}_{t=0}^{\infty}$, such that for each $t \in \{0,1,2,...\}$:*

(a) Factor demand (K_t^c, L_t^c) of public corporations satisfies $r_t = F_k(K_t^c, L_t^c) - \delta$ and $w_t = F_l(K_t^c, L_t^c)$.

(b) Household decision rules are solutions to optimization problem (H2).

(c) The value function $V(\boldsymbol{x};\Gamma)$ solves the Bellman equation for all $\boldsymbol{x}$.

(d) The labor market clears,

$$L = \int \eta l(\boldsymbol{x};\Gamma) d\Gamma = L_t^c + \int p_\theta(\theta'|\theta,\eta) l^d(\theta', e(\boldsymbol{x};\Gamma_{t-1})) d\Gamma_{t-1},$$

where $l^d(\theta,e) = \arg\max_{l^d}\{F(\theta e, l^d) - w_t l^d\}$.

(e) The bond market clears,

$$\int b'(\boldsymbol{x};\Gamma_{t-1}) d\Gamma_{t-1} = K_t^c.$$

(f) The government budget clears,

$$\int\int p(\theta',\eta'|\theta,\eta)\Big(\tau_{k,t} r_t \big[b'(\boldsymbol{x};\Gamma_{t-1}) + \theta' e'(\boldsymbol{x};\Gamma_{t-1})\big] + \tau_l\big(\eta' w_t l(\boldsymbol{x};\Gamma)\big)\Big) d(\theta',\eta') d\Gamma_{t-1}$$
$$= G + T_t.$$

(g) The distribution evolves according to the law of motion

$$\Gamma_t = P_t(\Gamma_{t-1}),$$

where the transition function $P_t : \mathcal{P} \to \mathcal{P}$ can be written explicitly as follows. For each set $A \times \Theta \times H$ in $\mathcal{B}(\mathbb{R}^3)$

$$\Gamma_t(A \times \Theta \times H) = \int p(\theta',\eta'|\theta,\eta) 1_{(\theta',\eta')\in\Theta\times H} 1_{a'(\theta',\eta'|x;\Gamma_{t-1})\in A} d(\theta',\eta') d\Gamma_{t-1}(\boldsymbol{x}),$$

where 1 denotes the indicator function and

$$a'(\theta',\eta'|\boldsymbol{x};\Gamma_{t-1}) = (1-\tau_{k,t})\big[r_t\big(\theta' e'(\boldsymbol{x};\Gamma_{t-1}) + b'(\boldsymbol{x};\Gamma_{t-1})\big)\big] + e'(\boldsymbol{x};\Gamma_{t-1}) + b'(\boldsymbol{x};\Gamma_{t-1}).$$

We can also define a stationary equilibrium for this setting.

Definition B.2. *A **Stationary General Equilibrium** is a dynamic general equilibrium in which $(\tau_{k,t},\Gamma_t) = (\tau_k,\Gamma)$ for all $t \in \{0,1,2,...\}$ and as a consequence all variables are time-invariant.*

As in the main text we will assume that the economy is initially in steady state, that is that Assumption 3 holds.

B.2 Optimal Capital Taxation

I now move to the derivation of the optimality condition for the general framework. For simplicity, I abstract from pre-announcement, that is I assume that $t^a = 0$.

B.2.1 Social Welfare Criterion

I consider a very general set of social welfare functions. This generalization nests the welfarist approach with Pareto weights on individual values.[B.1]

Denote time-t histories by $\mathbf{x}^t = (a_0, \theta_0, \eta_0, \theta_1, \eta_1, ..., \theta_t, \eta_t)$. Furthermore, for each $t \in \{0, 1, 2, ...\}$ denote by

$$P_t(\mathbf{x}^t) = \int_{\mathbf{z}_0 \leq \mathbf{x}_0} \int_{\mathbf{z}^t \leq \mathbf{x}^t} p(\mathbf{z}^t | \mathbf{z}_0) d\mathbf{z}^t d\Gamma_0(\mathbf{z}_0)$$

the probability distribution over time-t histories $\mathbf{x}^t$.

Let $c_t(\mathbf{x}^t)$ be the agents' optimal consumption and $l_t(\mathbf{x}^t)$ the agents' optimal labor choice in period t given history $\mathbf{x}^t$. The planner's objective is given by

$$(\tilde{P}) \qquad \max_{\tau_k \leq 1} W = \sum_{t=0}^{\infty} \beta^t \int \omega_t(\mathbf{x}^t) u(c_t(\mathbf{x}^t), l_t(\mathbf{x}^t)) dP_t,$$

where $\omega_t(\mathbf{x}^t)$ is a function that assigns real values to each individual history $\mathbf{x}^t$ in a way such that $\int \omega_t(\mathbf{x}^t) dP_t(\mathbf{x}^t)$ is constant across all t.

Denote by $\mathbf{x}_s(\mathbf{x}^t) = (a_s, \theta_s, \eta_s)$ the individual state in period s corresponding to history $\mathbf{x}^t$. Then:

1. whenever $\omega_t(\mathbf{x}^t) = \omega(\mathbf{x}_t(\mathbf{x}^t)) = \omega(\mathbf{x}_t)$ for all histories $\mathbf{x}^t$, we have

$$W = \sum_{t=0}^{\infty} \beta^t \int \omega(\mathbf{x}_t) u\big(c(\mathbf{x}_t; \Gamma_t), l(\mathbf{x}_t; \Gamma_t)\big) d\Gamma_t,$$

   that is a social welfare function with weights ω on states $\mathbf{x}$.

2. whenever $\omega_t(\mathbf{x}^t) = \omega(\mathbf{x}_0(\mathbf{x}^t)) = \omega(\mathbf{x}_0)$ for all histories $\mathbf{x}^t$, we have

$$W = \int \omega(\mathbf{x}_0) V(\mathbf{x}_0; \Gamma_0) d\Gamma_0,$$

   a welfarist social welfare function with Pareto weights ω on agents that are identified by initial states $\mathbf{x}_0$.

[B.1] There is some discrepancy in the literature in terms of terminology. I refer to a 'welfarist' social welfare function as a weighted sum of agents' values and to a 'utilitarian' social welfare function when those weights are the same for all agents. Others refer to the former as 'generalized utilitarian' and to the latter as 'pure utilitarian' social welfare function.

Observe that the two social welfare criteria coincide whenever for each t the weights $\omega_t(\mathbf{x}^t)$ are constant across all time-t histories $\mathbf{x}^t$, which is the case in the deterministic framework of the main text or in stochastic settings when the welfare criterion is utilitarian, i.e. with the criterion most often used in the Macroeconomics literature.

B.2.2 The Optimality Condition for the General Case

Marginal social welfare weights are now given by $g_t(\mathbf{x}^t) = \omega_t(\mathbf{x}^t)u_c(c_t(\mathbf{x}^t), l_t(\mathbf{x}^t))$. Denote the average marginal social welfare weight in period t by

$$\bar{g}_t = \int g_t(\mathbf{x}^t)dP_t$$

In a stochastic environment, $\bar{g}_t$ is generally time-varying. With a dynastic interpretation of my model g_t would measure by how much, relative to the average, the government values generation t. More generally, it measures the planner's relative valuation of an additional dollar available at time t.

Without loss of generality we can normalize the weights $\omega_t(\mathbf{x}^t)$ such that

$$\bar{G} = (1-\beta)\sum_{t=0}^{\infty}\beta^t \int g_t(\mathbf{x}^t)dP_t = 1. \tag{B.1}$$

The following proposition states the optimality condition in the general framework under the general welfare objective.

Proposition B.1. ***Optimality Condition in the General Framework.*** *Let Assumptions 1-3 be satisfied. If the pre-existing capital income tax rate $\tau_k < 1$ solves the Planner's problem ($\tilde{P}$), then*

$$\begin{aligned}\tau_k =& \frac{\beta^{t^a}(1-\bar{G}^k)}{\hat{\varepsilon}_{K,1-\tau_k}} - \frac{\bar{\varepsilon}_{L,1-\tau_k}}{\hat{\varepsilon}_{K,1-\tau_k}}\frac{\alpha^l}{\alpha^k}\left[\mathrm{E}_\Gamma[\tau_l'] + \mathrm{Cov}_\Gamma\left(\tau_l', \frac{y^l\bar{\varepsilon}_{l,1-\tau_k}}{Y^l\bar{\varepsilon}_{L,1-\tau_k}}\right)\right] \\ &+ \frac{\bar{\varepsilon}_{w,1-\tau_k}}{\hat{\varepsilon}_{K,1-\tau_k}}\frac{\alpha^l}{\alpha^k}\left[(1-\tau_k)\bar{\mathcal{G}}^k - (1-\bar{\tau}_l')\tilde{\mathcal{G}}^l + \tau_k - \bar{\tau}_l'\right],\end{aligned} \tag{B.2}$$

where

$$\hat{\varepsilon}_{K,1-\tau_k} = (1-\beta)\sum_{t=0}^{\infty}\beta^t\varepsilon_{K_t,1-\tau_k}\bar{g}_t,$$

$$\bar{G}^k = \frac{1-\beta}{t^a}\sum_{t=0}^{\infty}\beta^t\int g_t(\boldsymbol{x}^t)\frac{y_t^k(\boldsymbol{x}^t)}{Y^k}dP_t,$$

$$\bar{\mathcal{G}}^k = \frac{1-\beta}{\bar{\varepsilon}_{w,1-\tau_k}}\sum_{t=0}^{\infty}\beta^t\epsilon_{w_t,1-\tau_k}\int g_t(\boldsymbol{x}^t)\frac{y^k(\boldsymbol{x}^t)}{Y^k}dP_t,$$

and

$$\tilde{\mathcal{G}}^l = \frac{1-\beta}{\bar{\varepsilon}_{w,1-\tau_k}} \sum_{t=0}^{\infty} \beta^t \varepsilon_{w_t,1-\tau_k} \int g_t(\boldsymbol{x}^t) \frac{(1-\tau_l'(y^l(\boldsymbol{x}^t))y_t^l(\boldsymbol{x}^t)}{(1-\bar{\tau}_l')Y^l} dP_t.$$

Proof. See Appendix A.2.2. □

We observe that condition (B.2) is very similar to the optimality condition (5) in the main text. However, some modifications are required. Specifically, $\bar{\varepsilon}_K$ is replaced by $\hat{\varepsilon}_K$, $\bar{g}^k$ is replaced once with $\bar{G}^k$ and once with $\bar{\mathcal{G}}^k$, and $\tilde{g}^l$ is replaced with $\tilde{\mathcal{G}}^l$.

The object $\bar{G}^k$ is the lifetime analogue of $\bar{g}^k$. It measures by how much the planner values, on average, a marginal dollar in the hands of capital tax payers. By contrast, $\bar{\mathcal{G}}^k$ measures by how much the planner values a marginal dollar in the hand of those agents, who benefit from the increase in gross capital returns. Since the adjustment process of capital takes time, agents with high capital income in the beginning may have low capital income later and vice versa because idiosyncratic abilities are not any more perfectly persistent.

Similarly, the lifetime analogue of $\tilde{g}^l$,

$$\tilde{G}^l = (1-\beta) \sum_{t=0}^{\infty} \beta^t \int g_t(\mathbf{x}^t) \frac{(1-\tau_l'(y^l(\mathbf{x}^t))y_t^l(\mathbf{x}^t)}{(1-\bar{\tau}')Y^l} dP_t,$$

and $\tilde{\mathcal{G}}^l$ may differ because of potential differences in the planner's valuation of the payers of labor taxes and the beneficiaries from wage reductions.

Finally, observe that the average discounted capital elasticity $\hat{\varepsilon}_{K,1-\tau_k}$ discounts each time-t capital elasticity $\varepsilon_{K_t,1-\tau_k}$ by the product of β^t and the period-t average marginal social welfare weight $\bar{g}_t$.

In the stochastic setting, the main complication is that marginal social welfare weights are generally time-varying. Specifically, each agent generally values differently an additional dollar of consumption in two different periods $t \neq s$. However, the Corollary below shows that under the welfare criterion 1 above the two optimality conditions (5) and (B.2) coincide.

Corollary B.1. *Let Assumptions 1-3 be satisfied. Furthermore, assume that social welfare function is such that $\omega_t(\boldsymbol{x}^t) = \omega(\boldsymbol{x}_t(\boldsymbol{x}^t)) = \omega(\boldsymbol{x}_t)$. Then*

$$\begin{aligned}\tau_k =& \frac{\beta^{t^a}(1-\bar{g}^k)}{\bar{\varepsilon}_{K,1-\tau_k}} - \frac{\bar{\varepsilon}_{L,1-\tau_k}}{\bar{\varepsilon}_{K,1-\tau_k}} \frac{\alpha^l}{\alpha^k} \left[\mathrm{E}_\Gamma[\tau_l'] + \mathrm{Cov}_\Gamma\left(\tau_l', \frac{y^l \bar{\varepsilon}_{l,1-\tau_k}}{Y^l \bar{\varepsilon}_{L,1-\tau_k}}\right)\right] \\ &+ \frac{\bar{\varepsilon}_{w,1-\tau_k}}{\bar{\varepsilon}_{K,1-\tau_k}} \frac{\alpha^l}{\alpha^k} \left[(1-\tau_k)\bar{g}^k - (1-\bar{\tau}_l')\tilde{g}^l + \tau_k - \bar{\tau}_l'\right].\end{aligned}$$

Proof. Follows directly from Proposition B.1 and the fact that the welfare criterion implies $\hat{\varepsilon}_{K,1-\tau_k} = \bar{\varepsilon}_{K,1-\tau_k}$, $\bar{G}^k = \bar{\mathcal{G}}^k = \bar{g}^k$ and $\tilde{\mathcal{G}}^l = \tilde{g}^l$. □

As a consequence, with a utilitarian welfare criterion, the optimality condition for optimal capital tax rates is given by condition (5) even in the stochastic setting. Thus, the condition reveals the key elasticities that drive the optimal tax rate, independent of the true underlying data generating process. As long the statistics entering the right hand side are estimated accurately, the condition provides a test for optimality.

C Methodology to Recover Unmeasured Policy Elasticities

As discussed in the main text, the discounted capital elasticity $\bar{\epsilon}_{K,1-\tau_k}$, as well as the discounted aggregate and individual labor elasticities $\bar{\epsilon}_{L,1-\tau_k}$ and $\bar{\epsilon}_{l(k_0,\eta),1-\tau_k}$, are what Hendren (2016) refers to as "policy elasticities", which measure the causal effect of a concrete policy experiment. For example, $\epsilon_{K_t,1-\tau_k}$ summarizes the reaction of the capital stock in period t following an increase in the net-of-tax rate $1-\tau_k$ in period zero.

As in Saez and Stantcheva (2018), changes in the capital tax rate τ_k and in the transfer T induce different income and substitution effects on capital and labor supply that are all subsumed in the policy elasticities $\epsilon_{K_t,1-\tau_k}$ and $\epsilon_{L_t,1-\tau_k}$. The main difference in my framework is that the demand for production factors does not fully accommodate the changes in supply, such that in order to restore equilibrium on factor markets, interest rates and wages need to adjust, which in turn impacts demand and supply, and so forth. The elasticities $\epsilon_{K_t,1-\tau_k}$ and $\epsilon_{L_t,1-\tau_k}$ are the aggregate responses that capture this whole equilibrium adjustment process.[C.1] Consequently, in my framework the net-of-capital-tax elasticity of wages is positive, $\bar{\epsilon}_{w,1-\tau_k} > 0$, and hence (by Lemma 1) the net-of-capital-tax interest elasticity is negative, $\bar{\epsilon}_{r,1-\tau_k} < 0$, while in Saez and Stantcheva (2018) both are equal to zero by assumption.

As pointed out by Kleven (2021), "[policy elasticities] can be used only to measure the welfare effect of an actually implemented reform (compared to the counterfactual of no reform). They cannot be used to assess the welfare effect of any other counterfactual reform that could be implemented" (Kleven, 2021, Remark 2).

In the following, I describe the methodology laid out in Section 4 in more detail, that is I explain how one can express the unmeasured policy elasticities in terms of factor supply elasticities, for which we have actual evidence.

C.1 Relation between Equilibrium Labor- and Capital Elasticities

The policy elasticity of equilibrium effective labor in period t with respect to the net-of-capital tax rate can be decomposed as

$$\epsilon_{L_t,1-\tau_k} = \tilde{\epsilon}_{L_t,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{L_t,T_s}\epsilon_{T_s,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{L_t,\bar{r}_s}\epsilon_{r_s,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{L_t,w_s}\epsilon_{w_s,1-\tau_k},$$

where for any $i \in \{1-\tau_k, w_s, r_s, T_s\}$. I denote by

$$\tilde{\epsilon}_{L_t,i} = \frac{\partial \ln L_t}{\partial \ln i}$$

[C.1] Hendren (2016) considers a simple linear production structure that rules out general equilibrium price effects in the main text. However, he discusses general equilibrium effects in the online Appendix D, where he argues that "when policies have general equilibrium effects, one also needs to track the causal impact of the policy on prices [...] The causal effects are still the desired responses, but one needs to also know the general equilibrium effects of government policies."

the elasticity of labor *supply* in period t with respect to i. These elasticities are defined in a ceteris paribus way, holding all taxes, transfers and prices except i fixed (hence the partial instead of the total derivative).

It considerably simplifies the analysis if income effects on labor supply are ruled out. Specifically, Assumption 4 implies that labor supply in a given period t is only affected through changes in the wage in the same period as the following Lemma shows.[C.2]

Lemma C.1. ***Own- and Cross Price Elasticities of Labor Supply without Income Effects.*** *Let Assumptions 1-4 be satisfied. For all (k_0, η) and for all $t \geq 0$ we have*

$$\tilde{\epsilon}_{l_t(k_0,\eta),1-\tau_k} = \tilde{\epsilon}_{l_t(k_0,\eta),T_s} = \tilde{\epsilon}_{l_t(k_0,\eta),r_s} = 0 \; \forall s \quad \text{and} \quad \tilde{\epsilon}_{l_t(k_0,\eta),w_s} = 0 \; \forall s \neq t.$$

whereas

$$\tilde{\epsilon}_{l_t(k_0,\eta),w_t} = \frac{\gamma_l(k_0,\eta)\left(1 - p(y^l(k_0,\eta))\right)}{1 + \gamma_l(k_0,\eta) p(y^l(k_0,\eta))}.$$

Thereby,

$$\gamma_l(k_0,\eta) = \frac{v'(l(k_0,\eta))}{l(k_0,\eta) v''(k_0,\eta)}$$

denotes the Frisch elasticity of labor supply and

$$p(y^l) = -\frac{\partial \ln(1 - \tau_l'(y^l))}{\partial \ln(y^l)} = \frac{y^l \tau_l''(y^l)}{1 - \tau_l'(y^l)}$$

denotes the local rate of labor tax progressivity.

Proof. See Appendix A.2.3. □

These results together with Lemmas 1 imply the following corollary.

Corollary C.1. ***Relation between Policy Elasticities.*** *Let Assumptions 1-4 be satisfied. Then the net-of-capital-tax elasticity of equilibrium labor in period t is given by*

$$\epsilon_{L_t,1-\tau_k} = \frac{\frac{Y^k}{\sigma Y}\tilde{\epsilon}_{L_t,w_t}}{1 + \frac{Y^k}{\sigma Y}\tilde{\epsilon}_{L_t,w_t}} \epsilon_{K_t,1-\tau_k}$$

and the net-of-capital-tax elasticities of equilibrium factor prices are given by, respectively,

$$\epsilon_{r_t,1-\tau_k} = -\frac{\frac{Y^l}{\sigma Y}}{1 + \frac{Y^k}{\sigma Y}\tilde{\epsilon}_{L_t,w_t}} \epsilon_{K_t,1-\tau_k} \quad \text{and} \quad \epsilon_{w_t,1-\tau_k} = \frac{\frac{Y^k}{\sigma Y}}{1 + \frac{Y^k}{\sigma Y}\tilde{\epsilon}_{L_t,w_t}} \epsilon_{K_t,1-\tau_k}.$$

[C.2] With this assumption the Frisch-, Hicksian- and Marshallian elasticities of labor supply all coincide. Allowing for income effects would require estimates of all three elasticities rather than one single wage-elasticity of labor supply.

Proof. From Lemma C.1 we know that $\epsilon_{L_t,1-\tau_k} = \tilde{\epsilon}_{L_t,w_t}\epsilon_{w_t,1-\tau_k}$. Plugging in the expression of $\epsilon_{w_t,1-\tau_k}$ from Lemma 1 and rearraning terms gives the result. □

We have expressed the the equilibrium labor elasticities in terms of estimated statistics and the equilibrium capital elasticities $\epsilon_{K_t,1-\tau_k}$. Thus, what is left to do is to relate the latter to actual evidence.

C.2 The Equilibrium Capital Elasticities

The equilibrium capital elasticities can be decomposed in an analogous way to the equilibrium labor elasticity,

$$\epsilon_{K_t,1-\tau_k} = \tilde{\epsilon}_{K_t,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s}\epsilon_{T_s,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,r_s}\epsilon_{r_s,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,w_s}\epsilon_{w_s,1-\tau_k}, \tag{C.1}$$

where

$$\tilde{\epsilon}_{K_t,i} = \frac{\partial \ln K_t}{\partial \ln i}$$

denotes the elasticity of capital *supply* in period t with respect to i holding all other tax instruments and prices fixed.

Lemma 1 already expressed $\epsilon_{w_s,1-\tau_k}$ and $\epsilon_{r_s,1-\tau_k}$ in terms of estimated statistics and the unmeasured equilibrium capital elasticities. The following Lemma does the same for the tax elasticity of the transfer $\epsilon_{T_s,1-\tau_k}$.

Lemma C.2. ***Decomposition of Revenue Effect.*** *Let Assumptions 1-4 be satisfied. Then for all $s \geq 0$ the elasticty of the transfer with respect to the net-of-tax rate can be decomposed as*

$$\begin{aligned}\epsilon_{T_s,1-\tau_k} = & \underbrace{\frac{Y^k}{T}\left[-1_{t\geq t^a}(1-\tau_k) + \tau_k\epsilon_{K_s,1-\tau_k}\right]}_{\epsilon^{ex}_{T_s,1-\tau_k}} \\ & \underbrace{+\frac{Y}{T}\frac{\epsilon_{K_s,1-\tau_k}}{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L_s,w_s}}\tilde{\alpha}^k\alpha^l\left[\left(\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}) + \bar{\tau}_l'\right) - \tau_k\right]}_{\epsilon^{end}_{T_s,1-\tau_k}}.\end{aligned}$$

Proof. See Appendix A.2.4. □

The first line, $\epsilon^{ex}_{T_s,1-\tau_k}$, captures the revenue effect of capital tax reductions when prices are exogenous. The government mechanically loses $-1_{t\geq t^a}(1-\tau_k)\frac{Y^k}{T}$ in revenue. However, because capital tax reductions encourage investment, there would be a positive 'behavioral' effect of $\tau_k\frac{Y^k}{T}\epsilon_{K_s,1-\tau_k}$. In the absence of income effects on labor supply, this would be the only revenue effects if prices were constant.

The second line, $\epsilon^{end}_{T_s,1-\tau_k}$ captures the revenue effects due to changing factor prices. If labor supply were to be inelastic, this effect would be proportional to $\bar{\tau}_l' - \tau_k$. When labor supply is elastic the increase in wages accompanied by the investment increase, induces higher labor effort. The associated revenue effect is positive and proportional to $\mathrm{E}_\Gamma[\tau_l']\tilde{\epsilon}_{L_s,w_s} + \mathrm{Cov}_\Gamma\big(\tau_l', y^l\bar{\epsilon}_{l,1-\tau_k}\big)$.

Because of Lemma C.2 one can decompose the equilibrium capital elasticity as follows.

Corollary C.2. ***Decomposition of Equilibrium Capital Elasticity.*** *Let Assumptions 1-4 be satisfied. Then*

$$\begin{aligned}\epsilon_{K_t,1-\tau_k} =& \tilde{\epsilon}_{K_t,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s}\epsilon^{ex}_{T_s,1-\tau_k} \\ &+ \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,r_s}\epsilon_{r_s,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,w_s}\epsilon_{w_s,1-\tau_k} + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s}\epsilon^{end}_{T_s,1-\tau_k}.\end{aligned} \tag{C.2}$$

Proof. Follows directly from Lemma C.2. □

The first line describes the elasticity of the equilibrium capital stock if factor prices were to be constant. This is the policy elasticity entering the tax formulas in Saez and Stantcheva (2018). Note that this elasticity itself consists of two terms. The first one, $\tilde{\epsilon}_{K_t,1-\tau_k}$ is the elasticity of capital with respect to the net-of-capital-tax rate if the transfer were not to adjust. However, the fact that the policy experiment requires a budget neutral change in the transfer implies that also agents' reactions to the change in this transfer need to be accounted for. The terms in the second line capture that equilibrium prices, and hence also the equilibrium transfer, adjust. These price and transfer responses in turn affect agents' capital supply.

Lemma C.3. ***Tax-Elasticity of Capital Supply.*** *Let Assumptions 1-4 be satisfied and let $t^a = 0$. Then for all $t \geq 0$ we have that*

$$\tilde{\epsilon}_{K_t,1-\tau_k} = t\beta\frac{C}{K}\gamma_c,$$

where

$$\gamma_c = -\int \frac{c_0(k_0,\eta)}{C}\frac{u_c(k_0,\eta)}{c_0(k_0,\eta)\big(u_{cc}(k_0,\eta)\big)}d\Gamma$$

denotes the consumption weighted average inter-temporal elasticity of substitution. As a consequence we have

$$\tilde{\epsilon}_{K_t,1-\tau_k} = t\tilde{\epsilon}_{K_1,1-\tau_k}.$$

Proof. See Appendix A.2.5. □

This result implies that from the supply elasticity of capital in period one (or any other arbitrary period following the tax change) one can recover the whole path $\{\tilde{\epsilon}_{K_t,1-\tau_k}\}_{t=1}^{\infty}$.

Estimates. As discussed in the main text, arguably the currently best available estimates are those obtained by Jakobsen et al. (2020) using administrative Danish data. The authors exploit two different natural experiments emanating from a 1989 wealth tax reform, with which they estimate the elasticity of wealth with respect to wealth taxes for (i) households between the 97.6th and 99.3rd percentile of the wealth distribution and (ii) households in the top percentile of the wealth distribution. They refer to the former as the "moderately wealthy" and to the latter as the "very wealthy", a classification which I adopt in the following. While I used their estimates on the former group of households in the main text, I report both in what follows.

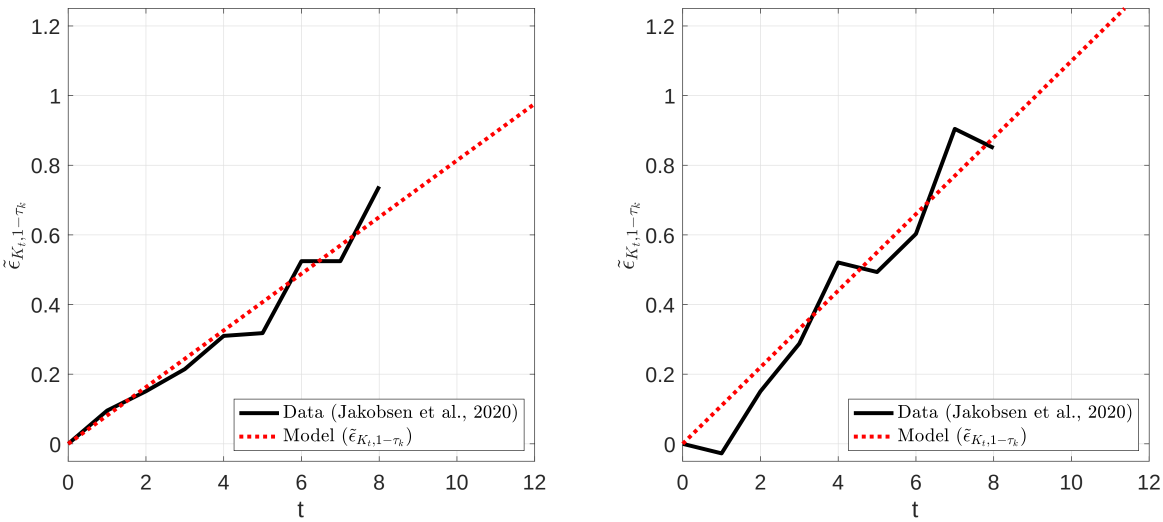

Figure C.1: Capital Supply Elasticity: data (solid line) from Jakobsen et al. (2020), Figures V (left panel) and VII (right panel); treatment on the treated; net-of-wealth-tax elasticities are translated to net-of-capital-tax elasticities using the return of $r = 6.58\%$; model (dotted line), $\tilde{\epsilon}_{K_t,1-\tau_k} = t\tilde{\epsilon}_{K_1,1-\tau_k}$.

The black solid lines in Figure 1 depict their estimated wealth elasticity with respect to the net-of-capital-tax rate for the first eight years following the reform, where the net-of-wealth-tax elasticities are translated into net-of-capital-tax elasticities using the benchmark return on capital ($r = 6.58\%$). By Lemma C.3, this elasticity is linear in time, $\tilde{\epsilon}_{K_t,1-\tau_k} = t\tilde{\epsilon}_{K_1,1-\tau_k}$, implying that in principle an estimate one year after the reform is sufficient to obtain the whole path of supply elasticities. However, in order to make use of all the available evidence, I regress their estimates $\{\hat{\epsilon}_{K_1,1-\tau_k}, \hat{\epsilon}_{K_2,1-\tau_k}, ..., \hat{\epsilon}_{K_8,1-\tau_k}\}$ on time (red dotted line). The left panel shows the estimates for the "moderately wealthy" and the right panel those of the "very wealthy". In both cases the model implied linear relationship does square very well with the data.

Lemma C.4. ***Policy Elasticity with Exogenous Prices.*** *Let Assumptions 1-4 be satisfied and let* $t^a = 0$. *In addition assume that* $F_{kl}(k,l) = 0$ *for all* $(k,l) \geq 0$. *Then*

$$\epsilon^{ex}_{K_t,1-\tau_k} = t\frac{1-\beta}{1-\beta(1-\tau_k r)}\tilde{\epsilon}_{K_1,1-\tau_k} \leq t\tilde{\epsilon}_{K_1,1-\tau_k} = \tilde{\epsilon}_{K_t,1-\tau_k}.$$

Proof. See Appendix A.2.6. □

The policy elasticity $\epsilon^{ex}_{K_t,1-\tau_k}$ of Lemma C.4 is the one entering the tax formulas in the exogenous factor price setting of Saez and Stantcheva (2018). Contrary to the supply elasticity $\tilde{\epsilon}_{K_t,1-\tau_k}$ of Lemma C.3, it takes into account the effect of budget neutral adjustments in the transfer on agents' savings decision. Observe that when the initial capital tax rate is $\tau_k = 0$ the policy elasticity with constant prices coincides with the pure supply elasticity to capital tax changes, that is $\epsilon^{ex}_{K_t,1-\tau_k} = \tilde{\epsilon}_{K_t,1-\tau_k}$. The reason is that in this case the additional savings induced by the lowering of the capital tax rate do not generate additional revenue and therefore no additional unearned income to the agents. However, when $\tau_k > 0$ the investment increase induced by the tax cut increases tax revenue and therefore the agents' transfer income over time. Agents want to partially consume their higher future government income, which reduces their savings. As a consequence, the policy elasticity is muted.

I derived Lemmas C.3 and C.4 by making use of envelope conditions. Specifically, using households' intra- and inter-temporal optimality conditions, one can recover the unmeasured elasticities $\{\tilde{\epsilon}_{K_t,1-\tau_k}\}_{t=9}^{\infty}$ and $\{\epsilon^{ex}_{K_t,1-\tau_k}\}_{t=1}^{\infty}$ with the estimates in Jakobsen et al. (2020). To derive the policy elasticities with endogenous prices, I use the same principle. Before stating the next proposition, which relates the unmeasured policy elasticities ($\epsilon_{K_t,1-\tau_k}$) to the estimated supply elasticity in period one ($\tilde{\epsilon}_{K_1,1-\tau_k}$), I shall briefly sketch its proof.

First, denoting the net return to capital by $\bar{r} = (1-\tau_k)r$, the expression (C.1) can be simplified to[C.3]

$$\begin{aligned}\epsilon_{K_t,1-\tau_k} &= \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,\bar{r}_s}\left[1 + \epsilon_{r_s,1-\tau_k}\right] \\ &\quad + \sum_{s=0}^{\infty} \tilde{\epsilon}_{K_t,T_s}\left[\epsilon_{T_s,1-\tau_k} + \frac{(1-\bar{\tau}_l')Y^l + (1-\mathrm{E}_\Gamma[\tau_l'])Y^l\tilde{\epsilon}_{L,w}}{T}\epsilon_{w_s,1-\tau_k}\right],\end{aligned}$$

which decomposes the policy elasticity into two weighted sums of price effects $\{\tilde{\epsilon}_{K_t,\bar{r}_s}\}_{s=0}^{\infty}$ and income effects $\{\tilde{\epsilon}_{K_t,T_s}\}_{s=0}^{\infty}$.[C.4] Second, again using the optimality conditions of the households' optimization problem, I derive expressions for all these price and income effects that again only depend on $\tilde{\epsilon}_{K_1,1-\tau_k}$, the capital supply response in the first period. Third, I use Lemma C.2 to substitute out $\epsilon_{T_t,1-\tau_k}$, and subsequently Corollary C.1 to substitute out $\epsilon_{L_t,1-\tau_k}$, $\epsilon_{r_t,1-\tau_k}$ and $\epsilon_{w_t,1-\tau_k}$. This yields an equation with the period-t policy elasticity $\epsilon_{K_t,1-\tau_k}$ on the left hand side and the whole the sequence of policy elasticities $\{\epsilon_{K_s,1-\tau_k}\}_{s=1}^{\infty}$ on the right hand side. Finally, I solve this infinite system of equations (for $t = 1, 2, 3, \ldots$) using a guess-and-verify approach. The guess is educated by the well known fact that in the neighborhood of the steady state the speed of convergence in the neoclassical growth model and related frameworks is constant.

The following proposition summarizes the result and thereby the main methodological contribution of this paper.

[C.3] Relative to the proof in the Appendix, here I already use that when the labor tax schedule features a constant rate of progressivity, we have $\mathrm{Cov}_\Gamma(\tau_l', y^l\tilde{\epsilon}_{l_s,w_s}) = 0$.

[C.4] Note that $\tilde{\epsilon}_{K_t,1-\tau_k} = \sum_{s=0}^{\infty}\tilde{\epsilon}_{K_t,\bar{r}_s}$.

Proposition C.1. ***Policy Elasticity with Endogenous Prices.*** *Let Assumptions 1-4 be satisfied and let $t^a = 0$. In addition assume that $F_{kl}(k,l) > 0$ for some $(k,l) > 0$. Then the path $\epsilon_{K_t,1-\tau_k}$ of equilibrium capital elasticities with respect to the net-of-capital-tax rate is given by*

$$\epsilon_{K_t,1-\tau_k} = \Big(1 - (\lambda(\boldsymbol{s}))^t\Big)\epsilon_{K_\infty,1-\tau_k} \quad \forall t \geq 0,$$

where the long-run capital elasticity is given by

$$\epsilon_{K_\infty,1-\tau_k} = \frac{\alpha^k}{\alpha^l}\frac{\sigma + \tilde{\alpha}^k \tilde{\epsilon}_{L,w}}{\tilde{\alpha}^k} < \infty$$

and $\lambda(\boldsymbol{s}) \in (0,1)$ is a constant that depends only on the vector of sufficient statistics

$$\boldsymbol{s} = \left(\tilde{\epsilon}_{K_1,1-\tau_k}, \tilde{\epsilon}_{L,w}, r, \tau_k, \sigma, \frac{Y^l}{Y}, \frac{Y^k}{Y}\right)$$

and satisfies

$$\frac{d\lambda(\boldsymbol{s})}{d\tilde{\epsilon}_{K_1,1-\tau_k}} < 0.$$

Proof. See Appendix A.2.7. □

Proposition C.1 expresses the policy elasticity $\epsilon_{K_t,1-\tau_k}$ in terms of actually estimated objects. The long-run capital elasticity is finite. The increase in capital supply following a reduction in τ_k is not fully accommodated for by capital demand as the marginal product of capital decreases. Consequently, the equilibrium interest rate declines, discouraging investment. In the long run, the equilibrium capital stock will hence settle at a finite level.

The capital supply elasticity $\tilde{\epsilon}_{K_1,1-\tau_k}$ determines how quickly the capital stock grows to its final level, that is the speed of convergence. Specifically, the higher $\tilde{\epsilon}_{K_1,1-\tau_k}$, the quicker the policy elasticity converges to its long run value $\epsilon_{K_\infty,1-\tau_k}$.

Figure 2 plots the path of policy elasticities implied by exogenous prices (Lemma C.4, dashed line) and endogenous prices (Proposition C.1, dash-dotted line). The left panel is the same as in Figure 2 of the main text, while the right panel depicts the policy elasticities when the supply elasticities of the "very wealthy" in Jakobsen et al. (2020) are targeted.

The discounted average elasticities that enter the optimality condition are then simply weighted averages over the paths of elasticities:

Corollary C.3. ***Discounted Average Capital Elasticity.*** *Let Assumptions 1-4 be satisfied and let $t^a = 0$.*

(a) Let $F_{kl}(k,l) = 0$ for all $(k,l) \geq 0$. Then the discounted average capital elasticity is given by

$$\bar{\epsilon}^{ex}_{K,1-\tau_k} = (1-\beta)\sum_{t=0}^{\infty}\beta^t \epsilon_{K_t,1-\tau_k} = \frac{\beta}{(1-\beta)\big(1-\beta(1-\tau_k r)\big)}\tilde{\epsilon}_{K_1,1-\tau_k}.$$

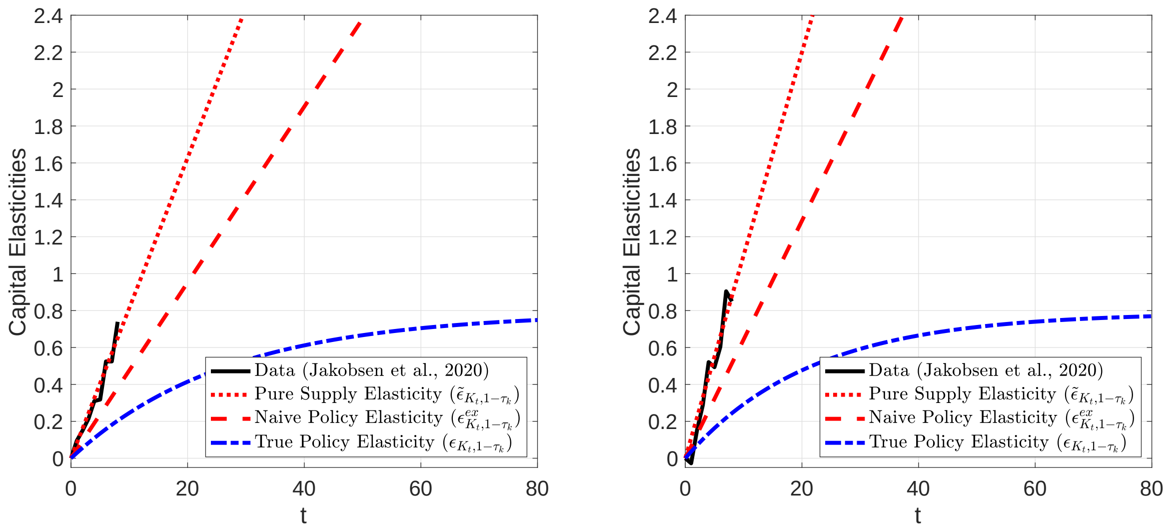

Figure C.2: Equilibrium Capital Elasticities: black solid line and red dotted line as in Figure C.1; red dashed line: policy elasticities with exogenous prices ($\sigma = \infty$); blue dash-dotted line: policy elasticities with endogenous prices ($\sigma = 0.6$, $\gamma_l = 0.5$). Left (right) panel based on responses of "moderately wealthy" ("very wealthy") in Jakobsen et al. (2020).

(b) Let $F_{kl}(k,l) > 0$ for some $(k,l) \geq 0$. Then the discounted average capital elasticity is given by

$$\bar{\epsilon}_{K,1-\tau_k} = (1-\beta)\sum_{t=0}^{\infty}\beta^t \epsilon_{K_t,1-\tau_k} = \frac{\beta(1-\lambda(\boldsymbol{s}))}{1-\lambda(\boldsymbol{s})\beta}\frac{\alpha^k}{\alpha^l}\frac{\sigma + \tilde{\alpha}^k\tilde{\epsilon}_{L,w}}{\tilde{\alpha}^k}.$$

Proof. Part (a) follows directly from Lemma C.4, part (b) from Proposition C.1. □

Part (a) of Corollary C.3 gives the discounted average capital elasticity $\bar{\epsilon}_{K,1-\tau_k}$ under the assumption of exogenous prices. This is the policy elasticity entering the formulas in Saez and Stantcheva (2018). By contrast, part (b) gives the discounted average capital elasticity when prices are endogenous.

C.3 Idiosyncratic Risk

In Appendix B, I show that for a large set of social welfare functions, the theoretical results derived in the main text extend to a more general model with uninsurable idiosyncratic labor- and capital income risk. However, the mapping of the unmeasured policy elasticities $\varepsilon_{K_t,1-\tau_k}$ and $\varepsilon_{L_t,1-\tau_k}$ to actually estimated supply elasticities, is derived within the deterministic modeling structure of the main text. In principle, these elasticities may differ with a stochastic income process. For example, it is well known that in the standard incomplete markets model of Aiyagari (1994) agents save partially for precautionary reasons and such savings motive may somewhat reduce the capital elasticity. However, as is convincingly argued in Krusell and Smith (1998) (see in particular their Sections III.D. and IV.C.), because most of the capital is held by rich agents, whose savings behavior is guided mainly by intertemporal concerns rather than by insurance motives, these modelling features have only a very small quantitative impact on the evolution of aggregate capital. I hence abstract from incorporating risk when deriving the mapping.

D Sensitivity Analysis

D.1 Sensitivity with Respect to the Return on Wealth

In this section, I perform the same quantitative analysis as in the main text assuming that the return on wealth is $r = 5\%$, respectively $r = 9\%$. Different assumption on capital depreciation are needed for the new interest rates to be consistent with equilibrium. In turn this affects the net capital income shares, whereas the gross capital income share of $\tilde{\alpha}^k = 0.4$ is unaffected. Specifically, with $r = 5\%$ we have $\delta = 0.054$ and $\alpha^k = 0.242$, while with $r = 9\%$ we have $\delta = 0.014$ and $\alpha^k = 0.365$.

Figure D.1 repeats the analysis of Figure 1 in the main text. In particular, the black lines translate the net-of-wealth-tax elasticities of wealth that are measured by Jakobsen et al. (2020) to net-of-capital-tax elasticities, where the capital tax only applies to the return on wealth. The left (right) panel shows the so obtained elasticities when a return of $r = 5\%$ ($r = 9\%$) is assumed. Naturally, different assumptions on the return change the implied net-of-capital-tax elasticities quite substantially. For example, while with $r = 5\%$ we have for the eight-year elasticity $\hat{\epsilon}_{K_8,1-\tau_k} \approx 0.5$, with $r = 5\%$ we have $\hat{\epsilon}_{K_8,1-\tau_k} \approx 1.1$.

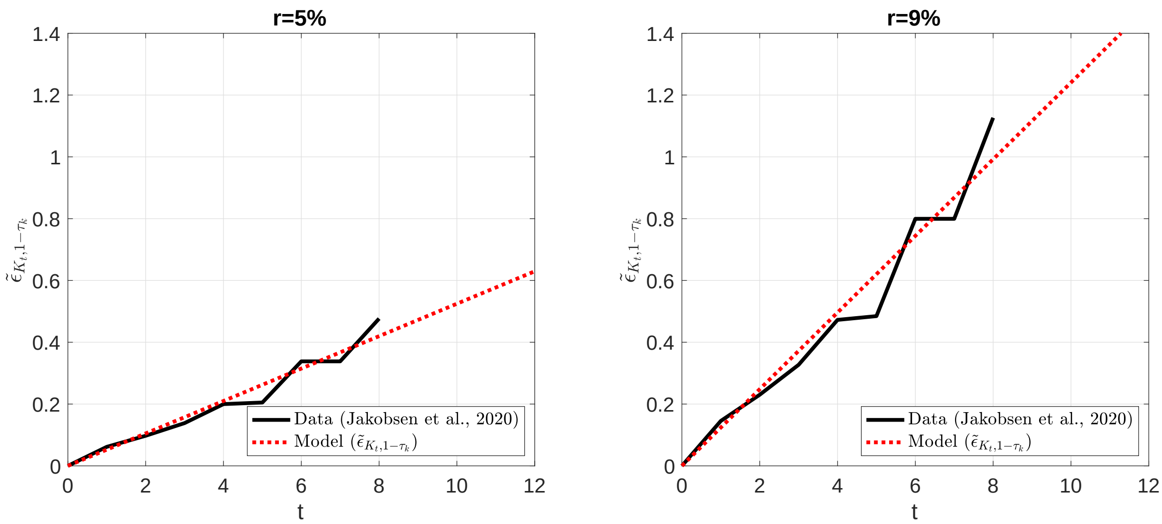

Figure D.1: Capital Supply Elasticity: data (solid line) from Jakobsen et al. (2020), Figure V; treatment on the treated; net-of-wealth-tax elasticities are translated to net-of-capital-tax elasticities using the return of $r = 5\%$ (left panel) and $r = 9\%$ (right panel); model (dotted line), $\tilde{\epsilon}_{K_t,1-\tau_k} = t\tilde{\epsilon}_{K_1,1-\tau_k}$.

This, in turn affects the deducted policy elasticities, which are depicted in Figure D.2 for the benchmark values of the substitution- and labor supply elasticities. The corresponding average discounted equilibrium elasticities are summarized in Table D.1. Assuming exogenous prices, the policy elasticity that is consistent with the estimated supply responses increases from 1.05 when $r = 5\%$ to 1.38 when $r = 9\%$. With endogenous prices and the benchmark capital-labor substitution elasticity elasticity of $\sigma = 0.6$, the policy elasticity increases from 0.30 ($r = 5\%$) to 0.49 ($r = 9\%$).

How the different elasticity values impact the marginal excess burden, is summarized in Table D.2. The total excess burden increases with the return on capital. It is 7 cents per mechanically raised

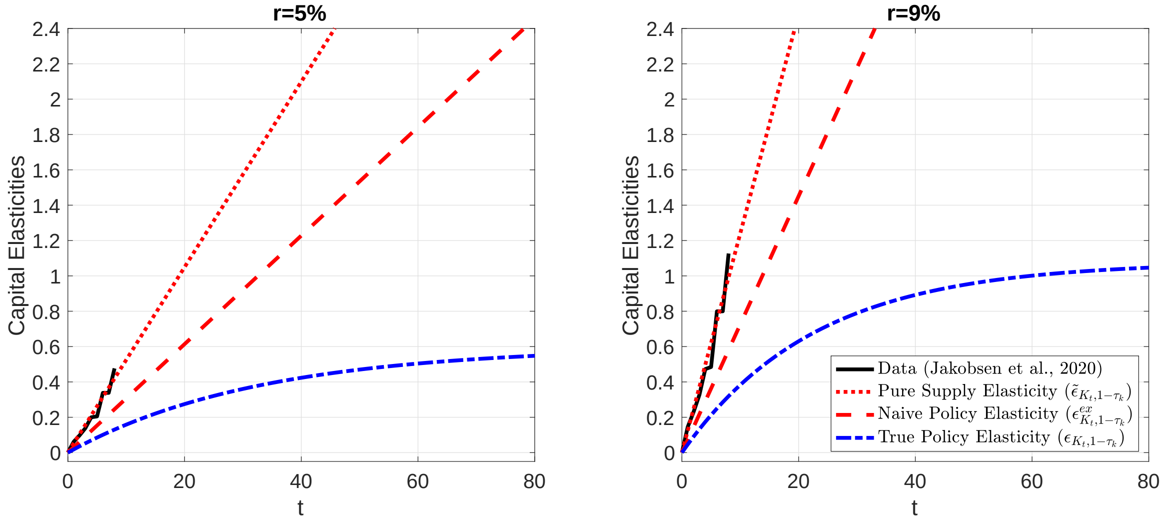

Figure D.2: Equilibrium Capital Elasticities: black solid line and red dotted line as in Figure 1; red dashed line: policy elasticities with exogenous prices ($\sigma = \infty$); blue dash-dotted line: policy elasticities with endogenous prices ($\sigma = 0.6$, $\gamma_l = 0.5$); assumed return $r = 5\%$ (left panel) and $r = 9\%$ (right panel).

Return on capital r	endogenous prices ($\sigma = 0.6$)	exogenous prices ($\sigma = \infty$)
5.00%	0.304	1.049
6.58%	0.385	1.237
9.00%	0.485	1.378

Table D.1: Discounted Average Elasticities: $\bar{\epsilon}_{K,1-\tau_k}$ for different values of the return (r); Frisch elasticity $\gamma_l = 0.5$.

dollar with $r = 5\%$ but 22 cents with $r = 9\%$. The composition of the various subcomponents is similar as with in the benchmark interest rate of $r = 6.58\%$.

Return on capital r	MEB_K	MEB_L	MEB_P	MEB
5.00%	0.2157	0.0215	−0.1642	0.0729
6.58%	0.2589	0.0196	−0.1497	0.1287
9.00%	0.3441	0.0190	−0.1456	0.2174

Table D.2: Decomposition of the Marginal Excess Burden: Components of the marginal excess burden (MEB); numbers in dollar per mechanical dollar in capital tax revenue raised; MEB_K: loss in capital income tax revenue due to lower savings; MEB_L: loss in labor income tax revenue due to lower labor supply; MEB_P: revenue impact of changing factor prices due to differential taxation of capital and labor; Frisch elasticity: $\gamma_l = 0.5$; capital-labor substitution elasticity: $\sigma = 0.6$.

Figure D.3 depicts the welfare effects of a marginal increase in the capital tax rate, the case of $r = 5\%$ in the left and the case of $r = 9\%$ in the right panel. In both cases, the equity effect exhibits a similar downward sloping shape as in the benchmark. The main difference is the due to the difference in the marginal excess burden that reduces the welfare gains when the interest rate is higher. Yet, in either case doe the bottom 60 percent of the US income distribution gain from capital tax increases. When $r = 5\%$ the status quo is optimal for households in the 71th percentile, while with $r = 9\%$ it is optimal for households around the 67th percentile. Thus, overall the qualitative – and to a large extent quantitative – features are similar as in the benchmark case.

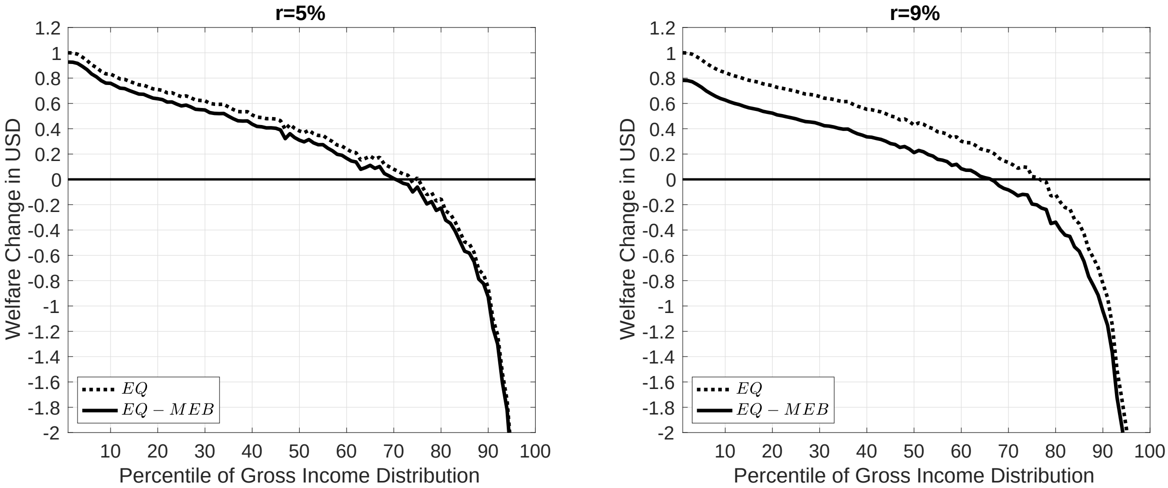

Figure D.3: Welfare Change: in USD per dollar of revenue mechanically raised; EQ: equity effect, MEB: marginal excess burden; value p on x-axis corresponds to the social welfare function that concentrates the whole welfare weight at percentile p of the total gross income distribution; Frisch elasticity of labor supply: $\gamma_l = 0.5$; substitution elasticity: $\sigma = 0.6$.

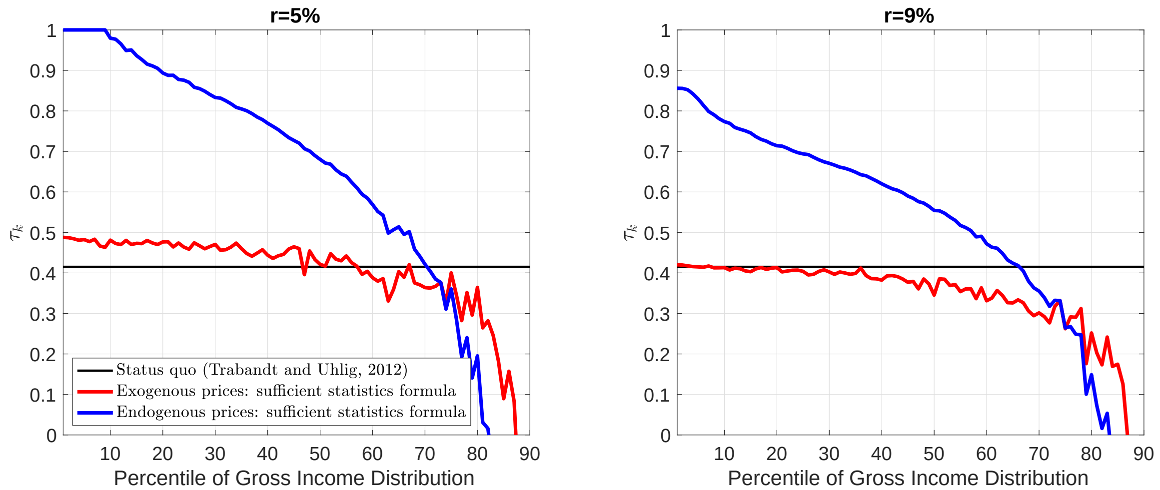

Figure D.4: Optimal Capital Tax Rates: sufficient statistics formula; value p on the x-axis corresponds to the social welfare function that concentrates the whole welfare weight at percentile p of the total gross income distribution; capital-labor substitution elasticities $\sigma = 0.6$ (endogenous prices) and $\sigma = \infty$ (exogenous prices); benchmark Frisch elasticity of labor supply ($\gamma_l = 0.5$).

Finally, Figure D.4 shows the 'optimal' tax rates, that is the tax rates predicted by the sufficient statistics formula (9) together with those predicted by the formula (10) that assumes exogenous factor prices. As is discussed in Section 5.3 and, in more detail, in Appendix E, this condition becomes somewhat inaccurate when it predicts tax rates far away from the status quo. Thus, especially for welfare functions that value only the lowest earners, the 'true' optimal tax rates can be expected to be slightly lower than those in the graph. Yet, we observe that overall the picture is similar as with the benchmark interest rate. Naturally, given the lower excess burden, optimal tax rates are higher when the interest rate is lower both in the case with exogenous and endogenous prices. A robust quantitative feature is that the bottom 60% of the US income distribution would like to see significant increases in the capital tax rate, though the desired tax rates are strongly declining in labor income.

D.2 Sensitivity with Respect to Capital-Labor Substitution and Labor Supply Elasticities

In this section, I provide a sensitivity analysis of my results with respect to variations in the substitution elasticity σ and the Frisch elasticity of labor supply γ_l. Specifically, on top of my benchmark elasticity of $\sigma = 0.6$, I consider $\sigma = 0.3$, which is at the low end of the empirical range, the Cobb-Douglas case ($\sigma = 1$), which is typically assumed in the parametric macroeconomics literature, as well as a very high value of $\sigma = 1.6$, which was put forward by Piketty (2014) and which is above all empirical estimates I am aware of. As for the Frisch elasticity of labor supply, I consider, on top of my benchmark value $\gamma_l = 0.5$, the case of completely inelastic labor supply ($\gamma_l = 0$) and the case of $\gamma_l = 1$, which should bracket most of the empirical evidence.

Discounted Average Tax-Elasticities of Capital and Labor. Table D.3 summarizes values of the discounted average elasticities of capital and labor that are consistent with the quasi-experimental evidence from Jakobsen et al. (2020) for the different combinations of substitution- and Frisch elasticities.

$\gamma_l \backslash \sigma$	0.3	0.6	1.0	1.6	∞
	Discounted average capital elasticity: $\bar{\epsilon}_{K,1-\tau_k} = (1-\beta)\sum_{t=0}^{\infty}\beta^t\epsilon_{K_t,1-\tau_k}$				
0.0	0.202	0.335	0.462	0.594	1.237
0.5	0.272	0.385	0.497	0.617	1.237
1.0	0.321	0.422	0.523	0.634	1.237
	Discounted average labor elasticity: $\bar{\epsilon}_{L,1-\tau_k} = (1-\beta)\sum_{t=0}^{\infty}\beta^t\epsilon_{L_t,1-\tau_k}$				
0.0	0.000	0.000	0.000	0.000	0.000
0.5	0.091	0.077	0.065	0.053	0.000
1.0	0.154	0.133	0.114	0.094	0.000

Table D.3: Discounted Average Elasticities of Capital and Effective Labor: $\bar{\epsilon}_{K,1-\tau_k}$ (upper panel) and $\bar{\epsilon}_{L,1-\tau_k}$ (lower panel) for different values of substitution elasticities (σ) and Frisch elasticities of labor supply (γ_l).

The upper panel summarizes the values of the discounted capital elasticity $\bar{\epsilon}_{K,1-\tau_k}$. While the Frisch elasticity has only a very small effect on capital equilibrium capital accumulation, there is considerable variability with regards to the substitution elasticity. Generally, $\bar{\epsilon}_{K,1-\tau_k}$ is increasing in σ. In the polar case with perfect factor substitutability it is $\bar{\epsilon}_{K,1-\tau_k} = 1.24$. The more complementary capital and labor are, the stronger the endogenous response of the gross interest rate due to tax changes. Since this endogenous equilibrium effect mitigates the change in the net return to capital, it has a moderating effect on the elasticity of capital. For the range of empirically plausible estimates, it is between one 0.2 and 0.63, that is substantially lower than the naive elasticity that one obtains when assuming constant factor prices.

The lower panel summarizes the discounted labor elasticities. In the absence of income effects, this elasticity is zero when factor prices are constant ($\sigma = \infty$). Naturally, it is also zero when labor supply is assumed to be inelastic ($\gamma_l = 0$). With the benchmark value for the Frisch elasticity of $\gamma_l = 0.5$ we have $\bar{\epsilon}_{L,1-\tau_k} \in (0.05, 0.09)$, while with the high end estimate of $\gamma_l = 1$ we have $\bar{\epsilon}_{L,1-\tau_k} \in (0.09, 0.15)$ for the range of plausible values of σ. Observe that the elasticity of equilibrium labor supply increases in the degree of complementarity between capital and labor, since the wage responses are stronger when complementarity is higher.

γ_l	MEB_K	MEB_L	MEB_P	MEB
		Low-end Substitution Elasticity: $\sigma = 0.3$		
0.0	0.143	0.000	−0.207	−0.064
0.5	0.193	0.024	−0.186	0.032
1.0	0.228	0.041	−0.171	0.098
		Benchmark Substitution Elasticity: $\sigma = 0.6$		
0.0	0.238	0.000	−0.172	0.066
0.5	0.273	0.021	−0.158	0.136
1.0	0.299	0.036	−0.148	0.187
		Cobb-Douglas Case: $\sigma = 1.0$		
0.0	0.328	0.000	−0.142	0.186
0.5	0.353	0.017	−0.133	0.237
1.0	0.371	0.031	−0.126	0.276
		Very high-end Substitution Elasticity: $\sigma = 1.6$		
0.0	0.421	0.000	−0.114	0.307
0.5	0.438	0.014	−0.109	0.353
1.0	0.450	0.025	−0.104	0.371
		Constant Factor Prices: $\sigma = \infty$		
0.0	0.878	0.000	0.000	0.878
0.5	0.878	0.000	0.000	0.878
1.0	0.878	0.000	0.000	0.878

Table D.4: Decomposition of the Marginal Excess Burden: Components of the marginal excess burden (MEB) for different values of substitution elasticities (σ) and Frisch elasticities of labor supply (γ_l); numbers in dollar per mechanical dollar in capital tax revenue raised; MEB_K: loss in capital income tax revenue due to lower savings; MEB_L: loss in labor income tax revenue due to lower labor supply; MEB_P: revenue impact of changing factor prices due to differential taxation of capital and labor.

Marginal Excess Burden. Table D.4 summarizes the three components of MEB for the same combinations of substitution- and Frisch elasticities as above. The lowest panel corresponds to the case of exogenous factor prices ($\sigma = \infty$). As discussed in the main text, the exclusion of income effects implies that the decomposition is trivial in this case. Specifically, absent changes in the equilibrium wage, a change in the capital tax rate will not affect labor supply and hence keep labor income tax revenue constant ($MEB_L = 0$). Furthermore, assuming away factor price changes implies that that $MEB_P = 0$ too. Consequently, the total marginal excess burden consists exclusively of the revenue loss due to a reduction in agents' savings. This revenue loss of $MEB = MEB_K = 0.88$, however, is substantial.

By contrast, with any combination of substitution- and Frisch elasticity in the range of empirical evidence, is the marginal excess burden is below 40 cents per mechanically raised dollar. As discussed above, a rise in the capital tax rate increases the gross return to capital. This mitigates the equilibrium reduction in the net return $(1 - \tau_k)r$, which has a moderating effect on the investment decline. Consequently, the capital tax revenue loss coming from the investment reduction MEB_K is much lower than in the case with exogenous prices. Furthermore, when the substitution elasticity is finite, the capital tax induced reduction in wages lowers labor supply and thus negatively affects labor income tax revenue whenever the labor supply elasticity is positive ($\gamma_l > 0$). Naturally the contribution of MEB_L to the overall excess burden is increasing in the labor supply elasticity. Since the equilibrium wage decline is stronger when capital and labor are strong compliments, MEB_L is decreasing in the substitution elasticity σ. Finally, the decrease in wages and the accompanied increase in the gross return to capital have a direct revenue impact themselves. Specifically, since in the status quo capital is taxed at a higher average rate than labor ($\tau_k > \bar{\tau}_l'$), the effect of changing factor prices on revenue is positive, that is the price responses' contribution to the excess burden is negative ($MEB_P < 0$). This effect is significant, an order of magnitude higher in absolute value than MEB_L. Consequently, we have in all cases $MEB < MEB_K$.[D.1]

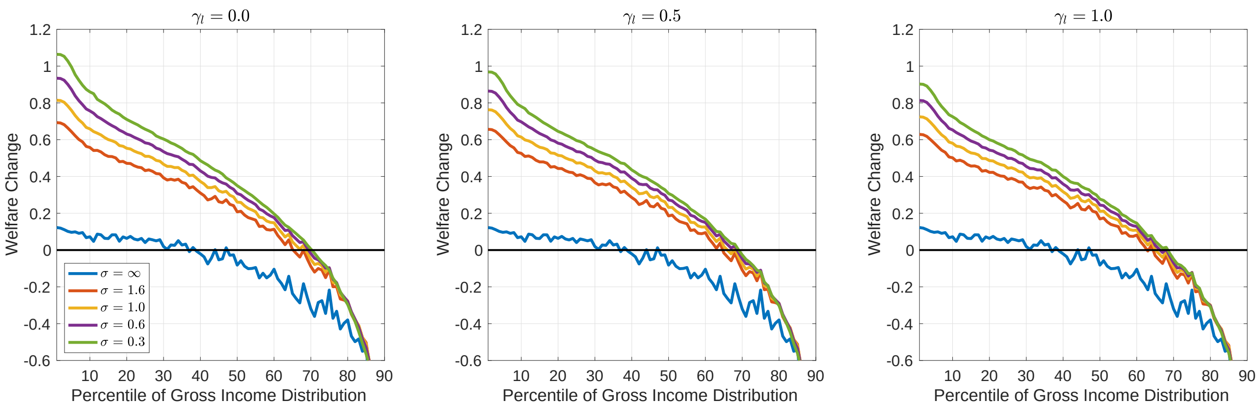

Figure D.5: Welfare Effect of a Marginal Tax Increase: in USD per dollar of revenue mechanically raised; varying substiution elasticity σ; $r = 6.58\%$; value p on the x-axis corresponds to the social welfare function that concentrates the whole welfare weight at percentile p of the total gross income distribution.

The three panels of Figure D.5 show the welfare gains of capital tax increases across the income distribution for the three different values of γ_l. Furthermore, within each panel I depict the gains for all the different values of the capital-labor substitution elasticity. We observe that the more complementary capital and labor are (the lower σ is) the lower is the marginal excess burden, implying higher welfare gains for the very bottom of the distribution. However, since also the depressing wage ef-

[D.1] The negative value of -0.064 when $\sigma = 0.3$ and $\gamma_l = 0$ would imply that on top of the mechanical dollar raised, the government would receive an additional 6 cents. However, this is a somewhat pathological case. Specifically, the very high capital-labor complementarity ($\sigma = 0.3$) implies very large price movements, with a large positive impact on government revenue $MEB_P << 0$ since capital is taxed at a higher rate than labor. However, note that if the labor elasticity was really zero, the government could achieve more revenue directly by taxing labor at higher rates than capital, in which case MEB_P and thus MEB would be positive.

fects are stronger with a lower σ the welfare change is decreasing more strongly as one moves to the right of the income distribution. Furthermore, the welfare gains of capital tax increases are declining in γ_l (compare different panels) because the wage depressing effect induces stronger labor supply reductions when γ_l is high, increasing the excess burden. In all cases do the lines cross the x-axis between the 63rd and the 70th percentile of the total income distribution.

Overall, a robust finding is that for all combinations of σ and γ_l do the bottom 60 percent of the total income distribution desire capital tax increases, while the top 30 percent desire capital tax decreases. The gains of tax increases at the very bottom of the income distribution, and hence the optimal Rawlsian tax rates, depend crucially on the substitution elasticity. In particular, lower values of σ imply higher Rawlsian tax rates because stronger endogenous factor price responses result in a lower marginal excess burden of capital taxation.

E Global Solution Method with a Nested Parametric Model

As described in Section 5.3, when the sufficient statistics formula (9) predicts tax rates far away from the status quo – as it does for example with a Rawlsian welfare objective – the policy prescription may be inaccurate. The reason is that the elasticities and the distributional statistics entering the formula are endogenous to the tax system. Usually, they are estimated locally, that is small variations around the status quo tax system are exploited in the empirical analysis. Thus, a potential concern is that these statistics may change when taxes are substantially altered.

To address this worry, in this section I use a parametric version of my model, which I calibrate such that it (locally) replicates all the statistics entering the formula. I then perform a sequence of counter-factual tax changes, that is I compute the whole transitional equilibrium path for all variables when the tax rate is changed once-and-for-all from its status quo to a new tax rate $\tau_k \in \{..., 0.39, 0.40, 0.41, 0.42, 0.43, ...\}$, where the transfer T adjusts to ensure budget balance period-by-period. I compute the welfare effect of each of this potential tax changes for each agent in the model economy and thus find the tax rate τ_k that would maximize each agent's welfare.

Note that the simplified model of the main text exhibits an indeterminate steady state wealth distribution. Such a model is unsatisfying when one wants to explore the reasons of inequality, in which case it is important for the model to *endogenously* generate a realistic wealth distribution (which versions of my more general model in Appendix B are capable of). However, for the exercise I perform in the present section, the indeterminate wealth distribution is an appealing feature. It allows me to isolate the approximation error due to the use of sufficient statistics formula (which I am interested in here) from the approximation error made when the wealth and income distribution is not perfectly matched (an issue parametric models have to deal with in general). Specifically, I can pick joint density of (k_0, η), such that the joint distribution of wealth and labor income – including its precise correlation structure – in the model economy is *exactly identical* to the one in the SCF. Each type of agent (k_0, η) in my model economy corresponds to one observation in the SCF and I pick the mass of this type to equal the corresponding sampling weight in the SCF. The model will then, for any chosen social welfare function, exactly replicate the same (initial) values of $\bar{g}^k$ and $\bar{g}^l$ that are used in the sufficient statistics analysis of the main text.

I use GHH preferences

$$u(c,l) = \frac{\left(c - \frac{l^{1+\frac{1}{\gamma_l}}}{1+\frac{1}{\gamma_l}}\right)^{1-\frac{1}{\tilde{\gamma}_c}}}{1 - \frac{1}{\tilde{\gamma}_c}},$$

which satisfy Assumption 4. I pick the benchmark Frisch elasticity of labor supply of $\gamma_l = 0.5$.

Note that with GHH preferences the parameter $\tilde{\gamma}_c$ does not exactly coincide with the (weighted average) intertemporal elasticity of substitution γ_c (see Lemma C.3 for the definition). The latter is heterogeneous across the population. In any case, I calibrate the parameter $\tilde{\gamma}_c$ such that, with small tax changes, the model generated path of net-of-tax elasticities of the equilibrium capital stock $\{\epsilon_{K_t,1-\tau_k}\}_{t=1}^{\infty}$ exactly replicates the dash-dotted blue line in Figure 2. The fact that this is possible is a numerical confirmation that my analytically derived long-rund capital elasticity $\epsilon_{K_\infty,1-\tau_k}$ as well as the analytically derived speed of convergence λ (see Proposition C.1) are correct. I obtain a value of $\tilde{\gamma}_c = 0.5$ and a consumption weighted average intertemporal elasticity of substitution of $\gamma_c = 0.4$. This is well in the middle of empirical estimates, which is further reassuring me that the quasi-experimental estimates of Jakobsen et al. (2020), which I use to discipline the policy elasticities, are reasonable.

I pick the discount factor $\beta = 0.963$, which is consistent with a steady state interest rate of $r = 6.58\%$ and a status quo-capital income tax rate of $\tau_k = 0.415$.

Technology is characterized by a CES production function

$$F(K,L) = \left(\alpha^{\frac{1}{\sigma}} K^{\frac{\sigma-1}{\sigma}} + (1-\alpha)^{\frac{1}{\sigma}} L^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}},$$

where $\sigma = 0.6$ and α is calibrated such that the gross capital share is $\tilde{\alpha}^k = 0.4$. As explained in the main text, given labor income, wealth and an interest rate of $r = 6.58\%$ this is consistent with a capital depreciation rate of $\delta = 3.85\%$ per annum.

I use the same parameterisation of the constant-rate-of-progressivity labor tax code as in the main text ($p = 0.181$ and τ_0 calibrated to match the labor income weighted average marginal tax rate of 22.5%). I calibrate the transfer T and – as a residual in the government budget constraint – expenditures G, such that the model matches a transfer-expense ratio of 71% as reported by the OECD.[E.1]

As for the ability distribution, I follow the strategy of Saez (2001). Since in the data we do not observe η, w and l separately but only total labor income $y^l(\eta) = \eta w l$, I compute each agent's ability η, which rationalizes her observed labor income. Specifically, optimizing households must satisfy the intra-temporal labor supply condition, which is equivalent to

$$l = \left[(1-\tau_0)(1-p)(y^l)^{1-p}\right]^{\frac{1}{1+\frac{1}{\gamma_l}}},$$

and thus provides a mapping from observed labor income y^l to unobserved labor supply l. Given the model implied steady state wage w (pinned down by r through the capital-labor ratio), the

[E.1] For this purpose I include in transfers subsidies form the labor income tax schedule at low labor incomes. I obtain a ratio of lump-sum transfers to aggregate income of $T/Y = 10.5\%$. However, note that the precise calibration of T does not affect my results. Specifically, to the extent that a different T changes savings behavior, the parameter $\tilde{\gamma}_c$ adjusts in order to generate the same net-of-tax-elasticities of equilibrium capital. As for the welfare analysis, I present my welfare gains in monetary amounts. What matters in this respect is how many additional dollars the household receives, which is independent of the initial transfer once $\tilde{\gamma}_c$ is recalibrated to generate the same net-of-tax elasticities of the equilibrium capital stock, and hence the same net-of-tax elasticity of government revenue (see Appendix C.2).

agent's labor productivity is then given by

$$\eta = \frac{y^l(\eta)}{wl}.$$

Using this parameterization, I then perform the policy experiments explained above. I rank households according to their total initial gross income and find the tax rate τ_k that maximizes the respective income percentiles' welfare. The result is the dash-dotted blue line in Figure 7. We see that the sufficient statistics formula approximates the optimal tax rates obtained with this global method remarkably well, though it somewhat overstates the tax rates that are optimal for the bottom of the income distribution.